Deliverable

3.2 Exploring the limits of earthquake predictability

Deliverable information	
Work package	WP3
Lead	University of Bologna
Authors	Paolo Gasperini, UNIBO, Marcus Herrmann, UNINA, Sebastian Heinzl, GFZ, Max Werner, UBRIS, Ian Main, UEDIN
Reviewers	Warner Marzocchi, UNINA
Approval	[Management Board]
Status	Final
Dissemination level	Internal
Will the data supporting this document be made open access? (Y/N)	Yes
If No Open Access, provide reasons	
Delivery deadline	31.08.2022
Submission date	31.08.2022
Intranet path	DOCUMENTS/DELIVERABLES/RISE-Deliverable_3.2.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821115

Table of contents

Summary

This deliverable collects all the scientific results obtained during the first 36 months of RISE in the ambit of Task 3.2 (Enhancing earthquake predictability).
Such task was planned to explore the limits of earthquake predictability. Hypotheses investigated here are not necessarily yet ready to be implemented as a forecasting model, but have the potential to reduce the limits of earthquake predictability. The task is subdivided into three subtasks as shown in Table 1.

Subtask	Short Description
Subtask 1	Search for precursory spatiotemporal seismicity patterns before and after strong earthquakes
Subtask 2	Analysis of the reliability of the magnitude-independence assumption
Subtask 3	Search for additional explanatory variables in the triggering properties of earthquakes

Table 1. Breakdown of RISE Task 3.2

In subtask1 precursory spatiotemporal seismicity patterns before and after strong earthquakes were searched, using the homogenized and higher resolution catalogue already available for Italy and Southern California and subsequently the new ones developed in Task 2.4 for other EU regions. In the recent literature, the b-value of the Gutenberg- Richter (GR) frequency-magnitude distribution was hypothesized to be a proxy of differential stress (DS, the difference between minimum and maximum stress eigenvalues) within the Earth’s crust. In particular low b-values seems to be associated with high levels of DS and vice versa high b-values with low DS. Thus, observations of low b-values might indicate the phase of preparation of an impending strong earthquakes while high b-values a quiet period. This hypothesis can be investigated by analysing the time evolution of b-value computed using seismic catalogues with homogeneously determined magnitude. At present only the catalogues of Italy since 1995 and of Southern California since 1981 appears to comply such requirement. Other parameters that were planned to be studied are the a-value (productivity) of the GR and the parameters of various models of seismic sequence time-decay (e.g. p and c of the Omori law) based on likelihood analysis.

In subtask2, the reliability of the magnitude-independence assumption, i.e., the earthquake magnitude of future earthquakes is independent and identically distributed (usually, according to a truncated Gutenberg-Richter law) has been explored. In particular, in this subtask it has been investigated if the magnitude-frequency distribution varies in space and time analyzing the new generation of earthquake catalogs, which have expanded the detection capabilities increasing the number of earthquakes of a factor of 10 or more.

In subtask3, systematic empirical studies to search for additional explanatory variables in the triggering properties of earthquakes were conducted. Obvious candidates include (i) surface heat flow, (ii) geodetic strain-rate, (iii) thickness of the seismogenic zone, (iv) lithology (inferred rigidity, rheology if available), (v) plate tectonic setting, (vi) inferred regional stress field, (vii) triggering susceptibility, (viii) time since last major earthquake (on well-characterised faults), and some variables that can be measured during a seismic sequence such as (i) source focal mechanism, (ii) aseismic after slip moment, (iii) stress drop, and (iv) Shake Map footprint. Specifically, dependencies between these variables and various clustering properties including (i) size/timing/location of largest triggered event, (ii) triggering productivity, (iii) foreshock statistics, (iv) swarm-like behaviour has to be searched. The research benefitted from advances in observational capabilities (T2.4) and exploit computational statistics to uncover hidden relationships.

1. Chapter 1 - Search for precursory spatiotemporal seismicity patterns before and after strong earthquakes

Forecasting methods taken from the literature and newly developed were applied to Italy by the retrospective testing using the HOmogeneized InstRUmental Seismic Catalog (HORUS) of Italy from 1960 to present partially developed also in Task 2.4 of RISE (Lolli et al., Seismol. Res. Lett. 91, 3208–3222, doi: 10.1785/0220200148). In particular, Gasperini et al. (2021) developed an alarm-based forecasting method based on the occurrence of strong foreshocks. This was tested using the Molchan diagrams and the Area Skill score approaches. Considering an alarm duration of three months, the algorithm retrospectively forecast more than 70 per cent of all shocks (mainshocks+aftershocks) with Mw ≥ 5.5 occurred in Italy from 1960 to 2019 with a total space–time fraction covered by the alarms of the order of 2 per cent. Considering the same space–time coverage, the algorithm is also able to retrospectively forecasts more than 40 per cent of the mainshocks only with Mw ≥ 5.5 of the seismic sequences occurred in the same time interval.
Biondini et al. (2022) applied to Italy the EEPAS (Every Earthquake a Precursor According to Scale) forecasting model. EEPAS is a is a space–time point-process model based on the precursory scale increase phenomenon and associated predictive scaling relations. It has been previously applied to New Zealand, California and Japan earthquakes with target magnitude thresholds varying from about 5 to 7. In all previous application, computations were done using the computer code implemented in Fortran language by the model authors. Biondini et al. (2022) developed a suite of computing codes completely rewritten in Matlab and Python. They first compared the two software codes to ensure the convergence and adequate coincidence between the estimated model parameters for a simple region capable of being analyzed by both software codes, then using the rewritten codes they optimized the parameters for a different and more complex polygon of analysis using the catalog data from 1990 to 2011. Finally, they performed a retrospective (pseudo-prospective) forecasting experiment of Italian earthquakes from 2012 to 2021 with Mw≥5.0 and compares the forecasting skill of EEPAS with other forecasting models using the standard test developed in the ambit of the Collaboratory for the Study of Earthquake Predictability (CSEP). The EEPAS approached demonstrated to be slightly worser than ETAS for short forecasting windows (3 months) and better for longer windows (up to 10 years).
Another forecasting approach is that followed by Gulia et. al. (2020, 2021) for the application of the Traffic Light System (TLS) to the pseudo-prospective forecasting of Ridgecrest Mw 7.1 earthquake of July 2021, based on the temporal variation of the b-value of the frequency-magnitude (Gutenberg-Richter) relation. In normally decaying aftershock sequences, the b-value of the aftershocks was found, on average, to be 10%–30% higher than the background b-value. A drop of 10% or more in "aftershock" b-values was postulated to indicate that the region is still highly stressed and that a subsequent larger event is likely. In this Ridgecrest case study, after analyzing the magnitude of completeness of the sequences, they were able to determine reliable b-values over a large range of magnitudes within hours of the two mainshocks. They then find that in the hours after the first Mw 6.4 Ridgecrest event, the b-value drops by 23% on average, compared to the background value, triggering a red foreshock traffic light. Spatially mapping the changes in b values, they identify an area to the north of the rupture plane as the most likely location of a subsequent event. After the second, magnitude 7.1 mainshock, which did occur in that location as anticipated, the b-value increased by 26% over the background value, triggering a green traffic light. Finally, comparing the 2019 sequence with the Mw 5.8 sequence in 1995, in which no mainshock followed, they find a b-value increase of 29% after the mainshock. Their results suggest that the real-time monitoring of b-values is feasible in California and may add important information for aftershock hazard assessment.
Gulia and Gasperini (2021) observed that artifacts often affect seismic catalogs. Among them, the presence of man-made contaminations such as quarry blasts and explosions is a well-known problem. Using a contaminated dataset reduces the statistical significance of results and can lead to erroneous conclusions, thus the removal of such nonnatural events should be the first step for a data analyst. Blasts misclassified as natural earthquakes, indeed, may artificially alter the

seismicity rates and then the b-value of the Gutenberg and Richter relationship, an essential ingredient of several forecasting models.
At present, datasets collect useful information beyond the parameters to locate the earthquakes in space and time, allowing the users to discriminate between natural and nonnatural events. However, selecting them from webservices queries is neither easy nor clear, and part of such supplementary but fundamental information can be lost during downloading. As a consequence, most of statistical seismologists ignore the presence in seismic catalog of explosions and quarry blasts and assume that they were not located by seismic networks or in case they were eliminated. They show the example of the Italian Seismological Instrumental and Parametric Database. What happens when artificial seismicity is mixed with natural one?

Published papers

Gasperini, P., E. Biondini, B. Lolli, A. Petruccelli and G. Vannucci (2020). Retrospective short-term forecasting experiment in Italy based on the occurrence of strong (fore) shocks, Geophys. J. Int, 225, 1192–120. doi: 10.1093/gji/ggaa592.

Gulia, L., and P. Gasperini (2021). Contamination of Frequency–Magnitude Slope (b-Value) by Quarry Blasts: An Example for Italy, Seismol. Res. Lett. 92, 3538–3551, doi: 10.1785/0220210080.

Gulia, L., and S. Wiemer (2021). Comment on "Two Foreshock Sequences Post Gulia and Wiemer (2019)" by Kelian Dascher- Cousineau, Thorne Lay, and Emily E. Brodsky, Seismol. Res. Lett., 92, 3251-3258, doi: 10.1785/0220200428

Gulia, L., S. Wiemer, and G. Vannucci (2020). Pseudoprospective Evaluation of the Foreshock Traffic-Light System in Ridgecrest and Implications for Aftershock Hazard Assessment, Seismol. Res. Lett. 91, 2828–2842, doi: 10.1785/0220190307.

In preparation

Biondini, E., D. Rhoades, and P. Gasperini (2022). Application of the EEPAS seismic forecasting method to Italy.

2. Chapter 2 - Analysis of the reliability of the magnitude-independence assumption

Spassiani & Marzocchi (2021) proposed to model the MFD of seismic events that nucleate in a confined area with an energy-dependent tapered Gutenberg–Richter (GR) relation, (TGRE). TGRE acknowledges the elastic rebound theory in the sense that the probability for another large event to nucleate in the same area within a short time interval has to be lower than according to the (tapered) GR relation. The validity and applicability of the TGRE model is demonstrated for the 1992 M7.3 Landers sequence, California. As expected by TGRE, it was shown that the on-fault MFD differs from the off-fault MFD (lower corner magnitude), evidencing the magnitude independence assumption. The TGRE fits the magnitude–frequency distribution (MFD) of on-fault seismicity better than the tapered GR model. An ETAS model with TGRE could improve OEF, i.e., finding the highest probability for a large earthquake not where the previous large earthquake occurred.
Herrmann & Marzocchi (2021) inspected the magnitude–frequency distribution (MFD) of high-resolution catalogs at the example of the 2019 M7.1 Ridgecrest sequence, 2009 M6.3 L'Aquila Sequence, and of whole Southern California. They found that the MFD of small earthquakes in these catalogs does usually not comply with the exponential Gutenberg–Richter (GR) relation. In fact, when using this relation rigorously, high-resolution catalogs do not seem to offer a crucial benefit over ordinary catalogs. This impediment is mostly due to an improper mixing of different

magnitude types, spatiotemporally varying detection capabilities, or distorted data processing. Common methods to apply the GR relation do not detect these discrepancies. These findings are relevant for both producers of high-resolution catalogs and modelers that use MFDs of such catalogs.

Herrmann et al. (2022) reanalyzed the 2016–2017 central Italy sequence using a high-resolution catalog and introduced an alternative perspective for studying MFD variability—using a spatiotemporal scale that considers the 3-D distribution of recorded seismicity. This approach is based on Piegari et al. (2022) of the same group: using a cluster analysis of a sequence using density-based algorithms to spatially isolate the most seismogenic zones; temporal periods are defined by the occurrence time of the largest events. They demonstrate that this approach proves beneficial in resolving the spatiotemporal variation of the MFD and b-value. For instance, they resolved what happened in the days before the largest event (Norcia) in its associated seismogenic zone. Rather than solely focusing on b-value estimates, they exploited more information from the MFD, e.g., by assessing and comparing its exponential-like part and reporting the b-value stability as function of Mc. They showed that the MFD behaves in a complex manner among the spatially isolated clusters throughout the sequence. Their findings reflect on the appropriate spatiotemporal scale to resolve the b-value and challenge existing approaches.

Manganiello et al. (2022) re-examined foreshock activity in southern California to investigate the existence and main characteristics of foreshock sequences that cannot be explained by ETAS, i.e., anomalous foreshock sequences. In other words, they looked for new insights on the evidence against the cascade model. They performed different statistical tests and considered the potential influence of subjective choices, such as the method to identify mainshocks and their foreshocks. They found anomalous foreshock sequences mostly for mainshock magnitudes below 5.5. These anomalies preferentially occurred in zones of high heat flow, which were already known to host swarm-like seismicity. Outside these regions, the foreshocks generally behave as expected by ETAS. These findings will contribute to an improving earthquake forecasting (e.g., by stimulating the discrimination of swarm-like from ETAS-like sequences) and the understanding of earthquake nucleation processes (e.g., anomalous foreshock sequences are not indicating a pre-slip nucleation process, but swarm-like behavior driven by heat flow).

Published papers:

Herrmann, M., & Marzocchi, W. (2021). Inconsistencies and lurking pitfalls in the magnitude–frequency distribution of high-resolution earthquake catalogs. Seismological Research Letters, 92(2A), 909–922. https://doi.org/10.1785/0220200337. https://zenodo.org/record/4428319

Piegari, E., M. Herrmann, & W. Marzocchi (2022). 3-D spatial cluster analysis of seismic sequences through density-based algorithms. Geophysical Journal International 230(3). 2073–2088. https://doi.org/10.1093/gji/ggac160. https://zenodo.org/record/6671514

Spassiani, I., & Marzocchi, W. (2021). An Energy-Dependent Earthquake Moment–Frequency Distribution. Bulletin of the Seismological Society of America, 111(2), 762–774. https://doi.org/10.1785/012020190. https://zenodo.org/record/5510040

Accepted papers:

Herrmann, M., E. Piegari, & W. Marzocchi (2022). b-value of what? Complex behavior of the magnitude distribution during and within the 2016–2017 central Italy sequence. Nature Communication (accepted). Preprint: https://doi.org/10.21203/rs.3.rs-1210699/v1

Submitted (under review) papers:

Manganiello, E., M. Herrmann, & W. Marzocchi (2022). New physical implications from revisiting foreshock activity in southern California. Geophysical Research Letters. (submitted). Preprint: https://www.essoar.org/doi/10.1002/essoar.10509908.2

3. Chapter 3 - Search for additional explanatory variables in the triggering properties of earthquakes

Systematic empirical studies to search for additional explanatory variables in the triggering properties of earthquakes were conducted. Obvious candidates include (i) surface heat flow, (ii) geodetic strain-rate, (iii) thickness of the seismogenic zone, (iv) lithology (inferred rigidity, rheology if available), (v) plate tectonic setting, (vi) inferred regional stress field, (vii) triggering susceptibility, (viii) time since last major earthquake (on well-characterised faults), and some variables that can be measured during a seismic sequence such as (i) source focal mechanism, (ii) aseismic afterslip moment, (iii) stress drop, and (iv) ShakeMap footprint. Specifically, we will search for dependencies between these variables and various clustering properties including (i) size/timing/location of largest triggered event, (ii) triggering productivity, (iii) foreshock statistics, (iv) swarm-like behaviour. The research will benefit from advances in observational capabilities (-> 2.4) and exploit computational statistics to uncover hidden relationships.

In this regard, Bayliss et al. (2020) has developed a Bayesian framework to make inferences of the effect of the explanatory variables listed above on the Epidemic-Type Aftershock Sequence (ETAS) model parameters. This allows them to have a comprehensive representation of the uncertainty by calculating a full posterior distribution for each quantity of interest. The novelty of their approach is to represent the ETAS model as a Latent Gaussian model (LGm). This allows them to use the Integrated Nested Laplace Approximation (INLA) methodology to obtain the posterior distribution of the parameters. The INLA methodology is an alternative to MCMC techniques designed to handle large LGm's having parameters with complex covariance structures, specifically, it has been used extensively to study the effect of spatially or temporally (or both) varying covariates on a phenomenon of interest. Applications of the INLA methodologies range from finance to biostatistics. This creates a theoretical framework to include covariates in the ETAS model and to compare models based on different combinations of those. Moreover, the INLA algorithm is deterministic which makes the result more reproducible than simulation based techniques such as MCMC. Finally, this theoretical framework is easily extendible to consider the parameters as spatially and/or temporally variable, by using Gaussian Markov Random Fields with the parameters of the covariance function determined by the data.

Probabilistic earthquake forecasts estimate the likelihood of future earthquakes within a specified time-space-magnitude window and are important because they inform planning of hazard mitigation activities on different timescales. The spatial component of such forecasts, expressed as seismicity models, generally rely upon some combination of past event locations and underlying factors which might affect spatial intensity, such as strain rate, fault location and slip rate or past seismicity. Bayliss et al. (2022) for the first time, extend previously reported spatial seismicity models, generated using the open source inlabru package, to time-independent earthquake forecasts using California as a case study. The inlabru approach allows the rapid evaluation of point process models which integrate different spatial datasets. they explore how well various candidate forecasts perform compared to observed activity over three contiguous five year time periods using the same training window for the seismicity data. In each case they compare models constructed from both full and declustered earthquake catalogues. In doing this, they compare the use of synthetic catalogue forecasts to the more widely-used grid-based approach of previous forecast testing experiments. The simulated-catalogue approach uses the full model posteriors to create Bayesian earthquake forecasts. They show that simulated-catalogue based forecasts perform better than the grid-based equivalents due to (a) their ability to capture more uncertainty in the model components and (b) the associated relaxation of the Poisson assumption in testing. They demonstrate that the inlabru models perform well overall over various time periods, and hence that independent data such as fault slip rates can improve forecasting power on the time scales examined. Together, these findings represent a significant improvement in earthquake forecasting is possible, though this has yet to be tested and proven in true prospective mode.

Aseismic afterslip is postseismic fault sliding that may significantly redistribute crustal stresses and drive aftershock sequences. Afterslip is typically modeled through geodetic observations of surface deformation on a case-by-case basis, thus questions of how and why the afterslip moment varies between earthquakes remain largely unaddressed. Churchill et al. (2022) compiled 148

afterslip studies following 53 M 6.0–9.1 earthquakes, and formally analyzed a subset of 88 well-constrained kinematic models. Afterslip and coseismic moments scale near-linearly, with a median Spearman's rank correlation coefficient (CC) of 0.91 after bootstrapping (95% range: 0.89–0.93). They inferred that afterslip area and average slip scale with coseismic moment as $M_0^{2/3}$ and $M_0^{1/3}$, respectively. The ratio of afterslip to coseismic moment (M_{rel}) varies from <1% to >300% (inter-quartile range: 9%–32%). M_{rel} weakly correlates with M_0 (CC: −0.21, attributed to a publication bias), rupture aspect ratio (CC: −0.31), and fault slip rate (CC: 0.26, treated as a proxy for fault maturity), indicating that these factors affect afterslip. M_{rel} does not correlate with mainshock dip, rake, or depth. Given the power-law decay of afterslip, studies that started earlier and spanned longer timescales to capture more afterslip are expected, but M_{rel} does not correlate with observation start time or duration. Because M_{rel} estimates for a single earthquake can vary by an order of magnitude, it is proposed that modeling uncertainty currently presents a challenge for systematic afterslip analysis. Standardizing modeling practices may improve model comparability, and eventually allow for predictive afterslip models that account for mainshock and fault zone factors to be incorporated into aftershock hazard models.
Strong earthquakes cause aftershock sequences that are clustered in time according to a power decay law, and in space along their extended rupture, shaping a typically elongate pattern of aftershock locations. A widely used approach to model earthquake clustering, the Epidemic Type Aftershock Sequence (ETAS) model, shows three major biases. First, the conventional ETAS approach assumes isotropic spatial triggering, which stands in conflict with observations and geophysical arguments for strong earthquakes. Second, the spatial kernel has unlimited extent, allowing smaller events to exert disproportionate trigger potential over an unrealistically large area. Third, the ETAS model assumes complete event records and neglects inevitable short-term aftershock incompleteness as a consequence of overlapping coda waves. These three aspects can substantially bias the parameter estimation and lead to underestimated cluster sizes. Grimm et al. (2022) combine the approach of Grimm et al. (Bull. Seismol. Soc. Am. 112, 474–493, doi: 10.1785/0120210097), who introduced a generalized anisotropic and locally restricted spatial kernel, with the ETAS-Incomplete (ETASI) time model of Hainzl (Bull. Seismol. Soc. Am. 112, 494–507, doi: 10.1785/0120210146), to define an ETASI space-time model with flexible spatial kernel that solves the abovementioned shortcomings. We apply different model versions to a triad of forecasting experiments of the 2019 Ridgecrest sequence and evaluate the prediction quality with respect to cluster size, largest aftershock magnitude and spatial distribution. The new model provides the potential of more realistic simulations of on-going aftershock activity, e.g. allowing better predictions of the probability and location of a strong, damaging aftershock, which might be beneficial for short term

Published papers

Bayliss, K., M. Naylor J. Illian & I.G. Main (2020). Data-driven optimization of seismicity models using diverse datasets: generation, evaluation and ranking using inlabru, J. Geophys. Res: Solid Earth, https://doi.org/10.1029/2020JB020226

Churchill, R. M., Werner, M. J., Biggs, J., & Fagereng, Å. (2022). Afterslip Moment Scaling and Variability from a Global Compilation of Estimates, Journal of Geophysical Research: Solid Earth, e2021JB023897. https://doi.org/10.1029/2021JB023897.

Grimm, C., Hainzl, S., Käser, M., Küchenhoff, H. (2022): Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. - Stochastic Environmental Research and Risk Assessment. https://doi.org/10.1007/s00477-022-02221-2

In press

Bayliss, K., Naylor, M., Kamranzad, F. & I. Main (2022). Pseudo-prospective testing of 5-year earthquake forecasts for California using inlabru, Nat. Hazards Earth Syst. Sci. https://doi.org/10.5194/nhess-2021-403

4. Appendix - Papers

Geophysical Journal International

Geophys. J. Int. (2021) **225,** 1192–1206
Advance Access publication 2020 December 16
GJI Seismology

doi: 10.1093/gji/ggaa592

Retrospective short-term forecasting experiment in Italy based on the occurrence of strong (fore) shocks

P. Gasperini,[1,2] E. Biondini,[1] B. Lolli,[2] A. Petruccelli[1,3] and G. Vannucci[2]

[1] *Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, I-40127, Italy. E-mail: paolo.gasperini@unibo.it*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, I-40128, Italy*
[3] *Swiss Seismological Service, ETH Zurich, Zurich, CH-8092, Switzerland*

Accepted 2020 December 8. Received 2020 November 4; in original form 2020 April 28

SUMMARY

In a recent work, we computed the relative frequencies with which strong shocks ($4.0 \leq M_{\mathrm{w}} < 5.0$), widely felt by the population were followed in the same area by potentially destructive main shocks ($M_{\mathrm{w}} \geq 5.0$) in Italy. Assuming the stationarity of the seismic release properties, such frequencies can be tentatively used to estimate the probabilities of potentially destructive shocks after the occurrence of future strong shocks. This allows us to set up an alarm-based forecasting hypothesis related to strong foreshocks occurrence. Such hypothesis is tested retrospectively on the data of a homogenized seismic catalogue of the Italian area against a purely random hypothesis that simply forecasts the target main shocks proportionally to the space–time fraction occupied by the alarms. We compute the latter fraction in two ways (i) as the ratio between the average time covered by the alarms in each area and the total duration of the forecasting experiment (60 yr) and (ii) as the same ratio but weighted by the past frequency of occurrence of earthquakes in each area. In both cases the overall retrospective performance of our forecasting algorithm is definitely better than the random case. Considering an alarm duration of three months, the algorithm retrospectively forecasts more than 70 per cent of all shocks with $M_{\mathrm{w}} \geq 5.5$ occurred in Italy from 1960 to 2019 with a total space–time fraction covered by the alarms of the order of 2 per cent. Considering the same space–time coverage, the algorithm is also able to retrospectively forecasts more than 40 per cent of the first main shocks with $M_{\mathrm{w}} \geq 5.5$ of the seismic sequences occurred in the same time interval. Given the good reliability of our results, the forecasting algorithm is set and ready to be tested also prospectively, in parallel to other ongoing procedures operating on the Italian territory.

Key words: Earthquake hazards; Earthquake interaction, forecasting, and prediction; Statistical seismology.

INTRODUCTION

Even if the deterministic prediction of earthquakes is presently not feasible and perhaps it will never be (Geller *et al.* 1997), several methods of probabilistic operational forecasting have been proposed in the last decades (see Jordan & Jones 2010 and Jordan *et al.* 2011 for an overview). Many of such methods take advantage of the well-known property of earthquakes to cluster in space and time (Mulargia & Geller 2003; Kagan 2014) and in particular of the possibility that relatively small shocks, occurring in advance (foreshocks) of destructive main shocks, might be used as precursory signal.

Jones & Molnar (1976, 1979) first observed that the property of worldwide strong earthquakes of being preceded by a few days or weeks of smaller shocks could have been used to predict somehow their occurrence. Jones (1984, 1985) noted that in California the occurrence of a weak shock increased of several order of magnitude the probability of occurrence of a main shock in the following hours or days and Agnew & Jones (1991) and Jones (1994) computed the probability of a major earthquake along the San Andreas fault in California, given the occurrence of a potential foreshock nearby the fault. The occurrence of foreshocks was then adopted as one of possible precursor of large earthquakes by the Southern San Andreas Working Group (1991) and Reasenberg (1999a,b) estimated the prospective frequency of potential foreshock being followed by stronger earthquakes in California and worldwide.

In Italy, Caputo *et al.* (1977, 1983) analysed earthquakes' swarms as forerunners of strong earthquakes, Grandori *et al.* (1988) proposed an alarm system based on the occurrence of a pair of foreshocks, Console *et al.* (1993) and Console & Murru (1996) studied the foreshock statistics and their possible relationship to earthquake prediction and Di Luccio *et al.* (1997) and Console *et al.* (1999) set up a forecasting hypothesis for the occurrence of earthquakes conditioned by prior events.

More recently, Gasperini *et al.* (2016), by the retrospective analysis of a homogeneous seismic catalogue of the Italian region, computed the relative frequencies with which strong shocks (defined as $4.0 \leq M_w < 5.0$) were followed in the same area by potentially destructive main shocks (defined as $M_w \geq 5.0$, 5.5, 6.0). In particular, they found that just after strong shocks, the relative frequency of potentially destructive main shocks in the same area increases with respect to quiet periods by a factor up to about 100 000. Then, as time goes by without any main shock occurring, such factor decreases logarithmically down to less than 10 for time windows of months to years. Within one day after the occurrence of a strong shock, the frequencies of main shocks with $M_w \geq 5.0$ and ≥ 5.5 range from 5 per cent to 2 per cent while within one month they range from 14 per cent to 6 per cent. Frequencies remain quite stable for about one hour after the strong shock and then start to decrease logarithmically at a rate of about 1 per cent for a doubling of the time elapsed from the strong shock. The frequencies of large main shocks ($M_w \geq 6.0$) are generally lower than 1 per cent except from about one month after a strong shock with $4.5 \leq M_w < 5.0$ when they become of the order of 4 per cent, but they decrease well below 1 per cent about two or three months after the strong shock if the main shock did not actually occur in the meantime. About 30 per cent of main shocks have been preceded by strong shocks in the day before, about 50 per cent one in the month before and about 60 per cent in the year before.

All such evidences suggest us to formulate an alarm-based forecasting hypothesis related to the simple occurrence of strong shocks in a given area. In this work, we first set up such hypothesis and then optimize it by the retrospective analysis of the HOmogenized instRUmental Seismic catalogue (HORUS) of the Italian area from 1960 to 2019 (Lolli *et al.* 2020) which is an improved and updated version of the seismic catalogue used by Gasperini *et al.* (2016).

In our knowledge, this is the first alarm-based forecasting experiments applied to the Italian region after the one by Grandori *et al.* (1988) cited above and after Console *et al.* (2010) and Murru *et al.* (2009) who converted to an alarm-based approach previous probabilistic forecasting studies by Console & Murru (2001) and Console *et al.* (2003, 2006). In fact, the latter studies, as well as others forecasting efforts in Italy (see Schorlemmer *et al.* 2010 and Marzocchi *et al.* 2014 for an overview), mostly based on the Epidemic-Type Aftershock Sequence (ETAS) model (Kagan & Knopoff 1987; Ogata 1988), were developed to reproduce at best the general behaviour of future seismicity, not to issue a warning of a possibly impending damaging earthquake.

The present forecasting hypothesis will be possibly submitted for prospective testing and validation to the testing facilities of the Collaboratory Study of Earthquake Predictability (Jordan 2006; Zechar *et al.* 2010).

SETTING UP THE FORECASTING HYPOTHESIS

We issue an alarm of duration Δt within a circular area (CA) of radius R every time a strong shock with $M_{\min} \leq M < M_{\max}$ occurs inside the CA. As target events to be forecasted we consider all the shock, with magnitude above a threshold $M_m \geq M_{\max}$.

We must note that after the actual occurrence of a target shock, the forecast of further target shocks in the same area and in the following weeks or months is somehow favoured by the strong aftershocks of the previous target event. Hence, we also verify the ability of our method to forecast only the first target shock of each sequence. We then consider also a declustered set of target shocks obtained by eliminating those target shocks occurred within a distance $D =$ 50 km and a time window of a year after another target shock of the sequence, even if they are larger than the first target shock of the sequence. This kind of declustering is somehow different with respect to that adopted for example in seismic hazard assessment (e.g. Gardner & Knopoff 1974; Reasenberg 1985) in which each sequence is usually represented by the largest shock, even if it is not the first one in the sequence. We choose the declustering space and time windows based on our experience on past Italian seismic sequences but we also checked visually that none possible secondary main shock remains not declustered. Also note that the chosen declustering windows approximately correspond to those determined by the algorithm of Gardner & Knopoff (1974) for $M = 5.5$.

As source areas we consider a regular tessellation of the Italian territory made of partially overlapping CAs with fixed radius R. Starting from an initial CA, centred at latitude 47° and longitude 7°, we compute the centres of the neighbour CAs by moving with steps $D = R\sqrt{2}$ both in longitude (from 7° to 19°) and in latitude (from 47° to 36°) covering then the whole Italian area with partial overlapping (Fig. 1).

Based on the results of our previous analysis (Gasperini *et al.* 2016), we choose a radius $R = 30$ km, as a good compromise between the opposing demands of having short spatial resolution and a sufficiently high number of earthquakes within each CA, so obtaining a total of 695 partially overlapping CAs. However, as the completeness of the seismic catalogue is poor in offshore areas, we consider in our analysis only the CAs within which at least one earthquake with $M_w \geq 4.0$ occurred inland from 1600 to 1959 (so as to be independent of the seismicity from 1960 to 2019 that will be used for the retrospective testing and optimization of the forecasting method), according to the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

According to Gasperini *et al.* (2016), we consider as target shocks the earthquakes with $M_w \geq 5.0$, ≥ 5.5 and ≥ 6.0, which, in Italy, usually cause moderate, heavy and very heavy damage to buildings and none, a few and many victims respectively. Larger thresholds cannot be investigated because only three shocks with $M_w \geq 6.5$ (1976 Friuli with $M_w = 6.5$, 1980 Irpinia with $M_w = 6.8$ and 2016 Norcia with $M_w = 6.6$) occurred during the time interval covered by our seismic catalogue.

We count a success if a target shock occurs during one or more alarm time windows Δt and within one or more CA. On the contrary we count a missed forecast if a target shock occurs outside any alarm window of any CA. According to Molchan (1990, 1991), we compute the miss rate as

$$\nu = \frac{N - h}{N} \tag{1}$$

where h is the number of target events successfully forecasted and N is the total number of target events.

We also compute the total time duration d_c of alarms as the union of all alarm windows within each CA. This can also be computed by multiplying the window length Δt by the number n of issued alarms and then subtracting the sum of time intersections between

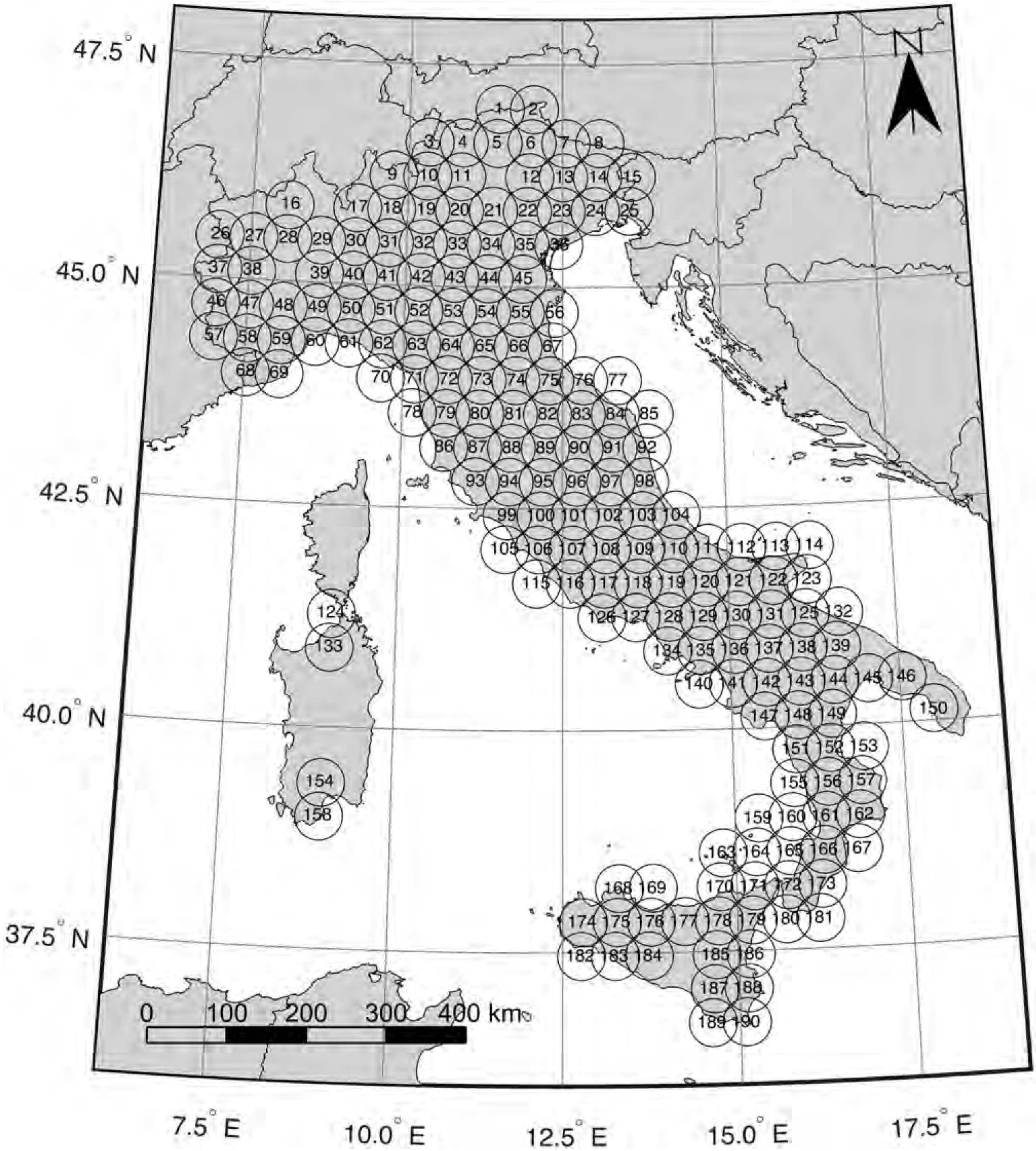


Figure 1. Tessellation of the Italian territory used for the retrospective forecast experiment. CA with $R = 30$ km within which at least one earthquake with $M_w \geq 4.0$ occurred inland from 1600 to 1959 according to the CPTI15 catalogue (Rovida *et al.* 2020).

alarm windows $\cap t_s$

$$d_c = \cup \Delta t = n \Delta t - \sum \cap t_s \tag{2}$$

The fraction of time occupied by alarms within each CA is then computed as

$$\tau_c = \frac{d_c}{T} \tag{3}$$

where T is the total duration of the forecasting experiment.

Finally, the overall fraction of space-time occupied by alarms is computed as the average of τ_c over all CAs

$$\tau_u = \frac{1}{M} \sum \tau_c \tag{4}$$

where M is the number of CAs. Note that such definition of fraction of space–time occupied by alarms is consistent with strong shocks occurring in the overlapping region of two adjoining CAs because in such case we sum the alarm fraction of time τ_c for both CAs.

Table 1. Magnitudes of completeness of the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

Magnitude threshold Mc	Time interval of completeness	ΔT (yr)
$M_w \geq 4.5$	1880–1959	80
$M_w \geq 5.0$	1880–1959	80
$M_w \geq 5.5$	1780–1959	180
$M_w \geq 6.0$	1620–1959	340

Following Shebalin *et al.* (2011), we also compute the fraction of space–time occupied by alarms by weighting each alarm with the long-term rate of earthquakes within each CA. We compute such rate based on the data of the CPTI15 catalogue (Rovida *et al.* 2016, 2020) using different completeness thresholds Mc for different time intervals from 1620 to 1959 (Table 1). We count the numbers of earthquakes $N(Mc)$ above each magnitude threshold Mc occurred within each CA and within the corresponding time interval of completeness $\Delta T(Mc)$. Then we compute for each magnitude threshold the expected rate λ (event yr^{-1}) of earthquakes with $M_w \geq 4.0$, assuming the b-value of the frequency–magnitude distribution (Gutenberg & Richter 1944) equal to 1 (Rovida *et al.* 2020):

$$\lambda = \frac{N(Mc)}{\Delta T(Mc)}\, 10^{M_c-4.0} \tag{5}$$

In each CA, we then compute the average λ_{ave} of rates $\lambda > 0$ from different magnitude thresholds. For those CAs for which such average frequency cannot be computed because there are no earthquakes within the completeness time window of any magnitude threshold, we assign the minimum rate computed overall.

Finally, the weighted fraction of space–time occupied by alarms is computed from all CAs as

$$\tau_w = \frac{\sum \lambda_{ave}\tau_c}{\sum \lambda_{ave}} \tag{6}$$

See the details of such computations for each CA in Table S1 of the Supporting Information.

DATA SET USED FOR TESTING AND OPTIMIZATION

To test and optimize our algorithm, we apply it retrospectively to the HORUS catalogue of Italian instrumental seismicity from 1960 to 2019 (Lolli *et al.* 2020). For the time interval from 1960 to 1980, HORUS coincides with the data set prepared by Lolli *et al.* (2018) and that can be downloaded from the electronic supplement of such paper. For the period from 1981 to 2019, it is obtained by merging various data sources and homogenizing the magnitudes to M_w as described by Gasperini *et al.* (2012, 2013). The catalogue used here is updated up to the end of 2019 but we have implemented an automatic procedure able to continuously update such catalogue in near real-time (with daily to hourly updates) through the downloading of new data from online sources and the application of magnitude conversions (Lolli *et al.* 2020). We provide the final catalogue on the web (https://doi.org/10.13127/HORUS) for public dissemination and the possible prospective testing of the present and other forecasting methods.

The magnitude completeness threshold for the period 1960–1980 has been assessed by Lolli *et al.* (2018) to be about 4.0 whereas, according to Gasperini *et al.* (2013), it is definitely lower for the successive time periods. Such thresholds might be definitely larger in offshore areas owing to the large distances from the closest seismic stations, which are usually located on land (excepting for a few instruments deployed on the sea bottom). This is the reason why we only consider earthquakes with $M_w \geq 4.0$ occurred within the 190 CAs containing one inland earthquake at least. As our interest is to forecast earthquakes that potentially threaten lives and goods, we also limit the analysis to shocks shallower than 50 km. We show in Fig. S1 of Supporting Information the spatial distribution of inland earthquakes from the HORUS catalogue (Lolli *et al.* 2020) with $M_w \geq 4.0$ and depth < 50 km used for testing and optimization and in Fig. S2 in the Supporting Information the time distribution of magnitudes of all inland earthquakes with depth < 50 km.

The catalogue provides uncertainties for all magnitude estimates, ranging from less than 0.1 (for M_w estimated by moment tensor inversion) to about 0.5 (for M_w proxies from body wave magnitude mb observed by a few stations). In general, magnitude and location errors have the effect to increase the randomness of the catalogue and then to penalize skilled forecasting methods with respect to unskilled ones.

Owing to the Gutenberg Richter (1944) law, errors tend on average to overestimate all magnitudes because there are more earthquakes below a given threshold which can be overestimated than earthquakes above the same threshold which can be underestimated. The larger the error the larger the overestimation.

On the other hand, magnitude errors are generally larger for small earthquakes because the latter are observed by less stations and because accurate method of magnitude determination, like moment tensor inversion, cannot be applied to them. This means that in general small earthquakes are overestimated more than larger ones and then that foreshocks are overestimated more than target shocks.

One possible consequence in the present case is that errors in magnitude might improperly increase the number of alarms and then the space–time fraction occupied by alarms, particularly in earlier times when the coverage of seismic networks was coarser, so that to slightly underestimate the real skill of the method. Conversely the number of target shocks should not be affected much by magnitude errors because in HORUS catalogue the most (about 80 per cent) of $M_w \geq 5.0$ are accurately computed by moment tensor inversions.

TESTING AND OPTIMIZING THE FORECASTING HYPOTHESIS

We here follow the approach proposed by Zechar & Jordan (2008, 2010) based on the so-called 'Molchan error diagram' (Molchan 1990, 1991; Molchan & Kagan 1992). The latter consists of a plot (e.g. Fig. 2) of the miss rate ν (eq. 1) as a function of the fractions of space–time occupied by alarms τ (τ_u of eq. 4 or τ_w of eq. 6). For a paradoxical forecasting method not issuing any alarm, the space–time occupied by alarms is 0 and no target events can be forecasted (all target event are missed) then it is represented by the point $(\tau, \nu) = (0, 100$ per cent$)$ at the upper left corner of the Molchan diagram. On the other hand, for a forecasting method issuing an alarm at any time and in any place, so occupying the entire space-time volume, no target events are missed and then the forecasting method is represented by the point $(\tau, \nu) = (100$ per cent$, 0)$ at the lower right corner of the diagram. The points on the diagonal line connecting such two points (e.g. the black continuous line in Fig. 2), with equation

$$\nu = 1 - \tau \tag{7}$$

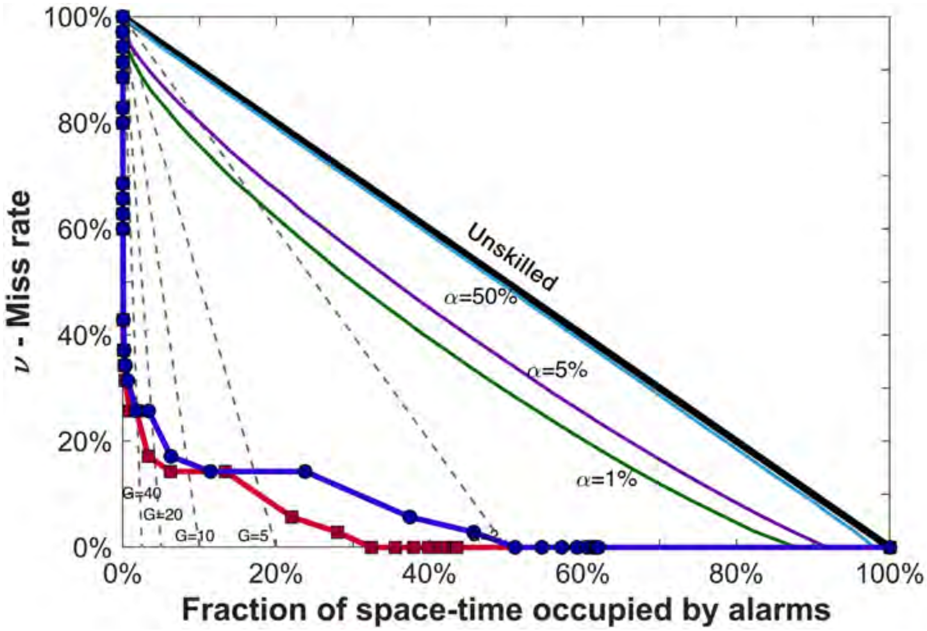


Figure 2. Molchan diagram for all target shocks with $M_{\mathrm{w}} \geq 5.5$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_{\mathrm{w}} < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the text). The black continuous line indicates a purely random forecasting method that separates skilled (below the line) from unskilled (above) forecasting methods. The light blue, violet and green lines indicate the confidence limits for $\alpha = 50$ per cent, 5 per cent and 1 per cent, respectively. The black dashed lines indicate probability gains $G = 2$, 5, 10, 20 and 40.

indicate the expected performance of a purely random forecasting method that simply forecasts target events proportionally to the space–time fraction occupied by the alarms.

On the diagonal line, the ratio between the success rate and the space-time fraction

$$G = \frac{1 - \nu}{\tau} \tag{8}$$

is 1 for any τ, while for a skilled forecasting method, located below the line, $G > 1$ represents the 'probability gain' factor with respect to the random case.

Following Zechar & Jordan (2008), τ (τ_u or τ_w) can be assumed as the probability of forecasting a target events by chance and then can be used to measure the performance of a forecasting method under the reasonable assumption that the probability of having exactly h successful forecasts over N targets is given by the binomial probability function

$$B\,(h|N\tau) = \binom{N}{h}\,(\tau)^h(1-\tau)^{N-h} \tag{9}$$

Then, the cumulative probability of having by chance h or more successful forecasts is

$$\alpha = \sum_{n=h}^{N} B\,(n|N\tau) = 1 - \sum_{n=0}^{h-1} B\,(n|N\tau) \tag{10}$$

Such statistic allows to measure the skill of a forecasting methods, given the miss rate ν and the fraction of space–time occupied by alarms τ. In particular, the lower the statistic the higher the skill. Moreover, by inverting eq. (10), we can compute the expected miss rate ν at a given τ, for a hypothetical forecasting method with given probability α, and then to plot confidence limits on the Molchan diagram (e.g. the blue, violet and green lines in Fig. 2).

This statistic can be used to validate a forecasting method using a prospective data set (collected after the final fixing of the forecasting hypothesis) but even to optimize the forecasting hypothesis by searching the values of the parameters of the forecasting algorithm (if any) for which the statistic is minimum, by using a retrospective data set.

A given forecasting method with fixed parameter values is represented by a single point (τ, ν) on the Molchan diagram. However, one can even consider curves (Molchan trajectories) connecting different points referred to the same general forecasting approach but obtained by varying one of the free parameters of the forecasting algorithm. In our case, we can vary the alarm time window Δt from 0 to the total duration T of the experiment. In this way, we span the total space–time occupied by the alarms and correspondingly the number of successful forecasts, which increase with increasing Δt.

In the light of such definition, the diagonal line in the Molchan diagram can be seen as the Molchan trajectory of a purely random forecasting method. If a forecasting method performs better than the random one, its trajectory mainly lies in the lower left half of the Molchan diagram below the random line.

Zechar & Jordan (2008, 2010) proposed to use as a measure of the performance of an alarm-based forecasting method the integral of the success rate function $1 - \nu_f(\tau)$ normalized to the alarm space–time coverage τ

$$a_f\,(\tau) = \frac{1}{\tau}\int_0^{\tau}\left[1 - \nu_f\,(t)\right]dt \tag{11}$$

As the integral corresponds to the area above the Molchan random trajectory, the statistic was named area skill (AS) score. The AS score is normalized so that its value ranges between 0 and 1: the larger the statistic the better the performance.

The expected value of the AS score for a purely random method can be derived by substituting the eq. (7) of the random line $\nu_f\,(t) =$

$1 - t$ in eq. (11). This gives

$$\langle a_f(\tau)\rangle = \frac{1}{\tau}\int_0^{\tau}[1-(1-t)]\,dt = \frac{1}{\tau}\,\frac{\tau^2}{2} = \frac{\tau}{2} \quad (12)$$

Such expectance function is represented in a plot as a function of τ by a straight line connecting the axes origin (0,0) with the point (100 per cent, 50 per cent) (e.g. the black line in Fig. 3). In such plot, the skilled forecasting methods lie above such random line.

Zechar & Jordan (2008, 2010) explored the AS score distribution and found that, for a continuous alarm function, the AS score at $\tau = 1$ is asymptotically Gaussian with a mean of 1/2 and a variance of 1/(12 N). They also found that the kurtosis excess is -6/(5 N) and hence, for N of the order of a dozen at least, the Gaussian approximation provides a good estimate of confidence bounds. Finally, they argued that even if the AS score can be computed for any τ, the power of the test tends to increase with increasing τ and therefore it is the best to use $a_f(\tau = 1)$ for hypothesis testing.

RESULTS OF RETROSPECTIVE TESTING

In Fig. 2, we show the Molchan trajectories for all target shocks (35) with $M_w \geq 5.5$ (not-declustered) preceded by strong shocks with $4.4 \leq M_w < 4.8$, by varying Δt from a width of a few seconds to the total duration $T = 60$ yr of the catalogue. Red and dark blue lines refer to the unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms, respectively (see in Table S2 in the Supporting Information the numerical values of plotted curves).

The adopted foreshock M_w range ($M_w = 4.6 \pm 0.2$) was chosen after a comparative analysis of the relative performance of various ranges with lower and upper magnitude bounds varying from the completeness threshold of the catalogue ($M_w = 4.0$) to the minimum magnitude of target shocks ($M_w = 5.0$). Such analysis was aimed at maximizing the overall AS score and at the same time minimizing the total number of alarms (Fig. 4).

Both the red and dark blue lines in Fig. 1 lie well below the $\alpha = 1$ per cent confidence curve (green) for all explored Δt. All the target shocks are successfully forecasted ($\nu = 0$) for $\Delta t = 20$ yr (corresponding to $\tau_u = 32$ per cent and $\tau_w = 51$ per cent) or larger. For $\Delta t = 1$ yr, about 83 per cent of target shocks (29) are successfully forecasted, with space–time coverages $\tau_u = 3.3$ per cent and $\tau_w = 6.3$ per cent. 40 per cent of target shocks (14) are forecasted with $\Delta t = 1$ d for which $\tau_u = 0.01$ per cent and $\tau_w = 0.03$ per cent. The AS diagram in Fig. 3 (see Table S2 in the Supporting Information for numerical values) confirms such good performance with the scores of the forecasting method (red and dark blue lines) well above the random expectation (black) and the 1 per cent confidence line (green) for any Δt. The overall AS scores $a_f(\tau_u = 1) = 0.96 \pm 0.05$ and $a_f(\tau_w = 1) = 0.94 \pm 0.05$, based on the Student's t-test, are significantly larger than the expectance of a random method (0.5) with significance level (s.l.) $\ll$0.01.

As noted above the aftershocks produced by the first target shocks of seismic sequences may significantly contribute to forecast subsequent target shocks with $M_w \geq 5.5$ within the same sequence. We then proceed to analyse in the same way the declustered set of target shocks with $M_w \geq 5.5$ obtained by discarding all target shocks occurred within a spatial distance $R = 50$ km and a time window of a year after the first and all subsequent $M_w \geq 5.5$ shocks of the sequence. This reduces the number of considered target shocks with $M_w \geq 5.5$ from 35 to 14.

In Figs 5 and 6, we report the same plots as in Figs 2 and 3 but for the (declustered) set of only the first target shocks with $M_w \geq 5.5$ of each sequence (see Table S3 in the Supporting Information for numerical values). The performance is worse than for the not-declustered set but remains well below the random line and the $\alpha = 1$ per cent confidence curve in the Molchan diagram of Fig. 5 and also well above the $\alpha = 1$ per cent confidence line of AS diagram of Fig. 6. Even in this case all 14 target shocks are successfully forecasted with $\Delta t = 20$ yr or larger. For $\Delta t = 1$ yr, 64 per cent of target shocks (9) are forecasted and 29 per cent (4) for $\Delta t = 1$ d. The overall AS score $a_f(\tau_u = 1) = 0.93 \pm 0.08$ and $a_f(\tau_w = 1) = 0.87 \pm 0.08$ are lower than for the not-declustered set but anyhow they are significantly larger than the expectance (0.5) of a random method with s.l. $\ll$0.01.

In Figs S3–S6 of Supporting Information, we report the same plots of Figs 2, 3, 5 and 6 for target shocks with $M_w \geq 5.0$ (numerical values in Tables S4 and S5, Supporting Information). The performance is definitely worse than for $M_w \geq 5.5$, but still better than the 1 per cent confidence limit. In particular, even for $\Delta t = 60$ yr, only 89 over 98 (91 per cent) target shocks for the not-declustered set and only 36 over 44 (82 per cent) for the declustered set are successfully forecasted. The reason is that even when Δt is equal to the total duration of the catalogue, in some CAs there remains a fraction of time (before the first strong shock) without any strong shock and then without any alarm. Actually, the maximum fraction of space–time occupied by alarms (τ_u) is only about 44 per cent of the total space–time and nine target shocks with $M_w \geq 5.0$ occurred in the remaining 56 per cent. Here, the last part of the Molchan trajectories, consisting of a linear decrease from the last point defined by the algorithm ($\tau_u = 44$ per cent and $\tau_w = 62$ per cent with $\nu = 9$ per cent for not-declustered and 18 per cent for declustered) to the lower left corner ($\tau = 100$ per cent, $\nu = 0$), can be interpreted as the application to the remaining earthquakes, not predicted by any foreshock, of a purely random forecasting method with success rate proportional to the fraction of the remaining space–time region not covered by our forecasting algorithm.

The overall AS scores are $a_f(\tau_u = 1) = 0.89 \pm 0.03$ and $a_f(\tau_w = 1) = 0.85 \pm 0.03$ for the not-declustered set and $a_f(\tau_u = 1) = 0.78 \pm 0.04$ and $a_f(\tau_w = 1) = 0.70 \pm 0.04$ for the declustered set. In all cases they are significantly larger than the expectance (0.5) of a random method with s.l. $\ll$0.01.

In Figs S7–S10 of the Supporting Information, we also report the plots for targets with $M_w \geq 6.0$ (see numerical values in Tables S6 and S7, Supporting Information). The performance is similar to that for $M_w \geq 5.5$ but as the number of target events is smaller (10 not-declustered and 7 declustered), the power of the tests and the reliability of possible inferences are relatively poorer. This is actually reflected by the fact that the confidence limits in this case are relatively close to the Molchan and AS trajectories.

All not-declustered targets are successfully forecasted with $\Delta t = 20$ yr, 80 per cent with $\Delta t =$1 yr and 50 per cent with $\Delta t =$1 d. For declustered targets, the corresponding forecasting rates are 100 per cent, 71 per cent and 43 per cent respectively. The overall AS scores are $a_f(\tau_u = 1) = 0.95 \pm 0.09$ and $a_f(\tau_w = 1) = 0.91 \pm 0.09$ for not-declustered and $a_f(\tau_u = 1) = 0.93 \pm 0.11$ and $a_f(\tau_w = 1) = 0.87 \pm 0.11$ for declustered. In all cases, they are significantly larger than the expectance (0.5) of a random method with s.l.$\ll$0.01.

One question that may come to mind when looking at the results of such space–time analysis is how much of the observed forecasting performance is due to spatial clustering and how much to time clustering. In order to try to answer such question, we made some further computations in which the time clustering is eliminated by

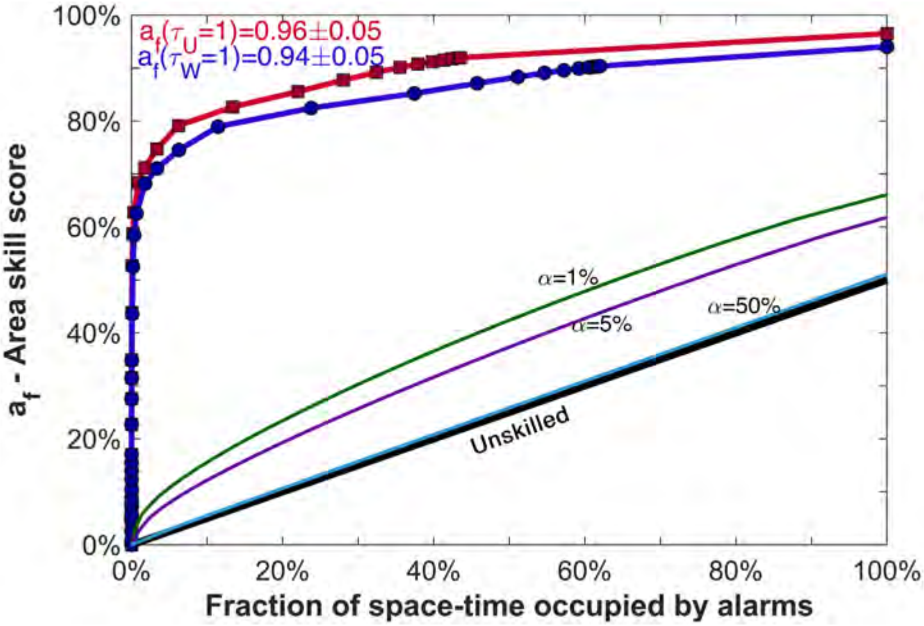


Figure 3. AS score diagram for all target shocks with $M_w \geq 5.5$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms, respectively (see the text). The black continuous line indicates the performance of a purely random forecasting method that separates skilled (above the line) from unskilled (below) forecasting methods. The light blue, violet and green lines indicate the confidence limits for $\alpha = 50$ per cent, 5 per cent and 1 per cent, respectively.

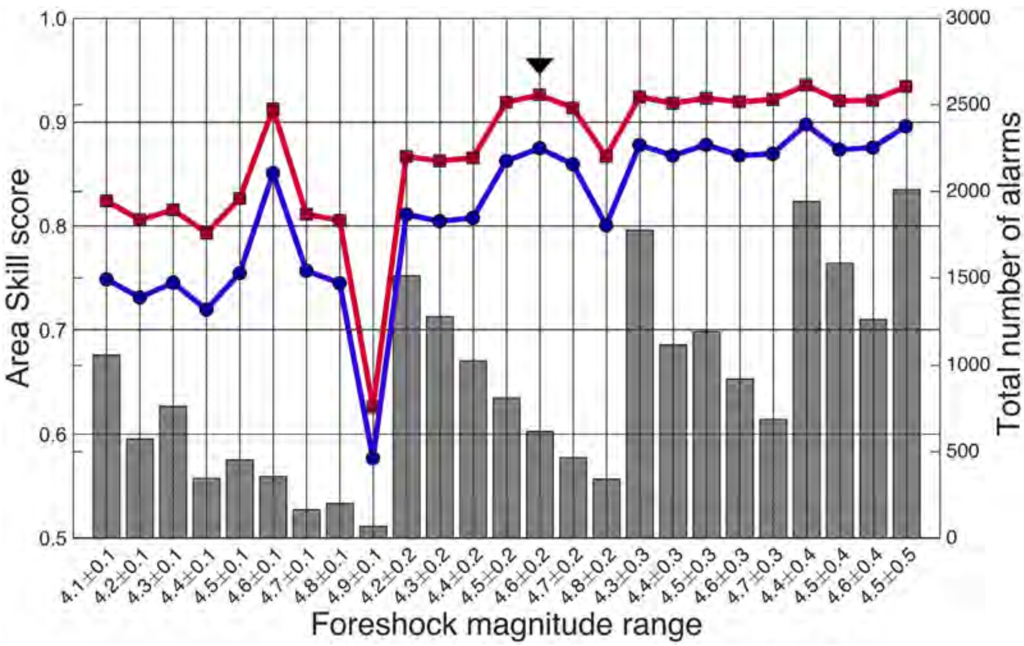


Figure 4. AS score computed for declustered targets with $M_w \geq 5.5$, using unweighted (red line) and weighted (blue) fractions of space–time occupied by alarms, and total number of alarms (grey bars) as a function of the foreshock magnitude range. The arrows indicate the range $M_w = 4.6 \pm 0.2$, chosen as best compromise between high AS score and low number of alarms.

assuming in each CA a permanent alarm for the entire duration of the forecasting experiment ($T = 60$ yr). We computed the time-independent Molchan and AS score trajectories by adding step by step one CA at a time, starting from the CA with highest weight (highest long-term seismic activity) and then going on, up to add all CAs. At each step, the unweighted and weighted fractions of space occupied by alarms are computed by simply taking $\tau_c = 1$ in eqs (4) and (6), respectively, for the included CAs and $\tau_c = 0$ for the not included CAs.

The results of such time-independent analysis for declustered (first) target shocks with $M_w \geq 5.5$ is shown in Figs 7 and 8. Even if they are not fully comparable with the time-dependent analysis of Figs 5 and 6 because the trajectories depend on the adopted ordering of the CAs, from the most to the least active, we can note

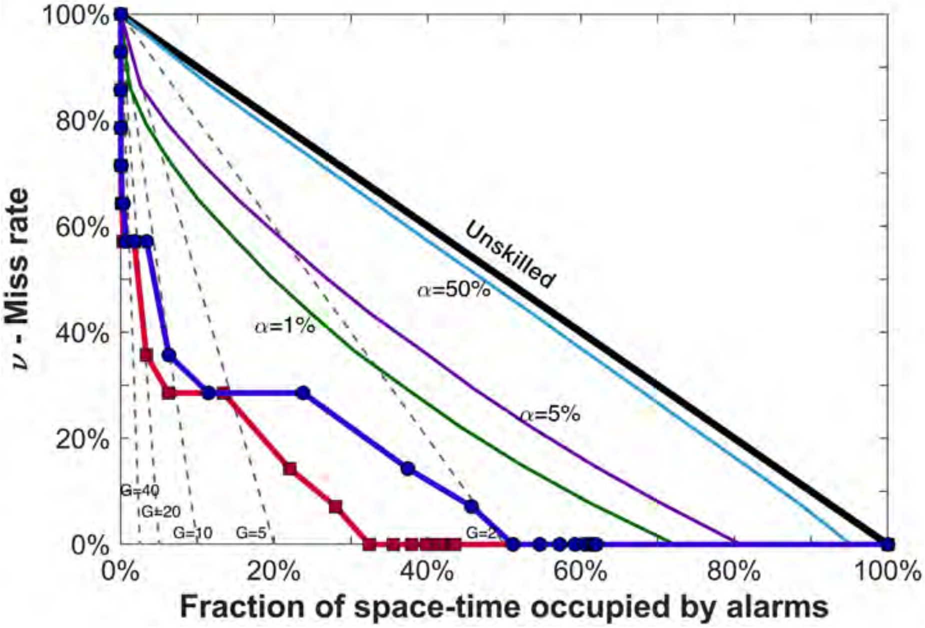


Figure 5. Same as Fig. 2 for declustered (first) target shocks with $M_w \geq 5.5$.

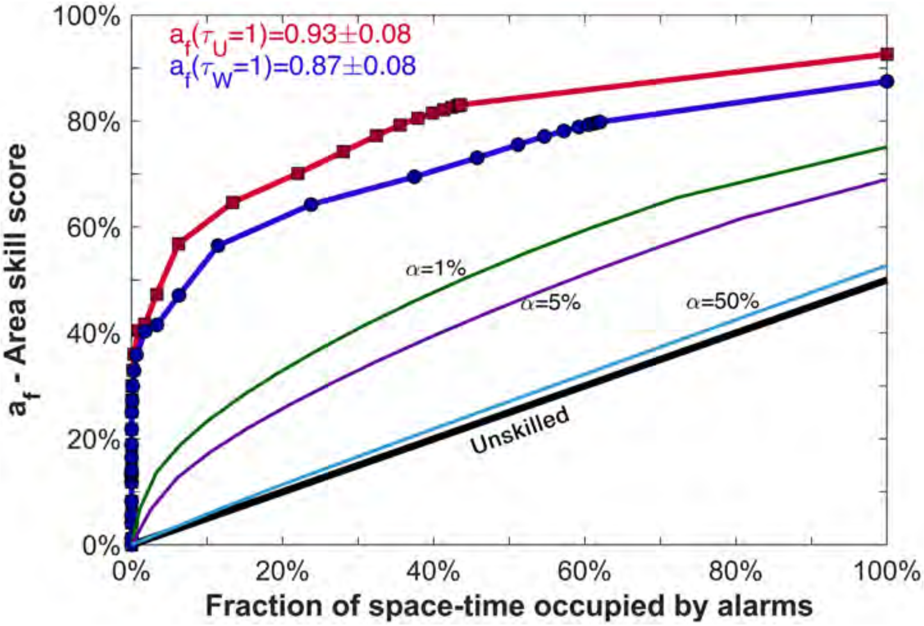


Figure 6. Same as Fig. 3 for declustered (first) target shocks with $M_w \geq 5.5$.

that the skill of time-independent analysis appears definitely lower, particularly at small τ and for the weighted trajectories (blue lines). This can be easily explained by the higher time clustering at short times (and then at small τ) and by the fact that the weights based on the long-term seismic activity penalize more the CAs where the target shocks actually occurred in the last 60 yr.

The results for declustered (first) target shocks with $M_w \geq 5.0$ and ≥ 6.0 are reported in Figs S11–S14 of Supporting Information. For $M_w \geq 5.0$, the comparison of Figs S11 and S12 in the Supporting Information with the time-dependent analysis of Figs S5 and S6 in the Supporting Information is similar to the case for $M_w \geq 5.5$ described before. For $M_w \geq 6.0$, the comparison of Figs S13 and S14 in the Supporting Information with the time-dependent analysis of Figs S9 and S10 in the Supporting Information, apart for small τ, apparently indicates an overall higher skill for the time-independent analysis with respect to the time-dependent one. This is due to the fact that for $M_w \geq 6.0$ all declustered target shocks occurred in CAs with very high long-term seismic activity and that, as noted above, time-independent and time-dependent statistics are not fully comparable between them.

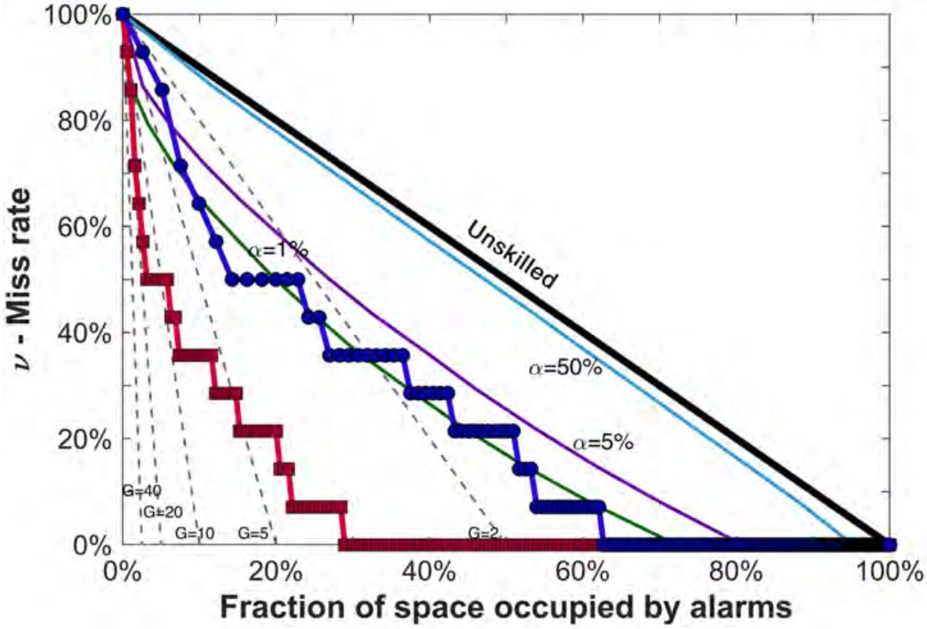


Figure 7. Same as Fig. 2 for time-independent analysis of declustered (first) target shocks with $M_{\mathrm{w}} \geq 5.5$.

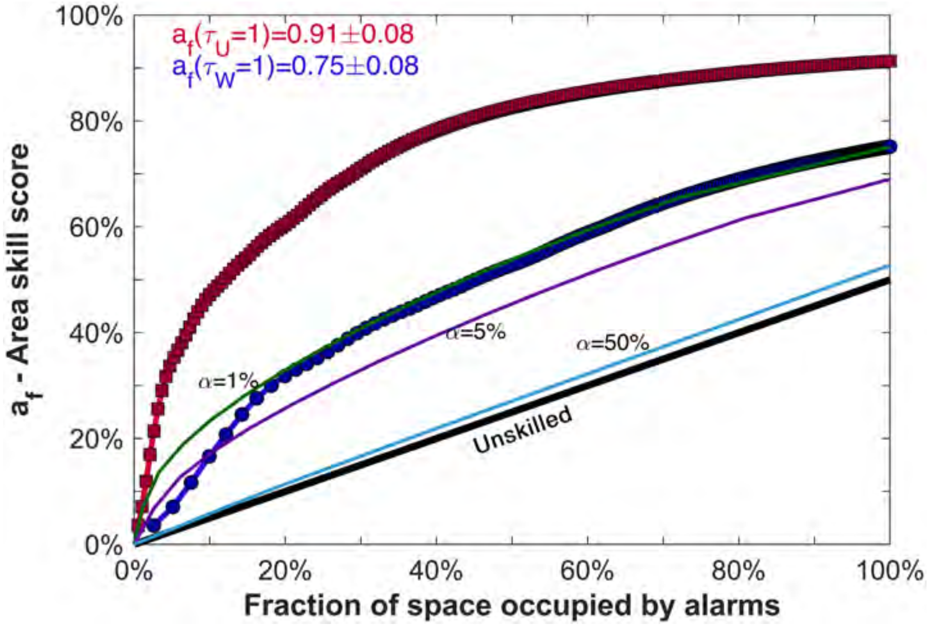


Figure 8. Same as Fig. 3 for time-independent analysis of declustered (first) target shocks with $M_{\mathrm{w}} \geq 5.5$.

OPTIMIZATION OF THE FORECASTING ALGORITHM

For a practical application of the forecasting method, it might be useful to determine the values of the algorithm parameter Δt for which the forecasting method is more efficient and useful for risk mitigation. To accomplish this purpose, we analyse the behaviour of some statistics that depend on the alarm time window Δt.

In Fig. 9 we report, for declustered targets and weighted fraction of space–time occupied by alarms (τ_w), the binomial probability (eq. 9), that is the probability that the observed number of successful forecasts is obtained by chance, as a function of Δt. The lower the probability the higher the strength of the forecast. In general, probabilities are relatively low within a wide range going from one day to some years. For $M_{\mathrm{w}} \geq 5.0$ (red line), very low probabilities are observed around $\Delta t = 2 \div 10$ d. For $M_{\mathrm{w}} \geq 5.5$ (blue line) and $M_{\mathrm{w}} \geq 6.0$ (green line) the minimum probabilities are larger than the ones for $M_{\mathrm{w}} \geq 5.0$, and they remain relatively low from a few hours to a few months. Within such ranges, the forecasting ability of our method reaches its higher efficiency.

The behaviour of the probability gain G (eq. 8) as a function Δt (Fig. 10) shows, for all the three magnitude thresholds, monotonically descending trends from more than 100 000 at very short Δt (less than a minute) to slightly more than 1 at very long Δt (tens

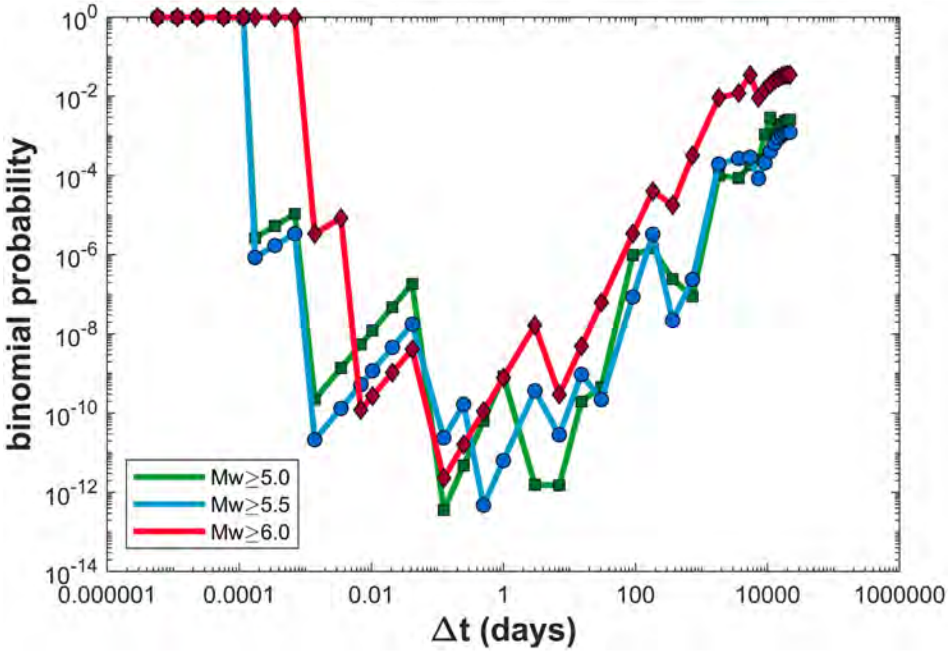


Figure 9. Binomial probability density for declustered (first) target shocks and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

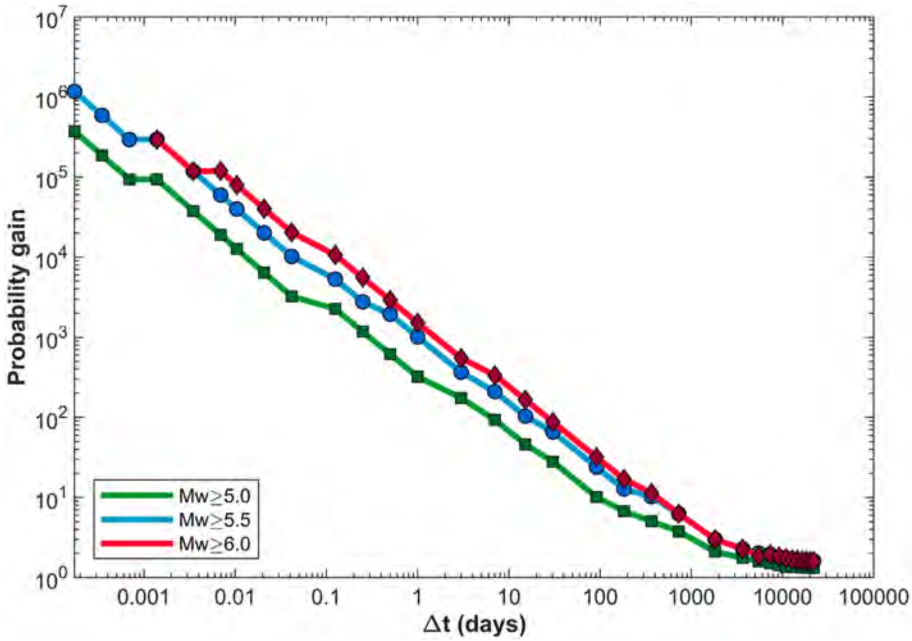


Figure 10. Probability gain for declustered (first) target shocks and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

of years). Such curves also show relatively milder slopes in correspondence of steep decreases of binomial probabilities in Fig. 9 (i.e. around 0.001 d and a few days).

In Fig. 11, we show the miss rate ν as a function of Δt. In general, it decreases with increasing Δt. The (negative) trends—with respect to $\log_{10}\Delta t$—are in between the -5 per cent and -10 per cent per decade, for Δt ranging from a few seconds to about 1 yr. Then they start to decrease more rapidly (about -20 per cent per decade) reaching 0 for $M_w \geq 5.5$ and ≥ 6.0 and 19 per cent for $M_w \geq 5.0$ at very large Δt.

The behaviour of the same statistic for the full set of target events (not-declustered) is reported in Figs S15–S17 of the Supporting

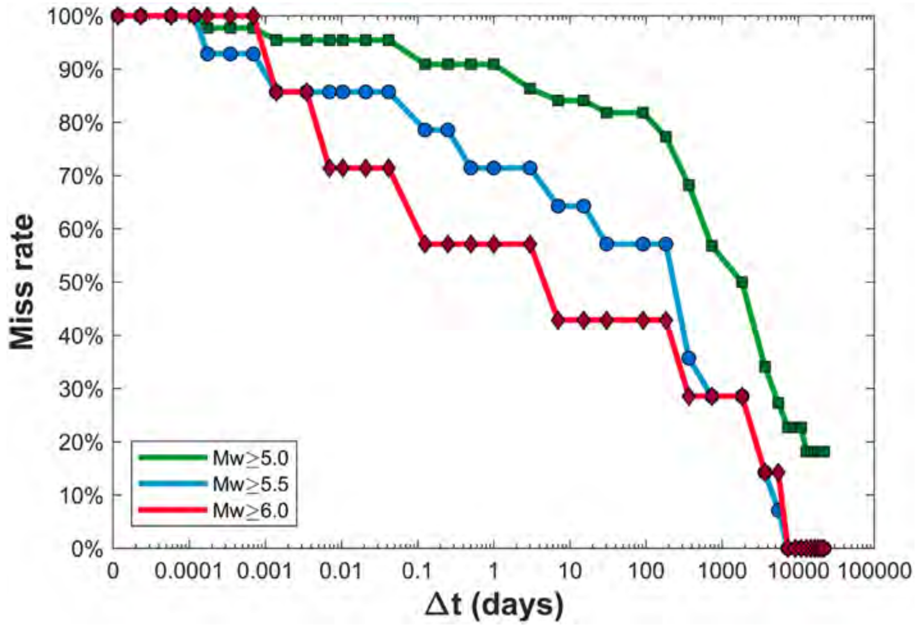


Figure 11. Miss rate for declustered (first) target shocks and different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Information. It is similar to those of the declustered set but the binomial probabilities are lower, the probability gains are higher and the miss rates decrease more rapidly with Δt.

Another aspect to be considered for the practical application of the forecasting method is the dependence on Δt of the fractions of space–time occupied by alarms τ_u and τ_w (Fig. 12). A long alarm interval Δt (with a corresponding long fraction of space–time occupied by alarms τ) allows to forecast more target earthquakes but at the same time it has relatively lower probabilities of occurrence than a shorter Δt. Furthermore, a longer duration of alarms would impact more with life activities of the population in the involved area. Even if any decision on the possible practical application in real situations would eventually require a careful evaluation by decision makers even considering a cost-benefits analysis (e.g. van Stiphout *et al.* 2010; Hermann *et al.* 2016), we examine here as an example the choice of $\Delta t = 3$ months (0.25 yr). This choice, in most cases, results in a fairly trade-off between a good efficiency and a narrow space–time fraction covered by alarms $\tau \approx 2$.

We can see in Table 2 that in this case the method is able to retrospectively forecast more than 50 per cent of not-declustered target shocks with $M_w \geq 5.0$ and more than 70 per cent of those with $M_w \geq 5.5$ and ≥ 6.0. We also report in Table 2 the statistic of the numbers of successful alarms with respect to the total number of alarms indicating higher rates for target with $M_w \geq 5.0$. About one-fifth of alarms actually forecast an earthquake, while the fraction of successful alarms definitely decreases for larger targets and further decreases for declustered sets down to about 1 per cent. Note that several alarm time windows are actually overlapped and then the total duration of alarms is shorter than the simple sum of alarm windows (eq. 2).

The performance of the method is definitely worse for the first target shocks (declustered set) but it improves by increasing the magnitude of target shocks. Actually, 4 over 7 first target shocks with $M_w \geq 6.0$ over the last 60 yr in Italy are retrospectively forecasted in this way.

We tested the stability with time of the forecasting performance by subdividing the seismic catalog in two equal parts of 30 yr: before and after 1990 january 01. The same computations of Table 2 for $\Delta t = 3$ months for intervals 1960–1989 and 1990–2019 are reported in Tables 3 and 4 respectively. The rates of successfully forecasted target shocks (declustered or not) are similar in the two periods whereas the space–time fraction occupied by alarms is definitely lower in the most recent period, consistently with the higher ratios between successful and total alarms. We could argue that smaller magnitude errors in most recent times, owing to the continuous improvement of the Italian seismic network, reduce the amount of false alarms and then increase the observed skill of the forecasting method with respect to the previous period.

In Tables 5 and 6, we report the lists of retrospective forecast of the first (declustered) target shocks with $M_w \geq 5.5$ and ≥ 6.0, respectively, occurred in Italy from 1960 to 2019 (also see the results for the declustered first shocks with $M_w \geq 5.0$ in Table S8 in the Supporting Information and the results for not-declustered targets with $M_w \geq 5.0$, 5.5 and 6.0 in Tables S9–S11, respectively, of the Supporting Information).

We can note that for two target shocks (1976 Friuli and 1990 Potentino) the forecast could have hardly been used by civil protection services to adopt safety countermeasures because the forecasting strong shocks occurred too shortly before the main shock (67 and 13 s, respectively). In other cases, the time delay between the forecasting shock and the main shock (going from a couple of hours to a few weeks) would have been sufficient to take some countermeasures.

We could note that a foreshock did actually occur a couple of days before the first main shock of 2012 May 20 ($M_w = 6.1$) in the area of Pianura Emiliana but its magnitude ($M_w = 4.2$) was only slightly below the lower threshold of $M_w = 4.4$ we adopted. The retrospective ability to predict $M_w \geq 6.0$ earthquakes might have been improved then by slightly reducing such lower threshold but at a cost of a general reduction of the performance of the algorithm, because of the increment of the number of alarms and of the fraction of space–time covered by alarms.

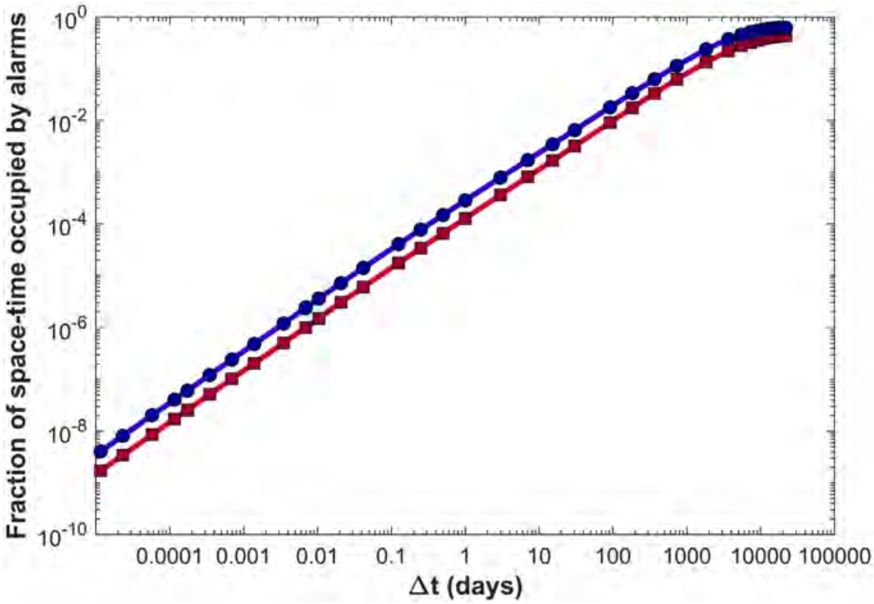


Figure 12. Unweighted (red) and weighted (dark blue) fraction of space–time occupied by alarms as a function of the alarm duration Δt.

Table 2. Retrospective forecasting performance of the algorithm for $\Delta t = 3$ months.

Target magnitude	≥5.0		≥5.5		≥6.0		τ_u (per cent)	τ_w (per cent)
			Not-declustered					
Forecasted/total shocks	55/98	56 per cent	26/35	74 per cent	7/10	70 per cent	0.9	1.9
Successful/total alarms	115/617	18.6 per cent	72/617	11.7 per cent	30/617	4.9 per cent	0.9	1.9
			Declustered					
Forecasted/total shocks	8/44	18 per cent	6/14	43 per cent	4/7	57 per cent	0.9	1.9
Successful/total alarms	13/617	2.1 per cent	9/617	1.5 per cent	8/617	1.3 per cent	0.9	1.9

Table 3. Same as Table 2 for the time interval 1960–1989.

Target magnitude	≥5.0		≥5.5		≥6.0		τ_u (per cent)	τ_w (per cent)
			Not-declustered					
Forecasted/total shocks	21/45	47 per cent	11/15	73 per cent	3/4	75 per cent	1.0	2.1
Successful/total alarms	45/336	12.9 per cent	22/336	6.6 per cent	9/336	2.7 per cent	1.0	2.1
			Declustered					
Forecasted/total shocks	3/25	12 per cent	3/7	43 per cent	2/3	67 per cent	1.0	2.1
Successful/total alarms	5/336	1.5 per cent	5/336	1.5 per cent	3/336	0.89 per cent	1.0	2.1

CONCLUSIONS

We analysed a simple algorithm to forecast shallow (depth < 50 km) main shocks ($M_\mathrm{w} \geq 5.0$, 5.5 and 6.0) that threaten the life and the goods of the population living on the Italian mainland territory, based on the previous occurrence within CA of 30 km of radius of widely felt strong shocks ($4.4 \leq M_\mathrm{w} < 4.8$) not particularly harmful in themselves. Based on a retrospective analysis of the HORUS seismic catalogue of Italy from 1960 to 2019 (Lolli *et al.* 2020) this method retrospectively forecast the majority of damaging earthquakes occurred in Italy in the past 60 yr by issuing alarms covering only a small fraction of the space–time coverage.

We estimated such fraction even considering the different levels of seismic activity in different areas of Italy by weighting more the alarm times in CA where the average seismicity rate, computed from the CPTI15 seismic catalogue (Rovida *et al.* 2016, 2020) from 1600 to 1959, is higher.

The retrospective testing using the Molchan diagram (Molchan 1990, 1991; Molchan & Kagan 1992) and the AS score (Zechar &

Table 4. Same as Table 2 for the time interval 1990–2019.

Target magnitude	$\geq$5.0		$\geq$5.5		$\geq$6.0		τ_u (per cent)	τ_w (per cent)
				Not-declustered				
Forecasted/total shocks	34/53	64 per cent	15/20	75 per cent	4/6	67 per cent	0.4	0.7
Successful/total alarms	70/281	24.9 per cent	50/281	17.8 per cent	21/281	7.5 per cent	0.4	0.7
				Declustered				
Forecasted/total shocks	5/19	26 per cent	3/7	43 per cent	2/4	50 per cent	0.4	0.7
Successful/total alarms	8/281	3.5 per cent	4/281	1.4 per cent	5/281	1.8 per cent	0.4	0.7

Table 5. Results of retrospective forecast of first main shocks (declustered targets) with $M_w \geq 5.5$ in Italy from 1960 to 2019, using $\Delta t = 3$ months (0.25 yr).

Year	Month	Day	Lat	Lon	M_w	t_a (d)		Epicentral area
1962	8	21	41.233	14.933	5.7	0.093	2.22 h	Irpinia
1968	1	15	37.700	13.100	5.7	0.425	10.2 h	Valle del Belice
1976	5	6	46.250	13.250	6.5	7.8×10^{-4}	67 s	Friuli
1979	*9*	*19*	*42.717*	*12.950*	*5.8*	*Missed*		*Valnerina*
1980	*11*	*23*	*40.800*	*15.367*	*6.8*	*Missed*		*Irpinia-Basilicata*
1984	*4*	*29*	*43.204*	*12.585*	*5.6*	*Missed*		*Umbria settentrionale*
1984	*5*	*7*	*41.666*	*13.820*	*5.9*	*Missed*		*Monti della Meta*
1990	5	5	40.650	15.882	5.8	1.5×10^{-4}	13 s	Potentino
1997	9	26	43.023	12.891	5.7	22.1		Appennino umbro-marchigiano
1998	*9*	*9*	*40.060*	*15.949*	*5.5*	*Missed*		*Appennino lucano*
2002	*10*	*31*	*41.717*	*14.893*	*5.7*	*Missed*		*Molise*
2009	4	6	42.342	13.380	6.3	6.5		Aquilano
2012	*5*	*20*	*44.896*	*11.264*	*6.1*	*Missed*		*Pianura Emiliana*
2016	*8*	*24*	*42.698*	*13.234*	*6.2*	*Missed*		*Monti della Laga*

Notes: t_a is the maximum time advance of the foreshock with respect to the main shock. '*Missed*' indicates that the target shock was not forecasted (in such cases all entries are in italics). Epicentral area identifiers are taken from the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

Table 6. Same as Table 2 for first main shocks with $M_w \geq 6.0$.

Year	Month	Day	Lat	Lon	M_w	t_a (d)		Epicentral area
1962	8	21	41.233	14.933	6.2	0.100	2.40 h	Irpinia
1976	5	6	46.250	13.250	6.5	7.8×10^{-4}	67 s	Friuli
1980	*11*	*23*	*40.800*	*15.367*	*6.8*	*Missed*		*Irpinia-Basilicata*
1997	9	26	43.015	12.854	6.0	22.5		Appennino umbro-marchigiano
2009	4	6	42.342	13.380	6.3	6.5		Aquilano
2012	*5*	*20*	*44.896*	*11.264*	*6.1*	*Missed*		*Pianura Emiliana*
2016	*8*	*24*	*42.698*	*13.234*	*6.2*	*Missed*		*Monti della Laga*

Jordan 2008) methods indicates that such approach clearly overperforms a purely random method with high or very high confidence, depending on the target shock magnitude threshold.

As the secondary main shocks during seismic sequences are definitely easier to be forecasted by this method because the aftershocks of the first main shock usually generate alarms at weakly (if not daily) rate, we also tested the ability of our approach to predict only the first main shock of each sequence. We found that the forecasting ability remains high even if being lower than that considering all main shocks.

Even if the true verification of the efficiency of the method will only be made on a prospective data set, we believe that such simple forecasting algorithm could be useful, like other operational forecasting approaches presently considered by the Italian Civil Protection Department, for planning preparation measures in the field (e.g. Marzocchi *et al.* 2014).

The latter approaches are mainly based on the ETAS model (Kagan & Knopoff 1987; Ogata 1988) and, as well as that of this work, showed to retrospectively forecast the evolution of Italian seismicity better than an inhomogeneous random process with spatial rates corresponding to past seismicity. On the other hand, Marzocchi & Zhuang (2011) showed that ETAS models is able to describe quite well even the observed foreshock activity. However, a comparison of the relative efficiency of our approach with ETAS models and even with other forecasting approaches (like e.g. the EEPAS method (Rhoades & Evison 2004) would require that the probabilistic formulation of the latter methods is adapted to the alarm-based one (e.g. by selecting a particular probability thresholds above which to declare an alarm). However, such adaptation is not trivial and hence, the question on which of the different approaches is better in predicting future damaging earthquakes remains not answered presently and has to be deferred to future papers comparing all methods in an alarm-based context by using, for example, the approach proposed by Shebalin *et al.* (2014).

One advantage of the present forecasting approach is that it is easy to implement and communicate because it does not require any other scientific analysis than the correct determination of the location and of the magnitude of the precursory shock. In principle every person could be informed very quickly by a notification sent by one of the already available mobile Apps which provide near real-time access to the INGV online earthquake list (http://terremoti.ingv.it/en#).

ACKNOWLEDGEMENTS

We thank the editor Margarita Segou, Jancang Zhuang, Rodolfo Console and two anonymous reviewers for thoughtful comments and suggestions that helped to improve the paper. This paper benefitted from funding provided by the European Union within the ambit of the H2020 project RISE (no. 821115), which in particular fully financed the PhD grant of one of the authors (EB).

REFERENCES

Agnew, D.C. & Jones, L.M., 1991. Prediction probabilities from foreshocks, *J. geophys. Res.,* **96,** 11959–11971.

Caputo, M., Gasperini, P., Keilis-Borok, V., Marcelli, L. & Rotwain, I., 1977. Earthquakes swarms as forerunners of strong earthquakes in Italy, *Ann. Geofis.,* **30,** 269–283.

Caputo, M., Console, R., Gabrielov, A.M., Keilis-Borok, V.I. & Sidorenko, T.V., 1983. Long-term premonitory seismicity patterns in Italy, *Geophys. J. R. astr. Soc.,* **75,** 71–75.

Console, R., Murru, M. & Alessandrini, B., 1993. Foreshock statistics and their possible relationship to earthquake prediction in the Italian region, *Bull. seism. Soc. Am.,* **83**(4), 1248–1263.

Console, R. & Murru, M., 1996. Probability gain due to foreshock following quiescence tested by synthetic catalogs, *Bull. seism. Soc. Am.,* **86**(3), 911–913.

Console, R. & Murru, M., 2001. A simple and testable model for earthquake clustering, *J. geophys. Res.,* **106,** 8699–8711.

Console, R., Di Luccio, F., Murru, M., Imoto, M. & Stavrakakis, G., 1999. Short term and short range seismicity patterns in different seismic areas of the world, *Nat. Hazard,* **19,** 107–121.

Console, R, Murru, M & Lombardi, A.M., 2003. Refining earthquake clustering models, *J. geophys. Res.,* **108,** 2468, doi:10.1029/2002JB002130.

Console, R, Murru, M & Catalli, F., 2006. Physical and stochastic models of earthquake clustering, *Tectonophysics,* **417,** 141–153.

Console, R., Murru , M. & Falcone, G., 2010. Retrospective Forecasting of M $\geq$ 4.0 Earthquakes in New Zealand, *Pure appl. Geophys.,* **167,** 693–707.

Di Luccio, F., Console, R., Imoto, M. & Murru, M., 1997. Analysis of short time-space range seismicity patterns in Italy, *Ann. Geofis.,* **40/4,** 783–798.

Gardner, J.K. & Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. seism. Soc. Am.,* **64,** 1363–1367.

Gasperini, P., Lolli, B., Vannucci, G. & Boschi, E., 2012. A comparison of moment magnitude estimates for the European–Mediterranean and Italian region, *Geophys. J. Int.,* **190,** 1733–1745.

Gasperini, P., Lolli, B. & Vannucci, G., 2013. Empirical calibration of local magnitude data sets versus moment magnitude in Italy, *Bull. seism. Soc. Am.,* **103,** 2227–2246.

Gasperini, P., Lolli, B. & Vannucci, G., 2016. Relative frequencies of seismic main shocks after strong shocks in Italy, *Geophys. J. Int.,* **207,** 150–159.

Geller, R., Jackson, D.D., Kagan, Y.Y. & Mulargia, F., 1997. Earthquakes cannot be predicted, *Science,* **275,** 1661, doi:10.1126/science.275.5306.1616.

Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34,** 185–188.

Grandori, G., Guagenti, E. & Perotti, F., 1988. Alarm systems based on a pair of short-term earthquake precursors, *Bull. seism. Soc. Am.,* **78**(4), 1538–1549.

Hermann, M., Zechar, J.D. & Wiemer, S., 2016. Communicating time-varying seismic risk during an earthquake sequence, *Seismol. Res. Lett.,* **87**(2A), 30–312.

Jones, L.M., 1984. Foreshocks (1966–1980) in the San Andreas system, California, *Bull. seism. Soc. Am.,* **74,** 1361–1380.

Jones, L.M., 1985. Foreshocks and time-dependent earthquake hazard assessment in southern California, *Bull. seism. Soc. Am.,* **75,** 1669–1680.

Jones, L.M., 1994. Foreshocks, aftershocks, and earthquake probabilities accounting for the Landers earthquake, *Bull. seism. Soc. Am.,* **84**(1), 892–899.

Jones, L. & Molnar, P., 1976. Frequency of fore- shocks, *Nature,* **262,** 677–679, 1976.

Jones, L. & Molnar, P., 1979. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults, *J. geophys. Res.,* **84**(B7), 3596–3608.

Jordan, T.H., 2006. Earthquake predictability, brick by brick, *Seismol. Res. Lett.,* **77**(1), 3–6.

Jordan, T.H. *et al.*, 2011. Operational earthquake forecasting: state of knowledge and guidelines for implementation, final report of the international commission on earthquake forecasting for civil protection, *Ann. Geophys.,* **54**(4), 315–391.

Jordan, T.H. & Jones, L.M., 2010. Operational earthquake forecasting: some thoughts in why and how, *Seismol. Res. Lett.,* **81/4,** 571–574.

Kagan, Y.Y., 2014. *Earthquakes,* AGU-Wiley, 283 pp.

Kagan, Y. & Knopoff, L., 1987. Statistical short-term earthquake prediction, *Science,* **236,** 1563–1567.

Lolli, B., Gasperini, P. & Rebez, A., 2018. Homogenization in terms of Mw of local magnitudes of Italian earthquakes occurred before 1981, *Bull. seism. Soc. Am.,* **108**(1), 481–492.

Lolli, B., Randazzo, D., Vannucci, G. & Gasperini, P., 2020, HOmogenized instRUmental Seismic catalog (HORUS) of Italy from 1960 to present, *Seismol. Res. Lett.,* **91,** 3208–3222.

Marzocchi, W., Lombardi, A.M. & Casarotti, E., 2014. The Establishment of an Operational Earthquake Forecasting System in Italy, *Seismol. Res. Lett.,* **85/5,** 961–969.

Marzocchi, W. & Zhuang, J., 2011. Statistics between mainshocks and foreshocks in Italy and Southern California, *Geophys. Res. Lett.,* **38,** L09310.

Molchan, G.M., 1990. Strategies in strong earthquake prediction, *Phys. Earth planet. Inter.,* **61,** 84–98.

Molchan, G.M., 1991. Structure of optimal strategies in earth- quake prediction, *Tectonophysics*. **193,** 267–276.

Molchan, G.M. & Kagan, Y.Y., 1992, Earthquake predictionand its optimization, *J. geophys. Res.,* **97,** 4823–4838.

Mulargia, F. & Geller, R.J., 2003. *Earthquake Science and Seismic Risk Reduction,* Kluwer, 338pp.

Murru, M., Console, R. & Falcone, G., 2009. Real time earthquake forecasting in Italy, *Tectonophysics,* **470,** 214–223.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.,* **83,** 9–27.

Reasenberg, P.A., 1985. Second-order moment of central California seismicity, 1969–1982, *J. geophys. Res.,* **90,** 5479–5495.

Reasenberg, P.A., 1999a. Foreshock occurrence before large earthquakes, *J. geophys. Res.,* **104**(B3), 4755–4768.

Reasenberg, P.A., 1999b. Foreshock occurrence rates before large earthquakes worldwide, *Pure appl., Geophys,* **155,** 355–379.

Rhoades, D.A. & Evison, F.F., 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale, *Pure appl. Geophys.,* **161,** 47–72.

Rovida, A., Locati, M., Camassi, R., Lolli, B. & Gasperini, P., 2016. CPTI15. The 2015 version of the Parametric Catalogue of Italian Earthquakes, *Istituto Nazionale di Geofisica e Vulcanologia*.

Rovida, A., Locati, M., Camassi, R., Lolli, B. & Gasperini, P., 2020. The Italian earthquake catalogue CPTI15, *Bull. Earth. Eng.,* **18,** 2953–2984.

Schorlemmer, D., Christophersen, A., Rovida, A., Mele, F., Stucchi, M. & Marzocchi, W., 2010. Setting up an earthquake forecast experiment in Italy, *Ann. Geophys.,* **53/3,** 1–9.

Southern San Andreas Working Group, 1991. Short-Term Earthquake Hazard Assessment for the San Andreas Fault in Southern California, U.S. Geol. Surv., Open-file Report 91-31.

Shebalin, P., Narteau, C., Holschneider, M. & Schorlemmer, D., 2011. Short-Term earthquake forecasting using early aftershock statistics, *Bull. seism. Soc. Am.,* **101,** 297–312.

Shebalin, P., , Narteau, C., Zechar, J. D. & Holschneider, M., 2014. Combining earthquake forecasts using differential probability gains, *Earth Planets Space,* **66,** 37.

van Stiphout, T., Wiemer, S. & Marzocchi, W., 2010. Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake, *Geophys. Res. Lett.,* **37**(6), 1–5.

Zechar, J.D. & Jordan, T.H., 2008. Testing alarm-based earthquake predictions, *Geophys. J. Int.,* **172,** 715–724.

Zechar, J.D. & Jordan, T.H., 2010. The area skill score statistic for evaluating earthquake predictability experiments, *Pure appl. Geophys.,* **167,** 893–906.

Zechar, J.D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J. & Jordan, T.H., 2010. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurr. Comput.,* **22,** 1836–1847.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Spatial distribution of inland earthquakes from the HORUS catalogue (Lolli *et al.* 2020) with $M_w \geq 4.0$ and depth < 50 km used for testing and optimization. Black dots indicate $4.0 \leq M_w < 5.0$, green dots $5.0 \leq M_w < 5.5$, blue dots $5.5 \leq M_w < 6.0$ and red dots $M_w \geq 6.0$.

Figure S2. Time distribution of magnitudes of inland earthquakes km from the HORUS catalogue (Lolli *et al.* 2020) with depth < 50 km used for testing and optimization. Black dots indicate $M_w < 5.0$, green dots $5.0 \leq M_w < 5.5$, blue dots $5.5 \leq M_w < 6.0$ and red dots $M_w \geq 6.0$.

Figure S3. Molchan diagram for all target shocks with $M_w \geq 5.0$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the main text). The black continuous line indicates the performance of a purely random forecasting method that separates skilled (below the line) from unskilled (above) forecasting methods. The light blue, violet and green lines indicate the confidence limits for α =50 per cent, 5 per cent and 1 per cent, respectively. The black dashed lines indicate probability gains $G = 2$, 5, 10, 20 and 40.

Figure S4. AS score diagram for all target shocks with $M_w \geq 5.0$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the main text). The black continuous line indicates the performance of a purely random forecasting method that separates skilled (above the line) from unskilled (below) forecasting methods. The light blue, violet and green lines indicate the confidence limits for α =50 per cent, 5 per cent and 1 per cent, respectively.

Figure S5. Same as Fig. S2 for declustered (first) target shocks with $M_w \geq 5.0$ (see the text).

Figure S6. Same as Fig. S3 for declustered (first) target shocks with $M_w \geq 5.0$ (see the text).

Figure S7. Same as Fig. S2 for all target shocks with $M_w \geq 6.0$ (not-declustered).

Figure S8. Same as Fig. S3 for all target shocks with $M_w \geq 6.0$ (not-declustered).

Figure S9. Same as Fig. S2 for declustered (first) target shocks with $M_w \geq 6.0$.

Figure S10. Same as Fig. S3 for declustered (first) target shocks with $M_w \geq 6.0$.

Figure S11. Same as Fig. S2 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.0$.

Figure S12. Same as Fig. S3 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.0$.

Figure S13. Same as Fig. S2 for time-independent analysis of declustered (first) target shocks with $M_w \geq 6.0$.

Figure S14. Same as Fig. S3 for time-independent analysis of declustered (first) target shocks with $M_w \geq 6.0$.

Figure S15. Binomial probability density for all target shocks (not-declustered) and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Figure S16. Probability gain for all target shocks (not-declustered) and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Figure S17. Miss rate for all target shocks (not-declustered) for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Table S1. List of centre coordinates of CA with radius of 30 km.

Table S2. Values of variables in Molchan and AS score plots of Figs 2 and 3 for $M_w \geq 5.5$ not-declustered targets.

Table S3. Same as Table S2 for $M_w \geq 5.5$ declustered targets (Figs 4 and 5).

Table S4. Same as Table S2 for $M_w \geq 5.0$ not-declustered targets (Figs S1 and S2).

Table S5. Same as Table S2 for $M_w \geq 5.0$ declustered targets (Figs S3 and S4).

Table S6. Same as Table S2 for $M_w \geq 6.0$ not-declustered targets (Figs S5 and S6).

Table S7. Same as Table S2 for $M_w \geq 6.0$ declustered targets (Figs S7 and S8).

Table S8. Results of retrospective forecast of first main shocks (declustered targets) with $M_w \geq 5.0$ in Italy from 1960 to 2019, using Δt= 3 months (0.25 yr).

Table S9. Results of retrospective forecast of not-declustered targets with $M_w \geq 5.0$ in Italy from 1960 to 2019, using Δt= 3 months (0.25 yr).

Table S10. Same as Table S9 for not-declustered targets with $M_w \geq 5.5$.

Table S11. Same as Table S9 for not-declustered targets with $M_w \geq 6.0$.

Contamination of Frequency–Magnitude Slope (*b*-Value) by Quarry Blasts: An Example for Italy

Laura Gulia*[1] and Paolo Gasperini[1,2]

Abstract

Artifacts often affect seismic catalogs. Among them, the presence of man-made contaminations such as quarry blasts and explosions is a well-known problem. Using a contaminated dataset reduces the statistical significance of results and can lead to erroneous conclusions, thus the removal of such nonnatural events should be the first step for a data analyst. Blasts misclassified as natural earthquakes, indeed, may artificially alter the seismicity rates and then the *b*-value of the Gutenberg and Richter relationship, an essential ingredient of several forecasting models.

At present, datasets collect useful information beyond the parameters to locate the earthquakes in space and time, allowing the users to discriminate between natural and nonnatural events. However, selecting them from webservices queries is neither easy nor clear, and part of such supplementary but fundamental information can be lost during downloading. As a consequence, most of statistical seismologists ignore the presence in seismic catalog of explosions and quarry blasts and assume that they were not located by seismic networks or in case they were eliminated.

We here show the example of the Italian Seismological Instrumental and Parametric Database. What happens when artificial seismicity is mixed with natural one?

Cite this article as Gulia, L., and P. Gasperini (2021). Contamination of Frequency–Magnitude Slope (*b*-Value) by Quarry Blasts: An Example for Italy, *Seismol. Res. Lett.* **92**, 3538–3551, doi: 10.1785/0220210080.

Introduction

Data analysis is a fundamental part of science, and statistical seismology made important steps forward both in understanding and forecasting earthquake dynamics in the last decades, thanks to the increasing development of seismic networks and data acquisition techniques. Besides the main parameters (location, time, and magnitude), indeed, each event is nowadays also characterized by several additional attributes describing the source as well as the origin of the event itself.

Although databases contain an ever-increasing number of the events' properties, some of them might be lost when data are downloaded from the websites using simplified web accesses. We here show the case of the Italian Seismological Instrumental and Parametric Database (ISIDe, ISIDe Working Group, 2007), in which the event type (earthquake, quarry blast, explosion, etc.) is indicated since 1 May 2012, but such info is lost during direct downloading from the website (see Data and Resources) in .txt format. Indeed, the user can customize the search in terms of time, magnitude range, location (longitude, latitude, and depth) but cannot discriminate the event type (see Data and Resources) because the presence of nontectonic events in the database is not clearly described in the website itself. As a consequence, nonnatural events such as quarry blasts, controlled, experimental and mining explosions, may be processed together with tectonic earthquakes.

How does such loss of information impact the statistical analysis? What happens when artificial seismicity is mixed with natural ones?

The maximum magnitude of quarry and mine blasts in Europe is usually assumed to be 2.5–3.0 (Giardini *et al.*, 2004; Gulia, 2010), which corresponds to the blast of about 100–500 kg of trinitrotoluene (TNT), assuming the standard energy release of about 4 MJ per kg of TNT and the Gutenberg and Richter energy–magnitude relation. A higher threshold has been observed in the United States, where the magnitude of quarry and mine blasts can occasionally exceed magnitude 4 (Stump *et al.*, 2002). Having low magnitudes, the nonnatural events enrich the number of small earthquakes in a catalog, falsifying the relative portion of microseismicity in respect to the higher magnitudes. This might alter the relative earthquake size distribution and then the *b*-value of the

1. Department of Physics and Astronomy, University of Bologna, Bologna, Italy, https://orcid.org/0000-0002-1474-230X (LG); https://orcid.org/0000-0002-5314-0563 (PG); 2. Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

*Corresponding author: laura.gulia@unibo.it

frequency–magnitude relationship by Gutenberg and Richter (1944). The Gutenberg and Richter relationship is a fundamental ingredient of several short-term forecasting models; among them, the epidemic-type aftershock sequence (ETAS) (Ogata, 1988, 1998) that is used in operational earthquake forecasting (e.g., Jordan *et al.*, 2011), the short-term earthquake probability model (STEP; Gerstenberger *et al.*, 2005) and the foreshock traffic light system (FTLS) (Gulia and Wiemer, 2019; Gulia *et al.*, 2020). Many authors pointed out that the *b*-value is a proxy of the state of the stress of a region (e.g., Wyss, 1973), being inversely correlated to it and showed the *b*-value capability to be a "stressmeter" of the Earth's crust at different scales from laboratory specimens (e.g., Scholz, 1968) to observations (Schorlemmer and Wiemer, 2005; Schorlemmer *et al.*, 2005; Gulia *et al.*, 2010, 2016, 2018; Tormann *et al.*, 2015; Petruccelli *et al.*, 2018; Petruccelli, Gasperini, *et al.*, 2019; Petruccelli, Schorlemmer, *et al.*, 2019). The *b*-value can have a precursory drop before the failure (Papadopoulos *et al.*, 2010; Nanjo *et al.*, 2012; Schurr *et al.*, 2014; Tormann *et al.*, 2015; Gulia *et al.*, 2016; Gulia and Wiemer, 2019; Huang *et al.*, 2020) suggesting that the seismicity evolution in terms of *b*-value should be routinely monitored. Such finding is confirmed in numerous laboratory studies, showing an increasing relative proportion of larger events as the system approaches failure (e.g., Goebel *et al.*, 2013).

The higher *b*-values, resulting from an artificial enrichment of the portion of low-magnitude events in regional catalogs, can mask the spatiotemporal variations, altering the alerts and be misinterpreted as a change in the natural phenomena.

Long-term models can also be affected by falsified seismic rates and *b*-values. In the probabilistic seismic hazard assessment (PSHA, usually based on Cornell, 1968), the Gutenberg and Richter relationship defines event rates used to compute expected levels of ground shaking. PSHA, indeed, assumes a Poissonian distribution of seismicity and such requirement is generally satisfied by declustering the input catalog (Gardner and Knopoff, 1974; van Stiphout *et al.*, 2010). For such reason, in the hazard assessment, rates are estimated on declustered catalogs (Wiemer *et al.*, 2009; Field *et al.*, 2014; Petersen *et al.*, 2018). Mizrahi *et al.* (2021) show that declustering can introduce a systematic bias to the size distribution of earthquakes, potentially biasing hazard assessment, and Iervolino (2019) proposes a generalization of the hazard integral to reintroduce aftershocks in PSHA. However, at present, the seismic rates are still estimated on declustered catalogs. Once the aftershocks are removed, the relative portion of quarry blasts, if present, increases as these latter usually occur almost uniformly in time, and the *b*-value of the grid node or zone are affected by artificial events, too. As a consequence, the rates of the highest magnitudes are underestimated. Although in Italy the input dataset is usually cut at magnitude 4, in other countries (e.g., Switzerland; Wiemer *et al.*, 2009) the threshold magnitude cutoff is lower.

Stress-based spatiotemporal models that describe the aftershocks productivity, can also be biased by quarry blasts; the expected rate of earthquakes in a given magnitude range (e.g., Dieterich, 1994) is indeed a function of the background seismicity.

It is important that high *b*-values can be observed in different natural settings, such as the volcanic regions (e.g., Wyss *et al.*, 1997, 2001; Roberts *et al.*, 2015) as well as in enhanced geothermal systems during the coinjection period (Bachmann *et al.*, 2012) and in hydrocarbon reservoirs during extraction of natural gas (Muntendam-Bos *et al.*, 2017). Being able to discriminate between natural, induced, and nonnatural *b*-values can help seismologists to understand and interpret the physical phenomena under investigation.

For all the reasons mentioned earlier, nonnatural events must be identified, mapped, and excluded from the catalogues before any meaningful statistical analysis can be performed. Statistical seismologists use catalogs assuming explosions have been eliminated but, as we showed before in the case of direct download from some websites, such events can erroneously be included in the catalogs.

Italy is an ideal testing region due to the simultaneous presence of a dense seismic network and several extraction sites. In 2014, the number of mining and quarrying active extraction sites in Italy was equal to 4612 (see Data and Resources) and a detailed map is available at the Italian Institute for Environmental Protection and Research (ISPRA) website (see Data and Resources) but no information is given on which of them use explosives.

Explosives are the primary source of energy for rock breaking in the mining, quarrying, and construction industries (Sanchidrián *et al.*, 2007; Hamdi *et al.*, 2008), particularly for the building materials. Underground mines are also excavated by explosions as well as salt and coal basins are mined by blasting. Explosives are also largely employed in civil engineering (e.g., tunnel and subway) and in offshore seismic prospecting.

Wiemer and Baer (2000) proposed a purely statistical tool to identify quarry and mine blasts based on the ratio between daytime and nighttime D/N events. In the case of all-natural events, such value should be around 1 ideally or more probably slightly lower, due to the lower magnitude detection threshold in nighttime owing to the lower level of anthropic seismic noise. On the contrary, the presence of nonnatural events should increase such ratio because mine blasts are usually performed during daytime.

In this work, we first show the D/N maps performed, by the tool proposed by Wiemer and Baer (2000), for two versions of ISIDe: the one downloaded directly from the website in .txt format and the one downloaded by the Istituto Nazionale di Geofisica e Vulcanologia (INGV) webservices (see Data and Resources), selecting the event type earthquake, in the period 2005–2020. Then, by the comparison with the related *b*-value

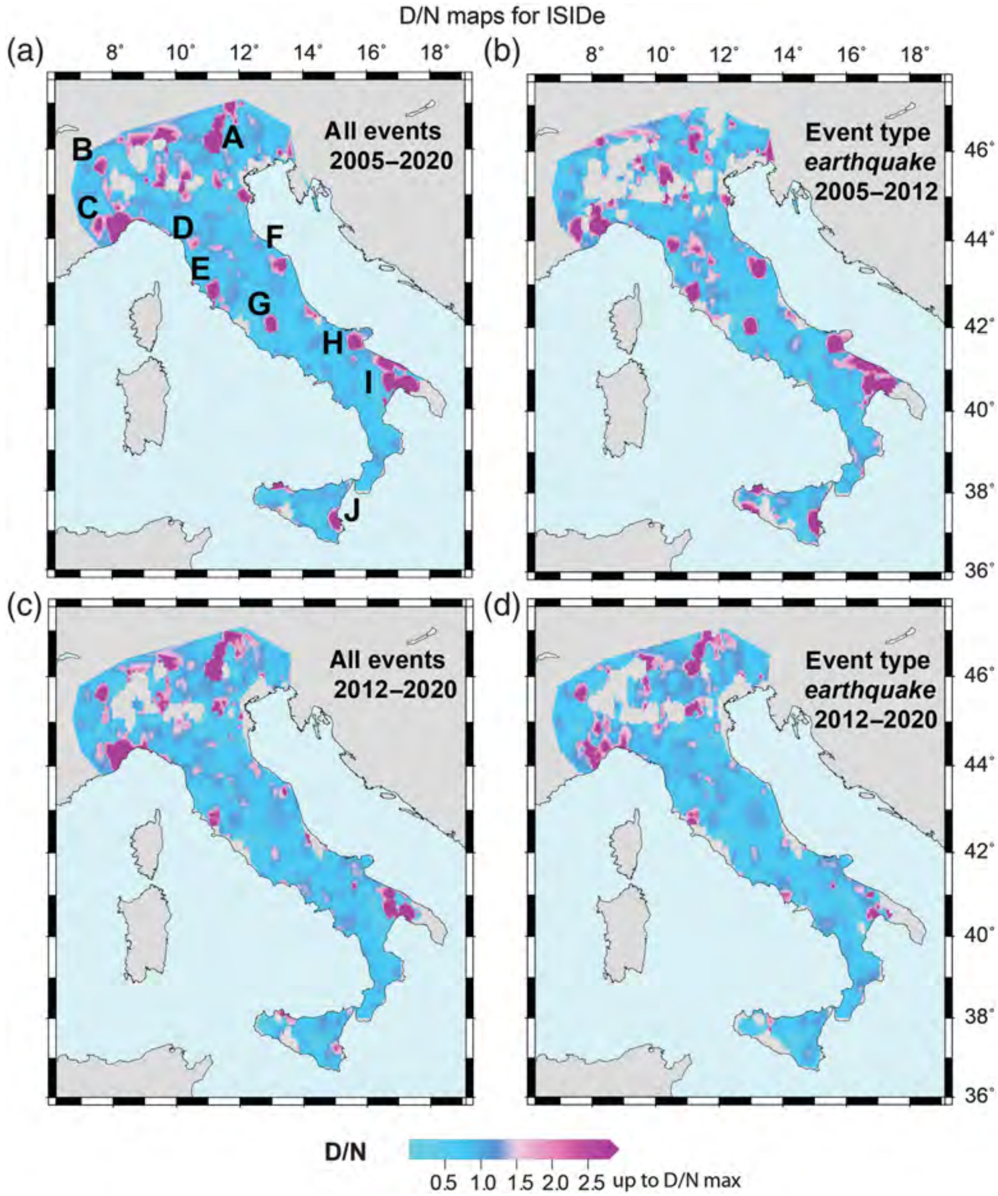


Figure 1. Maps of the daytime to nighttime ratio (D/N). (a) For the time interval 2005–2020 and for the whole dataset downloaded in .txt format (see Data and Resources). The letters A–J correspond to the excavation districts listed in Table 1. (b) For the time interval 16 April 2005–30 April 2012, in which the events are classified with event type earthquake. (c) For the time interval 1 May 2012–30 November 2020, for all the events downloaded in .txt format (see Data and Resources). (d) For the time interval 1 May 2012–30 November 2020, for all the events downloaded via webservices selecting the event type earthquake only. The maps are computed on a 10 km regularly spaced grid using the events within a 20 km radius from each node.

maps, we show the correspondence between unusually high b-values in the maps ($b > 1.4$–1.5) to the areas with the highest D/N: the presence of nonnatural events, mixed with natural ones, falsifies both the seismic rates and the b-value of the Gutenberg and Richter relationship.

Gulia (2010) mapped the D/N ratio for the available European regional catalogs, highlighting the presence of numerous quarry blasts; among them the Italian Seismicity Catalog (C.S.I. 1.1; Castello *et al.*, 2006), that contains about 100,000 earthquakes during period 1981–2002. From 1 May 2012, the indication of event types different from earthquake is routinely provided by the Italian Seismic Network: we thus repeat the same analysis on the dataset downloaded via webservice, selecting only the event type earthquake: are all the events contained in such version, earthquakes only? Are all the quarry blasts recorded by the network, correctly identified?

Data and Method

We compute the D/N ratio maps (Fig. 1a–d) for different time periods, on a 10 km regularly spaced grid using the events, within a 20 km radius from each node, taken from ISIDe (ISIDe Working Group, 2007). ISIDe contains the parameters of earthquake locations computed by the INGV National Seismic Network since 1985 but as input data for our estimates, we select events from 16 April 2005 (last accessed November 2020), when the Italian Seismic Network was reorganized and extended and the quality of hypocentral locations and magnitudes was definitely improved, to 30 November 2020.

The D/N is defined as follows:

$$Rq = NdLn/NnLd, \quad (1)$$

in which Nd is the total number of events in the daytime, Nn is the total number of events in the nighttime period, Ld is the number of hours in the daytime period, and Ln is the number of hours in the nighttime period. According to Wiemer and Baer (2000), an indicative value for the anomalous D/N is >1.5, well highlighted by the implemented color palette that, from around 1.5, abruptly changes from blue-sky to pink shades. We define as daytime the hours from 7 a.m. to 6 p.m. and nighttime the hours

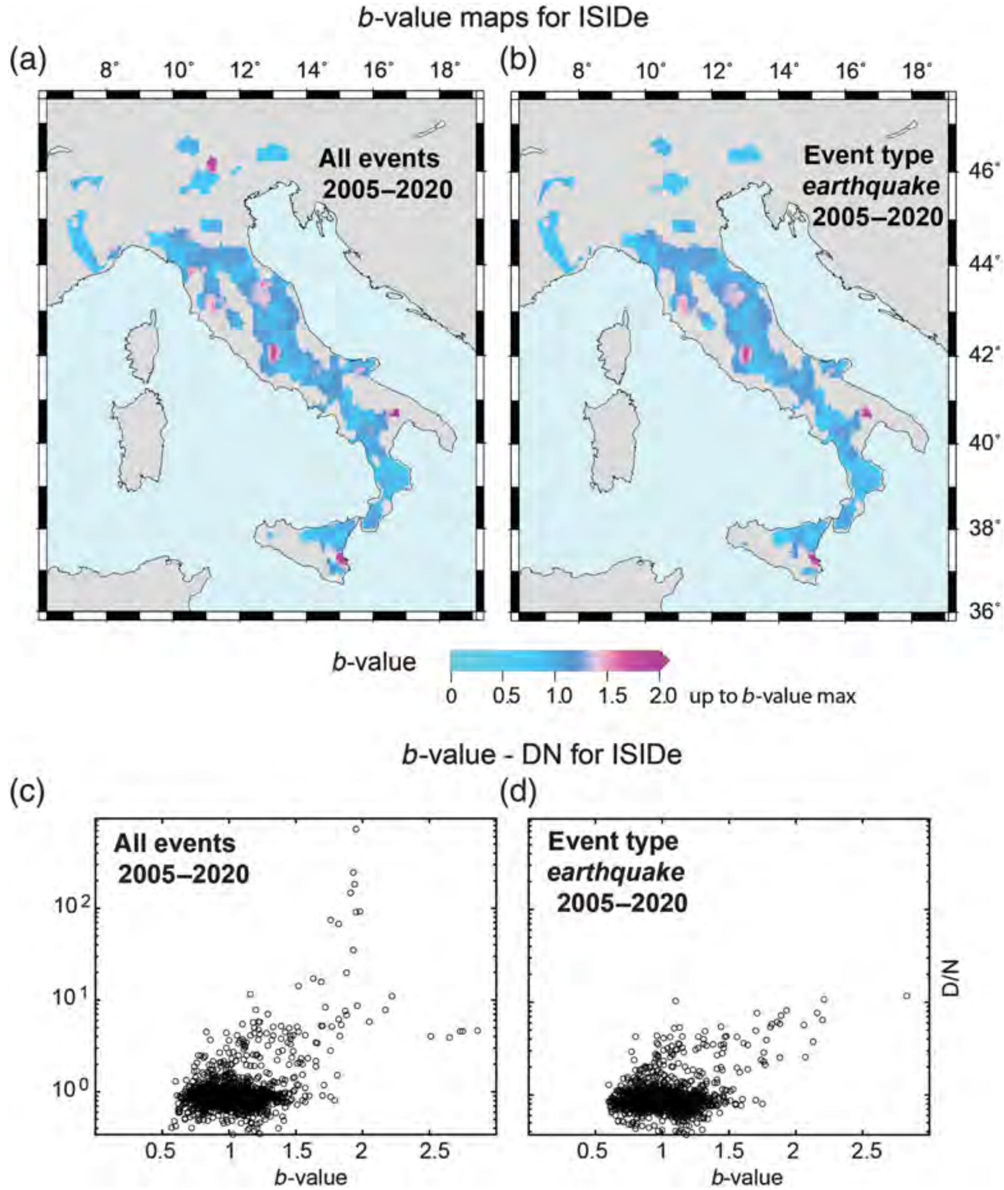


Figure 2. (a,b) Maps of the *b*-value for the time interval 16 April 2005–30 November 2020, computed on a 10 km regularly spaced grid using the events within a 20 km radius from each node for (a) the whole dataset downloaded in .txt format (see Data and Resources) and (b) the whole dataset downloaded via webservices selecting the event type earthquake only. (c,d) Plot of the D/N against the *b*-value for all the grid nodes adopted in the previous maps for (c) the whole dataset downloaded in .txt format (see Data and Resources) and (d) the whole dataset downloaded via webservices selecting the event type earthquake.

from 6 p.m. to 7 a.m. According to equation (1), the number of events in each time window are normalized to the number of hours (11 for day -*Ld*- and 13 for night -*Ln*-).

We then established three different time periods:

- 16 April 2005–30 November 2020: the whole dataset downloaded in .txt format (see Data and Resources);
- 16 April 2005–30 April 2012, when the event-type info is not yet available: natural and nonnatural events are mixed together and indicated with event type earthquake; they can be identified only by statistical analysis. Even a dataset downloaded via webservices, specifying the event type earthquakes, is contaminated by nonnatural events in the time period preceding May 2012; and
- 1 May 2012–30 November 2020: when the even-type info is available. For such time interval, we calculate two maps: one for the events downloaded in .txt format (see Data and Resources) and one selecting only the events identified as "earthquakes" via webservices (see Data and Resources).

We then calculate the *b*-value map for the two catalogs (the one containing all the events and the one contained the events classified as earthquake only) in the whole period 16 April 2005–30 November 2020 (Fig. 2a,b), using the same grid and the same radius adopted in Figure 1a–d. The magnitude of completeness is estimated at each grid node (maximum curvature, Wiemer and Wyss, 2000) with a 0.2 correction (Woessner and Wiemer, 2005) and we require a minimum sample size of 50 events above M_c to compute the *b*-value by the maximum-likelihood method. We could not estimate a *b*-value for all the grid nodes with an already associated D/N due to the minimum number of events above M_c required for the *b*-value.

For some of the grid nodes with anomalous D/N, we show the histograms of the hour of the events that represents a first and effective tool to identify the presence of quarry blasts in a catalog. Quarry and mine-rich regions, indeed, reveal a typical pattern, with a very large number of events during daytime hours.

TABLE 1
List of the Excavation Districts Labeled in Figure 1 with the Letters A–J and the Extracted Material

Letter	Excavation District	Material
A	Albiano–Trento	Porphyry
B	Issogne–Gressoney	Green marble, limestone, slate
C	Savona	Limestone and quartzite
D	Apuane–Garfagnana	White, red, and black marble
E	Maremma	Limestone
F	Cingoli–Marche	Limestone
G	Riofreddo	Limestone and basalt
H	Gargano	Marble and limestone
I	Altamura–Matera	Limestone and marble
J	Siracusa	Porphyry and basalt

Results

In all the four maps, anomalous high D/N are sparse on the whole Italian territory and most of them have been identified and described in Gulia (2010), corresponding to known excavation districts, such as Apuane–Garfagnana and Fabriano (respectively, areas D and F in Fig. 1a; see Table 1 for the list of the excavation districts in Fig. 1a). The user, in the case of a download in .txt format (see Data and Resources), would unconsciously download also a big number of nonnatural events that would then be erroneously processed as earthquakes (Fig. 1a).

Because the event types different from earthquake are specified only from 1 May 2012, even a dataset downloaded via webservices specifying the event type as *earthquake* would be contaminated by nonnatural events till such date: Figure 1b shows the D/N map from 2005 to 30 April 2012. Unluckily, copious nonnatural events are inevitably downloaded from webservices in any case also by a user who selects only the event type earthquake. The high-contaminated regions are about the same of Figure 1a.

The last two maps (Fig. 1c,d) show and compare the D/N, in the time period staring from the event type identification (May 2012), for the two catalogs. Here also the contaminated regions are about the same of Figure 1a,b, but often with a smaller size and value, due to the partial blasts' identification, that somewhat reduces the gap between the number of daily and nighttime events.

The similarity between these two maps (2012–2020) compared to the previous period (2005–2012, Fig. 1b), indeed, points out the improved capability of the network operators to detect and identify nonnatural events in some areas, however a very significant contamination still persists. Before analyzing in detail the most contaminated regions, we compare the D/N maps in Figure 1 with the two b-value maps in Figure 2a,b, for the whole time period and for the two catalogs: most of the regions with an unusually high b-value (>1.4–1.5) and a corresponding high D/N (>1.5) are well-known active excavations districts. Furthermore, the correspondence between high D/N and high b-values is well represented in Figure 2c,d, in which we plot the D/N and the b-value for all the grid nodes of the maps: the b-values in the range that is usually observed in different natural settings (0.6–1.2/1.3) are well correlated with the typical values of the D/N ratio, that is around 1 and lower. On the contrary, unusually high b-values correspond mainly to the highest D/N ratios.

In the Introduction, we wonder whether all the quarry blasts recorded by the network are correctly identified: by the observation of the above maps and plots, we may already claim they are not. However, hereafter, we will list and comment in detail the regions, labeled from A to J in Figure 1a, by the analysis of some specific grid nodes with anomalous D/N and the comparison, when possible, with the corresponding b-values, for the time periods 16 April 2005–30 April 2012 and 1 May 2012–30 November 2020.

The histograms containing the hour of the day of the total number of the events for the regions A, B, C, and E are shown in Figure 3, together with their seismicity maps. The relative D/N ratios are displayed too. In these four regions, most of the events have been recorded between 10 a.m. and 3 p.m., with a minimum around the lunch break (e.g., Fig. 3c,d,k,o), as already pointed out in Gulia (2010). The spatial clusters of events locate the active quarries (Castello and Pagagnone, 2016; see Data and Resources). In the regions A and B (Fig. 3a–h) the D/N of the second time period, that is when the catalog should contain only earthquakes, is even higher than in the previous one. On the contrary, the D/N in the regions C and E (Fig. 3i–p) decreases with time but remains higher than 1.5, indicating a partial identification of nonnatural events; however, the three very restricted, well-defined, and isolated spatial clusters in the seismicity map of the grid node in region E (Fig. 3n) are the best visual example, among the several ones we analyzed, of the highly suspected nonnatural origin of the events. In this grid node, the daytime events have been recorded mostly between 9.55 a.m. and 10 a.m. during spring and summer and between 10.55 a.m. and 11 a.m. during autumn and winter, indicating a one-hour shifted time of the blasting operations due to the daylight-saving time on spring–summer in Italy. Such peculiarity also characterizes the events in the time period following May 2012, in which the quarry blasts should had been identified and classified with the correct event type by the network.

For these four regions, where quarries and mines are active and the natural seismicity is very low, we could not calculate and compare the b-values of the two time periods.

But what happen in seismically active regions with working quarries and mines? That is the case of the regions labeled as D

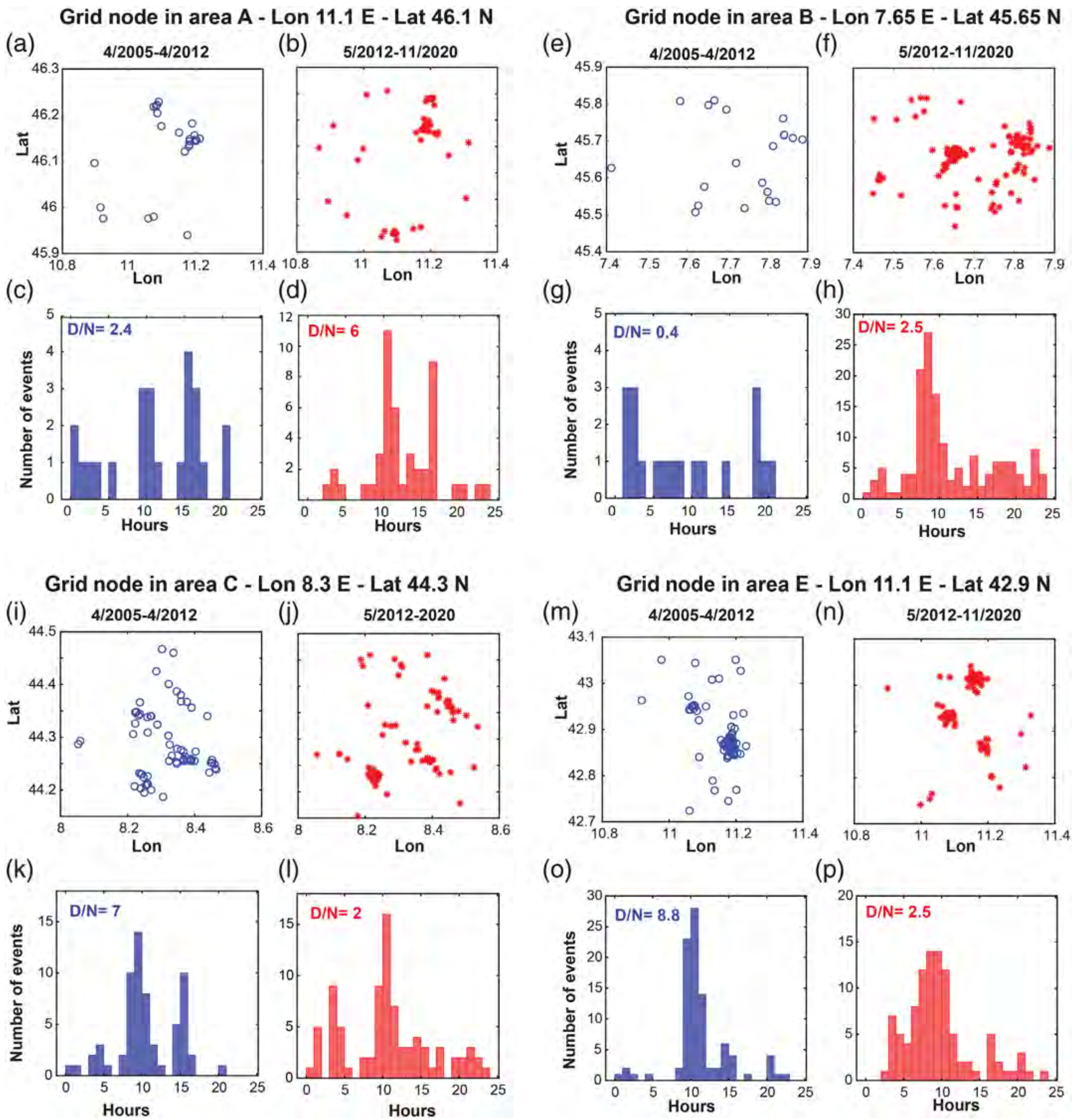


Figure 3. Spatial and statistical analysis of four grid nodes in the areas labeled as A, B, C, and E in Figure 1a. (a,b) Seismicity maps and histograms of the hours of the events (c,d) for a grid in area A for, respectively, the events downloaded (see Data and Resources) in the period 16 April 2005–30 April 2012 and for the events downloaded via webservices selecting the event type earthquake from 1 May 2012–30 November 2020. (e, f) Seismicity maps and histograms of the hours of the events (g, h) for a grid in area B for, respectively, the events downloaded (see Data and Resources) in the period 16 April 2005–30 April 2012 and for the events downloaded via webservices selecting the event type earthquake from 1 May 2012–30 November 2020. (i,j) Seismicity maps and histograms of the hours of the events (k,l) for a grid in area C for, respectively, the events downloaded (see Data and Resources) in the period 16 April 2005–30 April 2012 and for the events downloaded via webservices selecting the event type earthquake from 1 May 2012–30 November 2020. (m,n) Seismicity maps and histograms of the hours of the events (o,p) for a grid in area E for, respectively, the events downloaded (see Data and Resources) in the period 16 April 2005–30 April 2012 and for the events downloaded via webservices selecting the event type earthquake from 1 May 2012–30 November 2020.

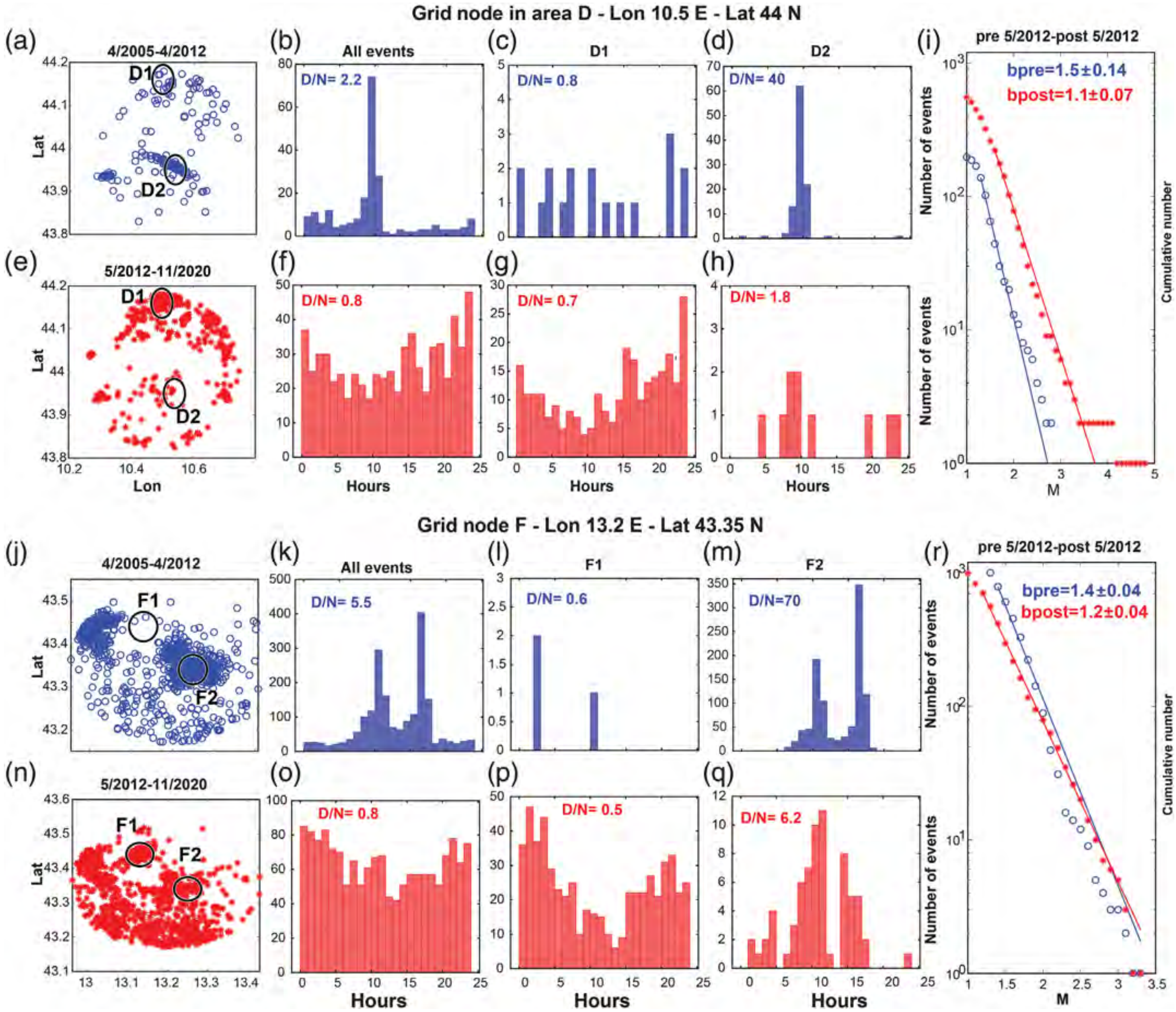


Figure 4. Spatial and statistical analysis of two grid nodes in the areas labeled as D and F. Grid node in area D in the time period 16 April 2005–30 April 2012: (a) seismicity map with two spatial clusters, D1 and D2, circled in black; histogram of the hour of events and relative D/N for (b) all the events in the grid node, (c) all the events in the D1 spatial cluster, and (d) all the events in the D2 spatial cluster. Grid node in area D in the time period 1 May 2012–30 November 2020: (e) seismicity map with two spatial clusters, D1 and D2, circled in black; histogram of the hour of events and relative D/N for (f) all the events in the grid node, (g) all the events in the D1 spatial cluster, and (h) all the events in the D2 spatial cluster. (i) Frequency–magnitude distributions for all the events in the grid node D from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks). Grid node in area F in the time period 16 April 2005–30 April 2012: (j) seismicity map with two spatial clusters, F1 and F2, circled in black; histogram of the hour of events and relative D/N for (k) all the events in the grid node, (j) all the events in the F1 spatial cluster, and (m) all the events in the F2 spatial cluster. Grid node in area F in the time period 1 May 2012–30 November 2020: (n) seismicity map with two spatial clusters, F1 and F2, circled in black; histogram of the hour of events and relative D/N for (o) all the events in the grid node, (p) all the events in the F1 spatial cluster, and (q) all the events in the F2 spatial cluster. (r) Frequency–magnitude distributions for all the events in the grid node F from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks).

and F in Figure 4 and G in Figure 5, that are excavation districts, also affected by natural seismicity.

In Figure 4, we show a seismically active area in northern Tuscany (area D in Fig. 1a): the Apuane–Garfagnana district, that is a well-known excavation district since the age of ancient Romans. The white marble, also known as "white gold," that artists like Michelangelo Buonarroti and Antonio Canova transformed into world heritage masterpieces, was mined here.

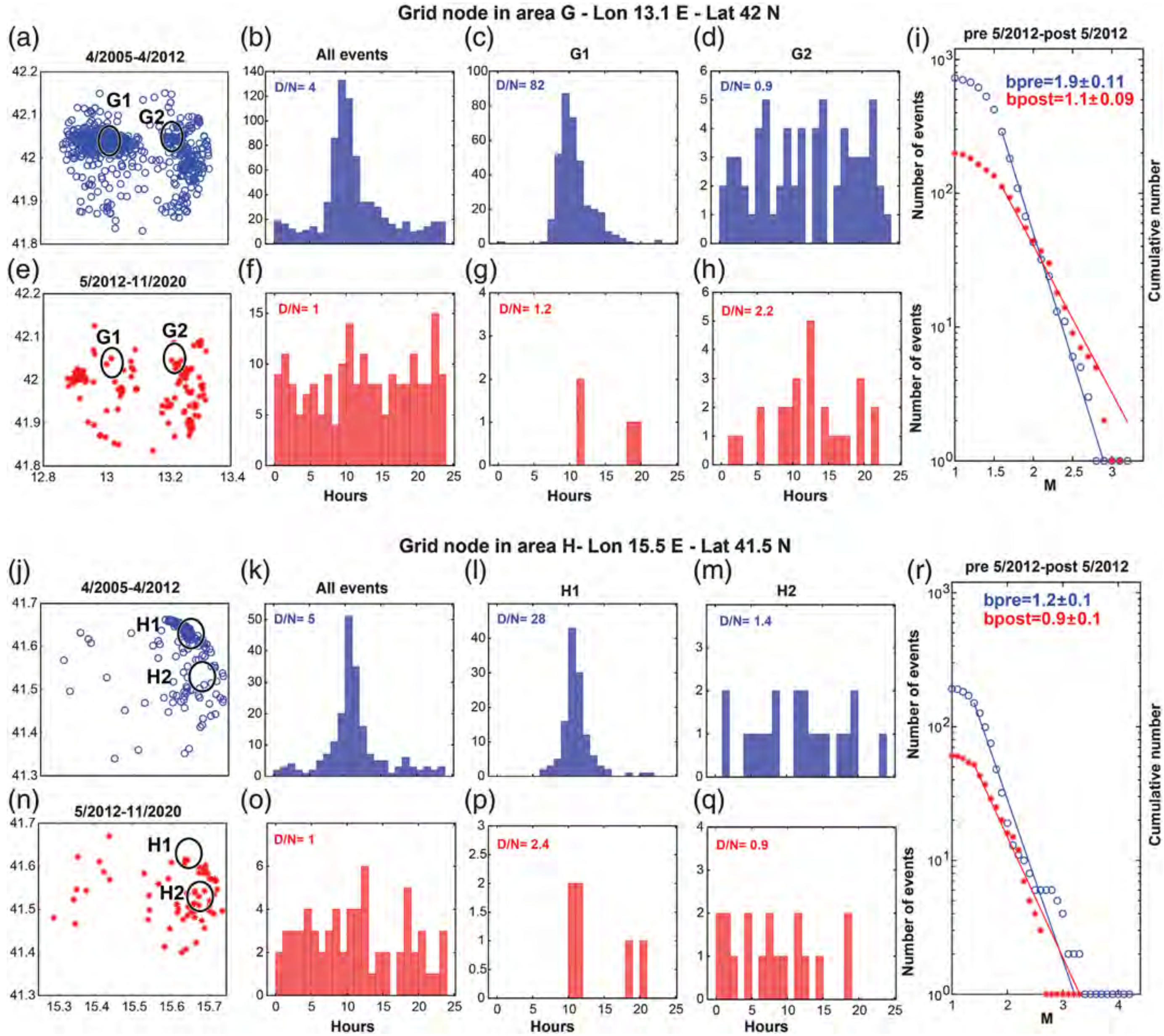


Figure 5. Spatial and statistical analysis of two grid nodes in the areas labeled as G and H. Grid node in area G in the time period 16 April 2005–30 April 2012: (a) seismicity map with two spatial clusters, G1 and G2, circled in black; histogram of the hour of events and relative D/N for (b) all the events in the grid node, (c) all the events in the G1 spatial cluster, and (d) all the events in the G2 spatial cluster. Grid node in area G in the time period 1 May 2012–30 November 2020: (e) seismicity map with two spatial clusters, G1 and G2, circled in black; histogram of the hour of events and relative D/N for (f) all the events in the grid node, (g) all the events in the G1 spatial cluster, and (h) all the events in the G2 spatial cluster. (i) Frequency–magnitude distributions for all the events in the grid node G from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks). Grid node in area H in the time period 16 April 2005–30 April 2012: (j) seismicity map with two spatial clusters, H1 and H2, circled in black; histogram of the hour of events and relative D/N for (k) all the events in the grid node, (j) all the events in the H1 spatial cluster, and (m) all the events in the H2 spatial cluster. Grid node in area H in the time period 1 May 2012–30 November 2020: (n) seismicity map with two spatial clusters, H1 and H2, circled in black; histogram of the hour of events and relative D/N for (o) all the events in the grid node, (p) all the events in the H1 spatial cluster, and (q) all the events in the H2 spatial cluster. (r) Frequency–magnitude distributions for all the events in the grid node H from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks).

In the first time period, from April 2005 till the end of April 2012, the events in this small area are spatially clustered (D2) and the histogram of the hour of the events (Fig. 4b) shows the typical pattern of quarry rich area: the events are concentrated between 9 a.m. and 11 a.m. In the following time period (from 1 May 2012 to 30 November 2020), the events are still spatially clustered (Fig. 4e) but toward the north (D1) and the histogram of the hour of the events (Fig. 4f) shows now the typical pattern of a slightly contaminated area. The nighttime hours have the highest peaks, but there are still peaks around 10 a.m. and 3 p.m. That is due to the simultaneous identification of most of blasts (The D/N of D2 passes from 40 to 1.8, e.g., Fig. 4d,h) and to a seismic sequence with a maximum M_L of 4.8 that hit the region on January 2013 (D1): its aftershocks increases the number of events in the grid nodes that passed from about 30 events per year to 90 whereas the overall b-value decreases from 1.5 (almost all blasts) to 1.1 (blasts and aftershocks; Fig. 4i). It is important that in the epicentral area of the M_L 4.8, the D/N remains below 1, even if a very small contamination is clear (hours 11 a.m. and 3 p.m., Fig. 4g), in fully agreement with Wiemer and Baer (2000): the detection threshold is generally lower during the day due to the ambient noise and, as a consequence, regions not containing quarries generally show a decrease in the number of events detected during the day and an increase during the night. Thus, we should expect a D/N lower than 1, in the case of natural seismicity and this grid node is a perfect illustration. The same grid node also illustrates the example of nonnatural events, in D2.

In this region, the b-value results from the mixing of natural and nonnatural events: there has been a strong improvement with time and most of blasts are now correctly identified by the network, even if a further effort is required to identify and remove all of them.

In Figure 4j–r, we show the case of a grid node in a seismically active area with only low-magnitude events (area F in Fig. 1a; the Cingoli district, Marche), that is another well-known and wide extraction district of the country, already described in Gulia (2010). In the seismicity maps (Fig. 4j,n), the events are spatially clustered in few main areas. We here choose two of them, named F1 and F2, to compare the evolution with time of a seismically active area (F1) with an extraction one (F2). Before May 2012, only three events have been recorded in F1 and more than 80% of the events in the whole grid node occurred during daily hours: the D/N of the node is 5.5 (Fig. 4k); after such date, the D/N passes to 0.8, showing a small blast contamination (Fig. 4o). The D/N of the natural seismicity in F1 is always below 1, whereas the D/N in F2 decreases from 70 to about 6: as already shown in the previous case, most of blasts are nowadays identified by the network, but several ones still remain.

The overall b-value of the events in the grid node, before May 2012, is 1.4 (Fig. 4r); a very high value respect to the one expected from this area, considering its prevalent style-of-faulting (Gulia and Wiemer, 2010). After such date, the b-value decreases to 1.2, possibly due to the increment of natural seismicity and the contemporary partial identification of nonnatural events. As for the grid node in region D, the overall b-value results from the mixture of natural and nonnatural events.

Figure 5 illustrates two interesting case studies: the first one is about a grid node in the region labeled as G in Figure 1a, Central Italy. Natural and nonnatural seismicity are mixed together, as revealed by the histograms of the hour of the events of two small spatial clusters (G1 and G2): before May 2012, 349 events out of 354, in G1, are daytime events and the relative b-value (1.9 in Fig. 5i) is unusually high, more than twice that typical, according to Gulia and Wiemer (2010), for the region. The G1 excavation area has been successfully identified after 2012, indeed ISIDe, in the following period, contains only four events, two during daytime and two during nighttime. Few blasts still remain also in the adjacent areas G2; however, this area has been successfully located by the network operators and most of nonnatural events identified. The overall b-value decreases to 1.1.

As well as for the previous case, in the grid node in area H (Fig. 5j–r) most of blasts have been successfully identified. The D/N of the whole area passes from 5 to 1 and the b-value decreases from 1.2 (resulting from many blasts and few natural events) to 0.9 (few blasts Fig. 5r).

The last two case studies, shown in Figure 6 (areas labeled as I and J in Fig. 1a), exhibit a similar evolution in time. Most of the seismicity before May 2012 is composed by nonnatural events. The D/N are, respectively, 2.6 and 43 (Fig. 6b,k) and the b-values 1.7 and 2.2 (Fig. 6i,r), among the highest values in the whole country. After May 2012, both D/N and b-value decreases, possibly due to the correct but still partial identification of nonnatural events.

Finally, because quarry blasts are performed during the day, the nighttime events should be all tectonic and scale with a lower b-value respect to the daytime events. We then divided the events according to the hour of the day for the above grid nodes with low-magnitude events only and show the comparison of the frequency–magnitude distributions for the two periods (Fig. 7). The theoretical expected behavior is fully confirmed for all the five nodes. The nighttime b-value are all well below the daytime ones, in some cases less than the half.

Discussion and Conclusion

Every day, beyond tectonic events, seismic networks detect several nonnatural earthquakes: among them, quarry and mine blasts are the most numerous anthropogenic recorded events. Often, such events are not identified and thus collected together with tectonic events.

Having low magnitudes, the artificial events enrich the number of small earthquakes in a catalog, contaminating the natural signals and seismicity datasets adulterating the relative portion of microseismicity respect to the higher

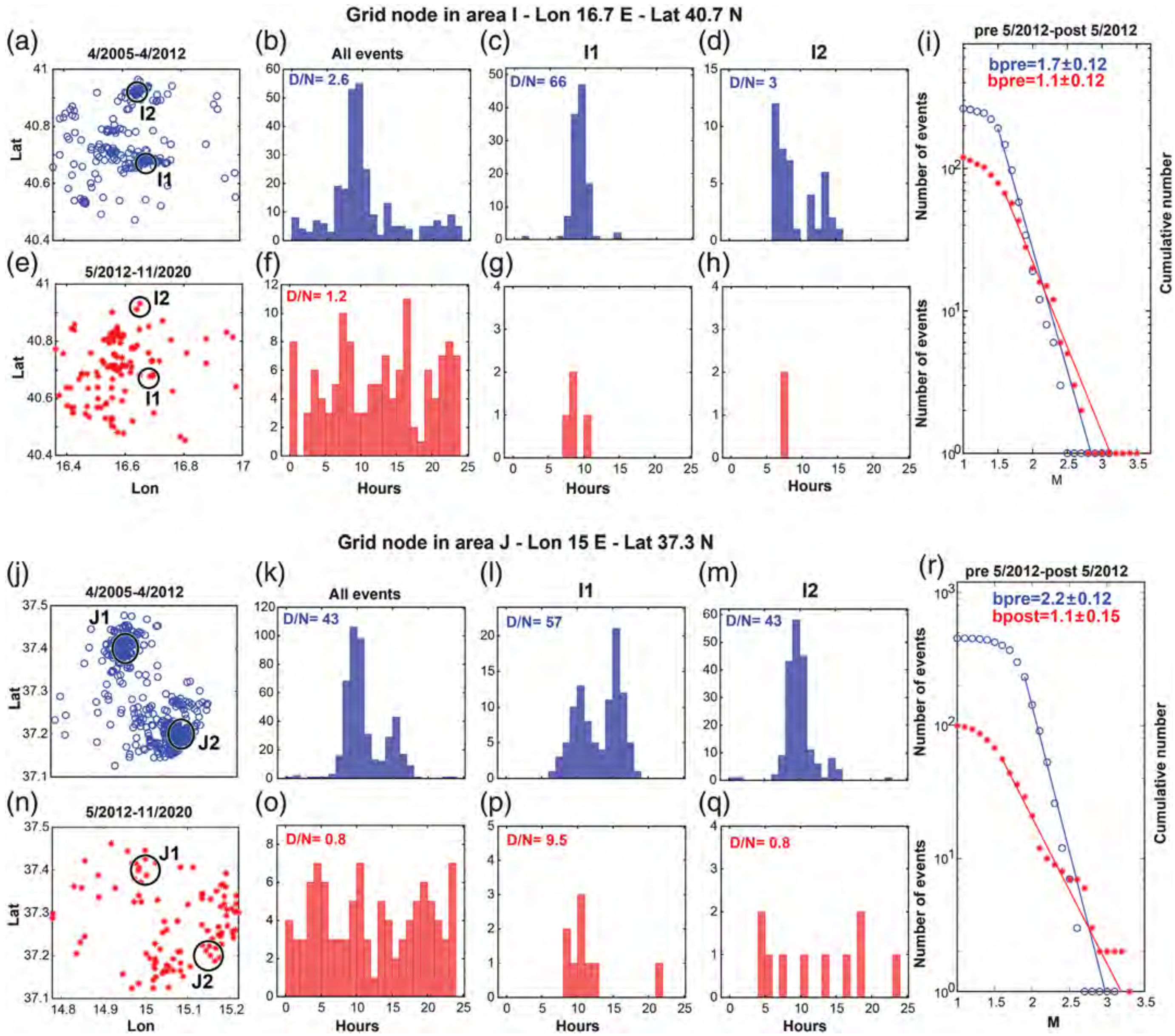


Figure 6. Spatial and statistical analysis of two grid nodes in the areas labeled as I and J. Grid node in area I in the time period 16 April 2005–30 April 2012: (a) seismicity map with two spatial clusters, I1 and I2, circled in black; histogram of the hour of events and relative D/N for (b) all the events in the grid node, (c) all the events in the I1 spatial cluster, and (d) all the events in the I2 spatial cluster. Grid node in area I in the time period 1 May 2012–30 November 2020: (e) seismicity map with two spatial clusters, I1 and I2, circled in black; histogram of the hour of events and relative D/N for (f) all the events in the grid node, (g) all the events in the I1 spatial cluster, and (h) all the events in the I2 spatial cluster. (i) Frequency–magnitude distributions for all the events in the grid node I from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks). Grid node in area J in the time period 16 April 2005–30 April 2012: (j) seismicity map with two spatial clusters, J1 and J2, circled in black; histogram of the hour of events and relative D/N for (k) all the events in the grid node, (j) all the events in the J1 spatial cluster, and (m) all the events in the J2 spatial cluster. Grid node in area J in the time period 1 May 2012–30 November 2020: (n) seismicity map with two spatial clusters, J1 and J2, circled in black; histogram of the hour of events and relative D/N for (o) all the events in the grid node, (p) all the events in the J1 spatial cluster, and (q) all the events in the J2 spatial cluster. (r) Frequency–magnitude distributions for all the events in the grid node J from 16 April 2005 to 30 April 2012 (blue circles) and from 1 May 2012 to 30 November 2020 (red asterisks).

magnitudes. The resulting seismic rate changes and the relative earthquake size distribution, or b-value of the Gutenberg and Richter (1944), are falsified.

The natural signal is then contaminated, impacting many short-term forecasting models, such as ETAS (Ogata, 1988, 1998) or the FTLS (Gulia and Wiemer, 2019), in which seismic

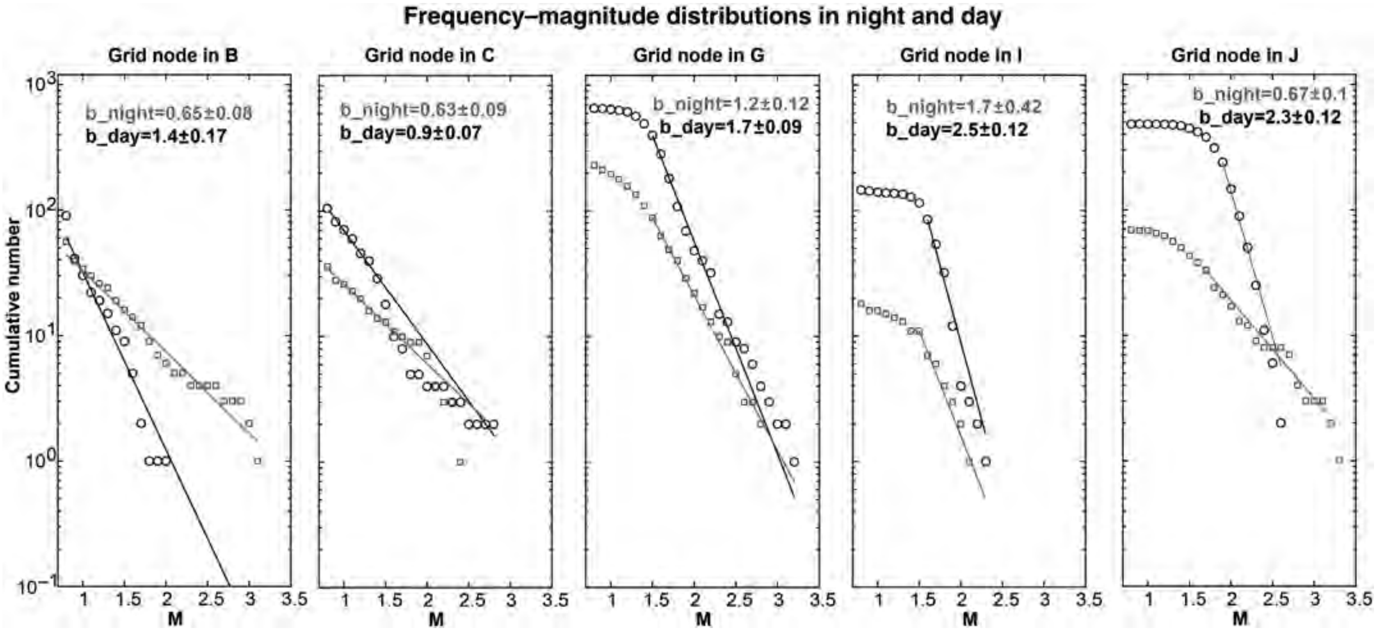


Figure 7. Comparison of the frequency–magnitude distributions for the five grid nodes in Figures 3–6 with low-magnitude events only (B, C, G, I, and J) for all the daytime (black circles) and nighttime (gray squares) events for the dataset downloaded via webservices (event type earthquake) from16 April 2005 to 30 November 2020.

rates and b-value, both inferred from the Gutenberg and Richter relationship, are a basic ingredient.

Long-term analysis, such as PSHA, can be also impacted by nonnatural events. Quarry blasts, if present, increase the b-value of the grid node or zone affected by artificial events, resulting in an underestimation of the highest magnitude rates.

In this work, we show the example of the ISIDe, in which the event types of the nonnatural events are available since 1 May 2012 only. In the custom search page (see Data and Resources), the user can set the starting and the end date, the magnitude as well as the latitude, longitude and the depth ranges, but the event type is not mentioned. Being the event type is not available among the custom search options in ISIDe, the user downloads also nonnatural events, not being conscious of this.

Other online available catalogs, for example, The Advanced National Seismic System Comprehensive Earthquake Catalog (ComCat; see Data and Resources) by U.S. Geological Survey (USGS), allows the user to also set the event type among the advanced options. The event type option for ISIDe can be set only when retrieving data via webservices.

We download and compare two versions of ISIDe, one downloaded at – see Data and Resources and one downloaded via webservices, specifying the event type earthquake. Are all the nonnatural events correctly recognized? If not, how many nonnatural events, misclassified as earthquakes, do impact the b-value?

As a first test, we spatially map the ratio of D/N events, proposed by Wiemer and Baer (2000) to investigate the presence of quarry blasts, for different time intervals of the two catalogs, showing that:

- in the whole period (16 April 2005–30 November 2020), the dataset downloaded (see Data and Resources) in .txt format is heavily contaminated by quarry blasts in the whole Italian territory (Fig. 1a);
- the period 16 April 2005–30 April 2012, in which the event type is always indicated as earthquake and thus the events are common to both catalogs, shows an even wider contamination that can be detected only by statistical analysis (Fig. 1b); and
- the following time period, that is 1 May 2005–30 November 2020, is still highly contaminated by nonnatural events in both catalogs (Fig. 1c,d). However, in the known extraction districts, there is a general improvement for the catalog downloaded via webservices. Some areas where extractions started after 2012 (e.g., area B, Fig. 3b) seem to be unknown.

We then spatially map the b-value for the two catalogs using the same grid and radius already adopted for the D/N maps (Fig. 2a,b): because we required a minimum number of events above M_c to calculate the b-value; not all the grid nodes with a D/N have a corresponding b-value. The regions with unusually high b-values (>1.4–1.5) well correspond to the regions with high D/N in Figure 1. To further highlight the correlation between high b-value and high D/N, we plot the two values for the same grid nodes (Fig. 2c,d), confirming the correspondence.

Some of the grid nodes with anomalous D/N have been analyzed in detail. For such areas, we show the seismicity maps, the histograms of the hour of the events (Figs. 3–6) and, when possible, the frequency–magnitude distributions before and after 1 May 2012. There has been a general improvement with time, and several quarry blasts are now correctly identified by the network's operators; however many ones still remain, affecting the *b*-value estimations and increasing the portion of low-magnitude events.

The seismicity of the grid node in Figure 4a–i offers an incisive and clear example of artificial *b*-value temporal fluctuations due to nontectonic events. It is an excavation area hit by an M_L 4.8 in 2013. If we compare the frequency–magnitude distributions of the two periods (Fig. 4i), we note an apparent 27% *b*-value decrease (from 1.5 to 1.1). According to several forecasting models and evidence from laboratory specimens, such decrease should suggest a change in the physical condition of the region, resulting somewhat in an alert for an impending strong earthquake.

Finally, because quarry blasts are performed during the day, we expect that all the nighttime events are natural earthquakes. We compare, in Figure 7, the frequency–magnitude distributions of D/N events for some of the previous analyzed grid nodes with no events with magnitude greater than 3.5; all the daytime *b*-value are higher than the nighttime ones.

Our analysis reveals the presence of numerous quarry blasts in the ISIDe (ISIDe Working Group, 2007) in the period 16 April 2005–30 April 2012, misclassified as earthquakes. After 1 May 2012, there is a general improvement in identifying the event type. However, many quarry blasts are still not correctly classified and such improvement is lost when the user downloads the event list in .txt format (see Data and Resources).

Data and Resources

The Italian Seismological Instrumental and Parametric Database (ISIDe, ISIDe Working Group, 2007) is available at http://terremoti.ingv.it/en/search and from the Istituto Nazionale di Geofisica e Vulcanologia (INGV) webservices for full download (webservices.ingv.it/fdsnws/event/1/). Both figures and calculations were performed by MATLAB, available at www.mathworks.com/products/matlab. The number of mining and quarrying active extraction sites is available at http://www4.istat.it/. The Italian Institute for Environmental Protection and Research (ISPRA) is available at https://www.isprambiente.gov.it/en/istitute. The INGV earthquake event is available at http://webservices.ingv.it/fdsnws/event/1/. The information about authorized quarries are available at https://www.regione.vda.it/territorio/territorio/attivita_estrattive/cave/cave_autorizzate_i.aspx and http://www.pianidibacino.ambienteinliguria.it/SV/03centa/varianti/DDG_2019_7664.pdf. The U.S. Geological Survey (USGS) earthquake catalog is available at https://earthquake.usgs.gov/earthquakes/search/. All websites were last accessed in February 2021.

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

Acknowledgments

The authors thank the editors, Allison Bent and Anastasia Pratt, and two anonymous reviewers for helping in improving and clarifying the article. The authors also thank Gianfranco Vannucci and Stefan Wiemer for their comments on the preliminary version of the manuscript, and Barbara Lolli and Daniele Randazzo for having provided the catalog file from WebServices with the indication of event types. This study was supported by the Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115 and partially funded by the Pianeta Dinamico-Working Earth INGV-MUR project.

References

Bachmann, C., S. Wiemer, B. P. Goertz-Allmann, and J. Woessner (2012). Influence of pore-pressure on the event-size distribution of induced earthquakes, *Geophys. Res. Lett.* **39,** no. 9, L09302 , doi: 10.1029/2012GL051480.

Castello, B., G. Selvaggi, C. Chiarabba, and A. Amato (2006). CSI catalogo della sismicità italiana 1981–2002, versione 1.1, Istituto Nazionale di Geofisica E Vulcanologia, Rome, Italy, available at http://www.ingv.it/CSI/ (last accessed October 2008) (in Italian).

Castello, C., and M. Pagagnone (2016). Minerali, miniere e cave del Parco Naturale Mont Avic (Comuni di Champdepraz e Champorcher - Valle d'Aosta - Alpi Occidentali), *Rev. Valdôtaine Hist. Nat.* **70,** 7–56, available at http://www.sfv.it/public/uploads/Revue/2016%2070/Revue2016_1_Castello%20&%20Paganone.pdf (last accessed April 2021) (in Italian).

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58,** 1583–1606.

Dieterich, J. H. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.* **99,** 2601–2618.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, *et al.* (2014). Uniform California Earthquake Rupture Forecast, version 3 (UCERF3)—The time-independent model, *Bull. Seismol. Soc. Am.* **104,** no. 3, 1122–1180.

Gardner, J. K., and L. Knopoff (1974). Is the sequence of earthquake in southern California, with aftershocks removed, Poissonian? *Bull. Seismol. Soc. Am.* **66,** 1271–1302.

Gerstenberger, M. C., S. Wiemer, L. M. Jones, and P.A. Reasenberg (2005). Real-time forecasts of tomorrow's earthquakes in California, *Nature* **435,** no. 7040, 328–331, doi: 10.1038/nature03622.

Giardini, D., S. Wiemer, D. Fäh, and D. Deichmann (2004). Seismic hazard assessment of Switzerland, *Report*, Swiss Seismological Service, ETH Zurich, 88 pp.

Goebel, T. H. W., D. Schorlemmer, T. W. Becker, G. Dresen, and C. G. Sammis (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments, *Geophys. Res. Lett.* **40,** 2049–2054, doi: 10.1002/grl.50507.

Gulia, L. (2010). Detection of quarry and mine blast contamination in European regional catalogues, *Nat. Hazards* **53,** 229–249, doi: 10.1007/s11069-009-9426-8.

Gulia, L., and S. Wiemer (2010). The influence of tectonic regimes on the earthquake size distribution: A case study for Italy, *Geophys. Res. Lett.* **37,** L10305, doi: 10.1029/2010GL043066.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** 193–199.

Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). The effect of a mainshock on the size distribution of the aftershocks, *Geophys. Res. Lett.* **45,** doi: 10.1029/2018GL080619.

Gulia, L., T. Tormann, S. Wiemer, M. Herrmann, and S. Seif (2016). Short-term probabilistic earthquake risk assessment considering time dependent b values, *Geophys. Res. Lett.* **43,** 1100–1108, doi: 10.1002/2015GL066686.

Gulia, L., S. Wiemer, and D. Schorlemmer (2010). Asperity-based earthquake likelihood models for Italy, *Ann. Geophys.* **53,** no. 3, 63–75, doi: 10.4401/ag-4843.

Gulia, L., S. Wiemer, and G. Vannucci (2020). Pseudoprospective evaluation of the foreshock traffic-light system in ridgecrest and implications for aftershock hazard assessment, *Seismol. Res. Lett.* **91,** no. 5, 2828–2842, doi: 10.1785/0220190307.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34,** 185–188.

Hamdi, E., N. B. Romdhane, J. du Mouza, and J. M. Le Cleac'h (2008). Fragmentation energy in rock blasting, *Geotech. Geol. Eng.* **26,** 133–146.

Huang, H., L. Meng, R. Burgmann, W. Wang, and K. Wang (2020). Spatio-temporal foreshock evolution of the 2019 M 6.4 and M 7.1 Ridgecrest, California earthquakes, *Earth Planet. Sci. Lett.* **551,** 116582, ISSN 0012-821X, doi: 10.1016/j.epsl.2020.116582.

Iervolino, I. (2019). Generalized earthquake counting processes for sequence-based hazard, *Bull. Seismol. Soc. Am.* **109,** no. 4, 1435–1450.

ISIDe Working Group (2007). Italian seismological instrumental and parametric database (ISIDe), Istituto Nazionale di Geofisica e Vulcanologia (INGV), Rome, Italy, doi: 10.13127/ISIDE.

Jordan, T. H., Y. Chen, R. Madariaga, I. Main, W. Marzocchi, G. Papadopulos, G. Sobolev, K. Yamaoka, and J. Zschau (2011). Operational earthquake forecasting: State of knowledge and guidelines for implementation, final report of the international commission on earthquake forecasting for civil protection, *Ann. Geophys.* **54,** no. 4, 315–391.

Mizrahi, L., S. Nandan, and S. Wiemer (2021). The effect of declustering on the size distribution of mainshocks, *Seismol. Res. Lett.* doi: 10.1785/0220200231.

Muntendam-Bos, A., J. Roset, and H. De Waal (2017). The effect of imposed production measures on gas extraction induced seismic risk, *Neth. J. Geosci.* **96,** no. 5, S271–S278, doi: 10.1017/njg.2017.29.

Nanjo, K. Z., N. Hirata, K. Obara, and K. Kasahara (2012). Decade-scale decrease in b value prior to the M9-class 2011 Tohoku and 2004 Sumatra quakes, *Geophys. Res. Lett.* **39,** L20304, doi: 10.1029/2012GL052997.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** no. 401, 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.* **50,** no. 2, 379–402, doi: 10.1023/A:1003403601725.

Papadopoulos, G. A., M. Charalampakis, A. Fokaefs, and G. Minadakis (2010). Strong foreshock signal preceding the L'Aquila (Italy) earthquake (Mw 6.3) of 6 April 2009, *Nat. Hazards Earth Syst. Sci.* **10,** 19–24.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, K. S. Rukstales, D. E. McNamara, R. A. Williams, A. M. Shumway, P. M. Powers, P. S. Earle, *et al.* (2018). 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes, *Seismol. Res. Lett.* **89,** no. 3, 1049–1061.

Petruccelli, A., P. Gasperini, T. Tormann, D. Schorlemmer, A. P. Rinaldi, G. Vannucci, and S. Wiemer (2019). Simultaneous dependence of the earthquake-size distribution on faulting style and depth, *Geophys. Res. Lett.* **46,** no. 20, 11,044–11,053, doi: 10.1029/2019GL083997.

Petruccelli, A., D. Schorlemmer, T. Tormann, A. P. Rinaldi, S. Wiemer, P. Gasperini, and G. Vannucci (2019). The influence of faulting style on the size-distribution of global earthquakes, *Earth Planet. Sci. Lett.* **527,** doi: 10.1016/j.epsl.2019.115791.

Petruccelli, A., G. Vannucci, B. Lolli, and P. Gasperini (2018). Harmonic fluctuation of the slope of the frequency - magnitude distribution (b-value) as a function of the angle of rake, *Bull. Seismol. Soc. Am.* **108,** 1864–1876, doi: 10.1785/0120170328.

Roberts, N. S., A. F. Bell, and I. G. Main (2015). Are volcanic seismic *b*-values high, and if so when? *J. Volcanol. Geoth. Res.* **308,** 127–141, ISSN 0377-0273, doi: 10.1016/j.jvolgeores.2015.10.021.

Sanchidrián, J. A., P. Segarra, and L. M. Lopez (2007). Energy components in rock blasting, *Int. J. Rock Mech. Min. Sci.* **44,** 130–147.

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. Seismol. Soc. Am.* **58,** 399–415.

Schorlemmer, D., and S. Wiemer (2005). Microseismicity data forecast rupture area, *Nature* **434,** 1086, doi: 10.1038/4341086a.

Schorlemmer, D., S. Wiemer, and M. Wyss (2005). Variations in earthquake-size distribution across different stress regimes, *Nature* **437,** no. 7058, 539–542, doi: 10.1038/nature04094.

Schurr, B., G. Asch, S. Hainzl, J. Bedford, A. Hoechner, M. Palo, R. Wang, M. Moreno, M. Bartsch, Y. Zhang, *et al.* (2014). Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake, *Nature* **512,** 299–302, doi: 10.1038/nature13681.

Stump, B. W., M. A. H. Hedlin, D. C. Pearson, and V. Hsu (2002). Characterization of mining explosions at regional distances: Implications with the International monitoring system, *Rev. Geophys.* **40,** no. 4, 2–21, doi: 10.1029/1998RG000048.

Tormann, T., B. Enescu, J. Woessner, and S. Wiemer (2015). Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake, *Nat. Geosci.* **8,** 152–158, doi: 10.1038/ngeo2343.

van Stiphout, T., S. Wiemer, and W. Marzocchi (2010). Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake, *Geophys. Res. Lett.* **37,** L06306, doi: 10.1029/2009GL042352.

Wiemer, S., and M. Baer (2000). Mapping and removing quarry blast events from seismicity catalogs, *Bull. Seismol. Soc. Am.* **90,** no. 2, 525–530.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western US and Japan, *Bull. Seismol. Soc. Am.* **90,** 859–869.

Wiemer, S., G. Giardini, D. Fäh, N. Deichmann, and S. Sellami (2009). Probabilistic seismic hazard assessment of Switzerland: Best estimates and uncertainties, *J. Seismol.* **13,** no. 4, 449.

Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** 684–698.

Wyss, M. (1973). Towards a physical understanding of the earthquake frequency distribution, *Geophys. J. Int.* **31,** no. 4, 341–359, doi: 10.1111/j.1365-246X.1973.tb06506.x.

Wyss, M., F. Klein, K. Nagamine, and S. Wiemer (2001). Anomalously high b-values in the South Flank of Kilauea volcano, Hawaii: Evidence for the distribution of magma below Kilauea's East Rift Zone, *J. Volcanol. Geoth. Res.* **106,** nos. 1/2, 23–37, ISSN 0377-0273, doi: 10.1016/S0377-0273(00)00263-8.

Wyss, M., K. Shimazaki, and S. Wiemer (1997). Mapping active magma chambers by *b* values beneath the off-Ito volcano, Japan, *J. Geophys. Res.* **102,** no. B9, 20,413–20,422, doi: 10.1029/97JB01074.

Manuscript received 25 March 2021
Published online 30 June 2021

Comment on “Two Foreshock Sequences Post Gulia and Wiemer (2019)” by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky

Laura Gulia*[1] and Stefan Wiemer[2]

Abstract

Dascher-Cousineau *et al.* (2020) apply the so-called foreshock traffic-light system (FTLS) model proposed by Gulia and Wiemer (2019) to two earthquake sequences that occurred after the submission of the model: the 2019 Ridgecrest ($\boldsymbol{M}_{\mathrm{w}}$ 7.1) and the 2020 $\boldsymbol{M}_{\mathrm{w}}$ 6.4 Puerto Rico earthquakes. We show in this comment that the method applied by Kelian Dascher-Cousineau *et al.* (2020) deviates in at least six substantial and not well-documented aspects from the original FTLS method. As a consequence, they used for example in the Ridgecrest case only 1% of the data available to estimate ***b***-values and from a small subvolume of the relevant mainshock fault. In the Puerto Rico case, we document here substantial issues with the homogeneity of the magnitude scale that in our assessment make a meaningful analysis of ***b***-values impossible. We conclude that the evaluation by Dascher-Cousineau *et al.* (2020) is misrepresentative and a not a fair test of the FTLS hypothesis.

Introduction and Context

Dascher-Cousineau *et al.* (2020, henceforth DC2020) applies the so-called foreshock traffic-light system (FTLS) model proposed by Gulia and Wiemer (2019, henceforth GW2019) to two earthquake sequences that occurred after the submission of the model: the 2019 M_{w} 7.1 Ridgecrest and the 2020 M_{w} 6.4 Puerto Rico earthquakes. We appreciate that DC2020 decided to evaluate our model and hypothesis pseudoprospectively on independent data and partially with their own code implementation. This is exactly how science needs to work: hypotheses proposed by one group need to be evaluated independently by others. For this reason, we also provided as part of GW2019 the source code used for the analysis. However, in our assessment documented here, the study by DC2020 contains substantial deviations from the originally proposed method, including demonstratable errors, which then lead the authors to partially incorrect conclusions.

Because DC2020 did not provide their source code nor their datasets as part of the publication, we requested them directly from the authors, who kindly supplied them for the Ridgecrest case. This comment addresses the deviations introduced by DC2020 in their study for each of the two mainshocks individually and draws some common conclusions.

Ridgecrest Case Study

For the first sequence (Ridgecrest), the analysis by DC2020 resulted in a red FTLS alert after the M_{w} 6.4 event and in an orange alert after the M_{w} 7.1 event. Meanwhile, in the same SRL issue, Gulia *et al.* (2020) published their own pseudoprospective assessment of this sequence, also reporting a red FTLS alert after the M_{w} 6.4 but a green alert following the M_{w} 7.1 two days later. The observed differences between these two articles in the FTLS setting and in the underlying b-value time series are a direct consequence of the substantial deviations from the GW2019 approach as implemented by DC2020. Here we document these deviations in methodology introduced by DC2020 step by step.

Correctly establish the reference *b*-value for the first mainshock

A critically important parameter to be established in the FTLS model is the local reference b-value because the FTLS decisions are based on the difference in percent between the sequence-specific b-values and the reference b-value. According to the GW2019 hypothesis, it is important to establish the reference b-value such that (1) it is only based on earthquake immediately near the initiating mainshock fault (i.e., within 3 km of the fault) because b-values vary substantially with space, and (2) it uses a long time series, to have the statically most robust estimate that averages over temporal variations. For the background of Californian sequences in Gulia *et al.* (2018), we start our analysis from 1981, when the network was greatly improved (e.g., Tormann *et al.*, 2014). Therefore, for our Ridgecrest analysis

1. Department of Physics and Astronomy, University of Bologna, Bologna, Italy, https://orcid.org/0000-0002-1474-230X (LG); 2. Swiss Seismological Service, ETH Zurich, Zurich, Switzerland, https://orcid.org/0000-0002-4919-3283 (SW)

*Corresponding author: laura.gulia@unibo.it

presented by Gulia *et al.* (2020), we use a 39 yr long background catalog. DC2020 started their analysis only in the year 2000, resulting in a factor of two reduction of the data used. This choice was made to avoid the influence of the 1992 M_w 7.3 Landers and 1999 M_w 7.1 Hector Mine aftershocks (Kelian Dascher-Cousineau, personal comm., 2020). However, these two sequences occurred, ~200 and 170 km from the Ridgecrest mainshock, respectively, distances well beyond the Gardner and Knopoff (1974) radii of influence for both magnitudes and neither of these mainshocks, had a noticeable impact on earthquake rates in the Ridgecrest area. We thus consider 1981 the better justified starting date, but this choice is indeed a subjective one and not fully automate in the approach yet.

Deviation 1.1: DC2020 use a catalog from the year 2000, although we advise (and do so in Gulia *et al.*, 2018, 2020) to use data from 1981.

Estimating reliable b-values also requires a robust, automated estimation of the magnitude of completeness. In GW2019, we use the so-called maximum curvature method by Woessner and Wiemer (2005) and apply it as suggested in their article: we first cut for robustness the catalog close to the overall catalog completeness and then re-estimate M_c for each timestep. We differentiate in purpose between the background b-values estimation, in which we apply as an overall M_c cut (M_c_maxCurve_overall—0.2), and aftershock sequence that are both data rich and have strongly varying M_c with time, in which we apply as on overall M_c cut of (M_c_maxCurve_overall, thus 0.2 higher). In both cases, we then re-estimate M_c in each time bin using M_c_maxCurv + 0.2. This procedure has been documented in the article and in detail in the source code.

DC2020 argued that the M_c pre-cutting approach outlined earlier is actually an error in our code that they detected (which it is not) and modified it such that they added an additional M_c increment I of +0.2 for also estimating the background b-values.

Deviation 1.2: DC2020 apply erroneously a "safety" M_c increment of +0.4 rather than +0.2 for the background b-value calculations.

These two deviations from our published method decrease the number of events available to establish the reference b-value with 3 km of the fault plane by 93%, from 1154 to 89, which then is well below the critical threshold of 250 events defined as a quality criterion in GW2019. Therefore, DC2020 select events in a circle around the M_w 6.4 epicenter (the alternative method used by GW2019 for inferior datasets) instead of along the actual fault plane.

Deviation 1.3: To establish the reference b-value, DC2020 sample events in circular region of ~10 km around the epicenter, but Gulia *et al.* (2020) use events in a box within 3 km of the rupture plane.

The combined impact of these three deviations is illustrated in Figure 1. Figure 1a,b shows the fault-plane projection of the M_w 6.4 event (black grid), superimposed is the catalog used by DC2020 to establish the background b-value (red dots). It is composed of the 250 events nearest the mainshock since 2000, events up to ~10 km from the epicenter. Shown in comparison is the dataset used by Gulia *et al.* (2020, blue dots), composed of events with a maximum distance of 3 km form the fault plane. DC2020 also used shallow events that are >3 km from the fault plane and thus not included in the GW2019 approach. As a consequence of these differences, the background b-value in DC2020 is 0.90 based on about seven events above completeness per year and averaged over 19 yr. Using the GW2019 approach, we compute $b = 0.97$ based on about 22 events per year, averaged over 39 yr.

Correctly selecting the mainshock fault plane and events between the mainshock

Among the two nodal planes defined by the focal mechanism, GW2019 proposed to use the one with the highest number of immediate aftershocks within 3 km of the fault because the method needs to run fully automatically and in near-real time. For the 31 sequences analyzed by Gulia *et al.* (2018) as well as for the three sequences analyzed in GW2019, we determined the mainshock plane based on the first 24 hr of aftershock data. This is a commonly used time interval sufficiently long to allow for stable detection of the active fault in most cases (see also Kanamori, 1977); however, we did not explicitly document this choice in GW2019. DC2020 decided to use a much shorter time interval of only 1 hr to establish the mainshock fault, resulting as explained later in the choice of the alternative fault plane.

The initial M_w 6.4 Ridgecrest mainshock was a complex rupture, and it took several days before geodetic, seismic, and relocated seismicity data provided a reliable view of this complex sequence. Ross *et al.* (2019) identified three simultaneous subevents and hypothesized that the rupture had been a cascading phenomenon. The purely statistical method used in GW2019 based on the first 24 hr of aftershocks selected the northwest-trending fault plane that represented the initial rupture (Fig. 1d, blue symbols). DC2020, on the contrary, selected the orthogonal plane (Fig. 1d, red dots). Given the complex rupture pattern, both choices are actually defendable. Deviations 1.1 and 1.2 apply on top for this part of the analysis. In addition, DC2020 did not limit the depth of selected events.

Deviation 2.1: DC2020 selected aftershock of the first hour rather than the first 24 hr to define the active fault. They thus selected the alternative fault plane for estimating the b-values of the aftershocks after the first mainshock.

Deviation 2.2: DC2020 do not limit the analysis to events with 3 km depth below and above the fault plane but extend the sampling down to 20 km.

As a consequence of these deviations, DC2020 compute on the alternative nodal plane a b-value for all the in-between events of $b = 0.83$, whereas Gulia *et al.* (2020) compute $b = 0.74$, based on a much larger data sets because of the lower M_c (Fig. 1e). Despite these five deviations, the overall result of

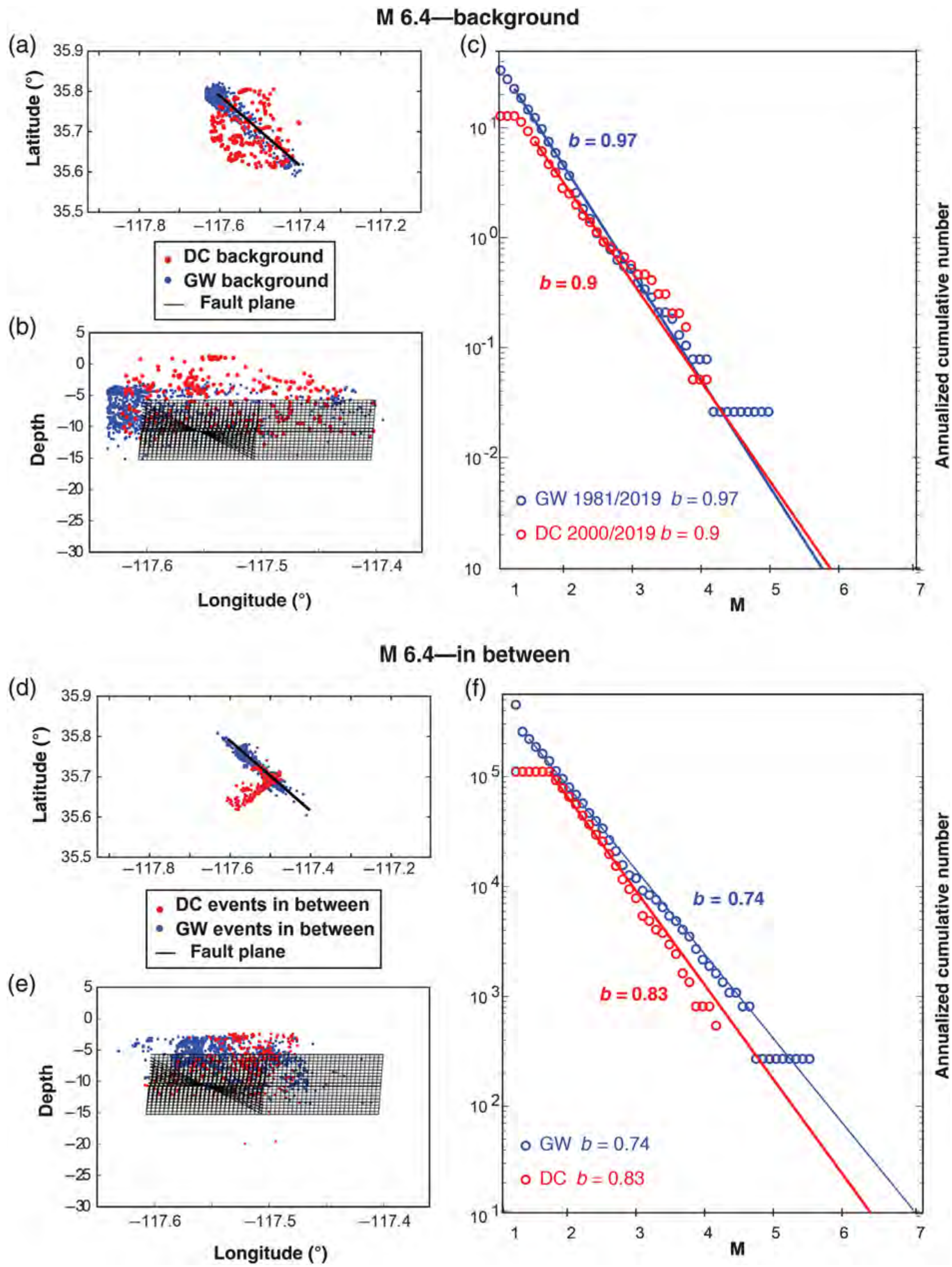


Figure 1. (a,b) Map of Ridgecrest region. Shown are the selected mainshock plane of the 4 July 2019 M_w 6.4 mainshock (black grid) and the selected background seismicity by Dascher-Cousineau *et al.* (2020, DC2020) (red dots) and by GW2020 (blue dots). (c) Annualized frequency–magnitude distribution (FMD) for the two datasets shown in panels (a,b). (d,e) Map of Ridgecrest region. Shown are the selected mainshock plane of the 4 July 2019 M_w 6.4 mainshock (black grid) and the selected "in-between" events by DC2020 (red dots) and by GW2020 (blue dots). (f) Annualized FMD for the two datasets show in panels (d,e).

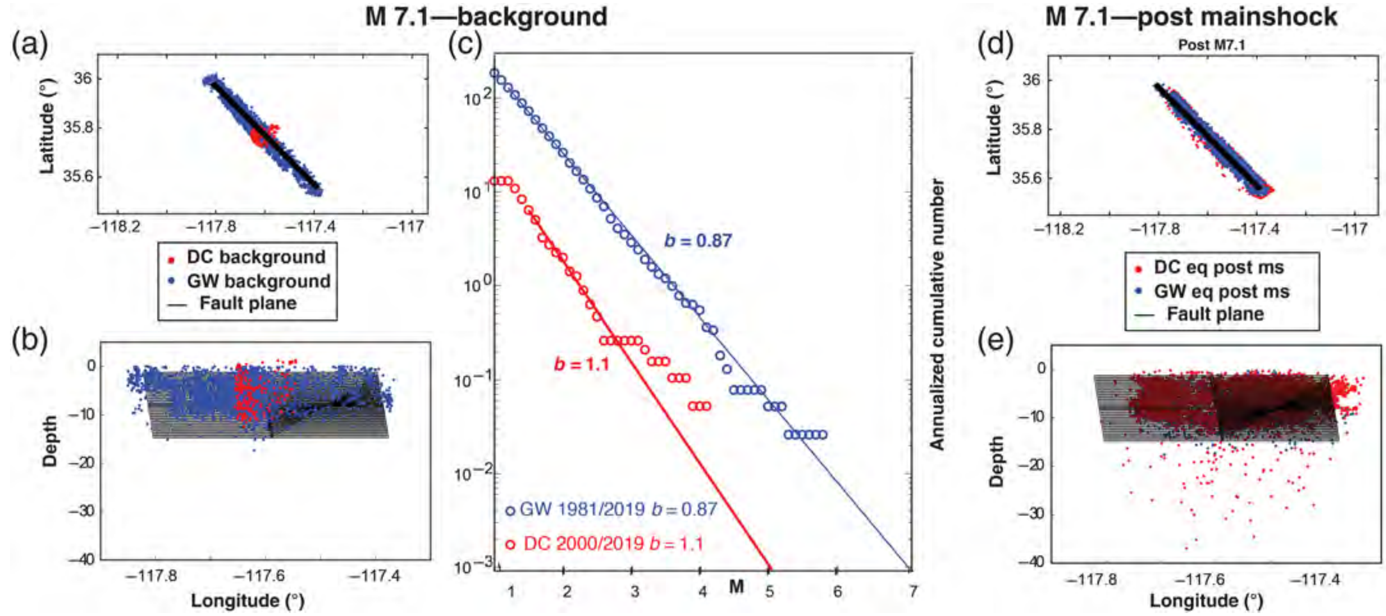


Figure 2. (a,b) Seismicity maps showing the fault plane (in black) and the events preceding the 6 July 2019 M_w 7.1 event selected by DC2020 (red dots) and by GW2020 (blue dots). (c) The relative FMD for the two datasets in panels (a,b). (d,e) Seismicity maps showing the fault plane (in black) and the events following the 2019 M_w 7.1 event selected by DC2020 (red dots) and by GW2020 (blue dots).

the FTLS assessment given by DC2020 remains unchanged: a red FTLS setting.

Correctly selecting the second mainshock fault plane, the new reference background, and aftershocks

According to the FTLS model, after a second and larger mainshock occurs as part of a sequence, the FTLS assessment process restarts: first, the new fault plane is determined based on the seismicity within 24 hr of this mainshock. Next, the background b-value is redetermined based on events within 3 km of this longer fault plane and then compared with the b-values of the aftershocks near the new fault to estimate the new FTLS status. This is typically the most data-rich part of the analysis because it involves larger fault areas and numerous aftershocks. Here, DC2020 also apply the deviations D1.1 (start date, 2000), D1.2 (M_c double counted for the background), and D1.3 (circular sampling instead of along the fault plane), but the resulting impact is much bigger because the M_w 7.1 fault is considerably longer.

Deviation 3.1: To establish the reference b-value for the M_w 7.1 fault, DC2020 sample events in circular region of ~3 km around the epicenter, but Gulia *et al.* (2020) use events in a box within 3 km of the about 60-km-long rupture plane.

As shown in Figure 2a, DC2020 select events that only cover a small subset of the fault, ~10%; added to this is the higher M_c and shorter catalog duration. The background b-value estimation of DC2020 thus is based only on about 1% of the data used by the GW2019 approach (Fig. 3c). As a consequence, the background b-value for the second event established by DC2020 is not unexpectedly very different from the one the GW2019 approach will compute (Fig. 2c): DC2020 estimate $b = 1.10$, and Gulia *et al.* (2020) estimate $b = 0.87$. The frequency–magnitude distribution (FMD) of DC2020, being based on a small data set, shows a substantial break in slope around magnitude 3 (Fig. 2c, red symbols). This difference in background b-value then results in very different changes in percent compared with the aftershock b-values and ultimately results in the difference in the FTLS setting observed between DC2020 and Gulia *et al.* (2020).

For the computation of the b-values of the aftershocks, DC2020 then correctly use events within 3 km of the mainshock fault (Fig. 2d,e), although deviations 1.2 and 2.2 still apply. However, although the absolute aftershock b-values are quite similar between the two articles, the all-important changes in percent normalized to the background b-values are very different (−10% for DC2020 à orange alert; +26% for Gulia *et al.*, 2020 à green alert), largely because of the different background b-values that they are normalized to ($b = 1.10$ vs. $b = 0.87$).

Because there are at least six substantial deviations from the GW2019 approach, it is no surprise that Gulia *et al.* (2020) report quite different results from DC2020 for the Ridgecrest sequence. We will discuss the appropriateness of these deviations and the meaningfulness of the comparison, given these deviations in the Discussions and Conclusions section.

Puerto Rico Case Study

The second case study discussed by DC2020 is the 7 January 2020 Puerto Rico event: DC2020 reported a red alert after the

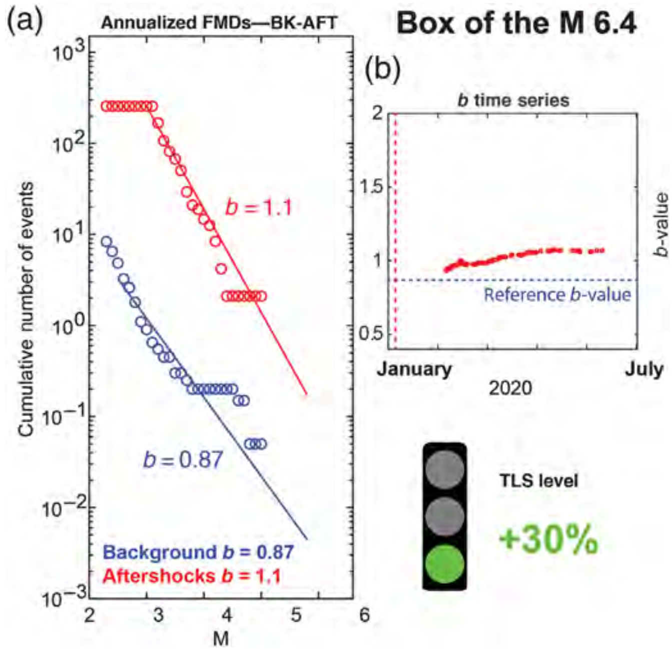


Figure 3. (a,b) Performance of the foreshock traffic-light system (FTLS) for the M_w 6.4 Puerto Rico event. (a) FMDs for the source of the M_w 6.4 event for two time periods: background in blue and maximum *b*-value reached in the first weeks of aftershocks. (b) *b*-value time series for the M_w 6.4. The blue-dashed line is the reference *b*-value, and the red-dashed vertical line indicates the time of the M_w 6.4 event. All the estimates are above the reference value. BK-AFT, Background-Aftershocks; TLS, terrestrial laser scanning.

mainshocks, indicating an upcoming larger event, which has not yet occurred on 24 February 2020 (and not until 19 December 2020), thus suggesting a false positive for the FTLS evaluation. As DC2020 themselves state, this case is not an actual test of the GW2019 hypothesis:

> *For the source region surrounding this event used for computing a b-value, we relax the nominal spatial window of 3 km from the source to 10 km to determine stable b-values. For this reason, the time series produced for the M_w 5.0 foreshock is not a strict test of the method proposed by Gulia and Wiemer (2019) but is nonetheless inter- interesting to consider.*

We add to this statement:

1. in, we explicitly exclude from the test offshore sequences because hypocenter accuracy and completeness are inevitably much inferior. In our assessment, the quality of offshore catalogs is typically too low to allow to select enough earthquakes near the rupture plane and with sufficient confidence.
2. DC2020 performed a time series on an M_w 5, a much lower magnitude than the minimum one (M_w 6) required for the model of GW2019. Because stress changes scale with magnitude, we have argued in Gulia *et al* (2018) that to apply the method to smaller magnitudes, only events close by should be considered (e.g., within 1 km of an M_w 5 earthquake).

Even though DC2020 in the Puerto Rico study did not test the GW2019 hypothesis in the first place, we also like to point out that their analysis is, in our opinion, flawed or biased. Data-quality issues related to the homogeneity of the estimate magnitudes across the magnitude scale were not considered, leading to arbitrary estimates of *b*-values, as explained next.

In a first step, we evaluated the FTLS method on the M_w 6.4 mainshock using the original published and unchanged method and selection criteria by GW2019 and the same catalog of the Puerto Rico National Seismic Network used also by DC2020 (although we could not check if it had been updated in between downloads). We select events within a 3 km distance from the fault plane of the M_w 6.4 event, applying a preliminary magnitude cutoff at a minimum level of completeness (here 2.3 + 0.2 correction factor). The results obtained without any modifications in the released code are shown in Figure 3. In our analysis of the M_w 6.4 event, the *b*-value increases by 30% after the mainshock, resulting in a green alert. The Puerto Rico sequence would thus represent an additional and further positive test of the GW2019 hypothesis; however, as explained later, the quality check applied in GW2019 estimates the FMDs not reliable enough to consider this a successful case study.

The challenge with magnitude-scale reporting homogeneity of the Puerto Rico catalog is illustrated in Figure 4, in which we show the overall *b*-value of earthquakes within ~50 km from the island of Puerto Rico for the period 2003–2019, plotted as a function of cutoff magnitude (red curve). This kind of plot is a simple check for both M_c and the homogeneity of reporting (Wiemer and Wyss, 2000; Woessner and Wiemer, 2005). The expected behavior is that the *b*-value is strongly underestimated as long as the catalog is incomplete, and when M_c is approached, the *b*-value levels of and a plateau emerge. The plots for the Puerto Rico catalog reveal no such plateau (the ones for Ridgecrest, e.g., do). Instead, it signals a very high sensitivity of *b*-values to the choice of M_c, with *b*-values ranging from <1.0 to 1.6, depending on the choice of M_c. Similar behavior is found for the M_w 6.4 mainshock region analyzing the 2020 data only (blue line in Fig. 4). Such a peak rather than a plateau is indicative of an upward bend of the FMD, typical, for example, if different procedures are used to estimate magnitudes in different magnitude bins.

The impact of this magnitude-scale compression on the FMD near the mainshocks is shown in Figure 4b. We selected events within 3 km distance of the M_w 6.4 mainshock fault. The resulting FMD does not only look nonlinear to the eye, but it also does not pass the nonlinearity filter (Tormann *et al.*, 2014) that we apply as a quality check in GW2019 to ensure

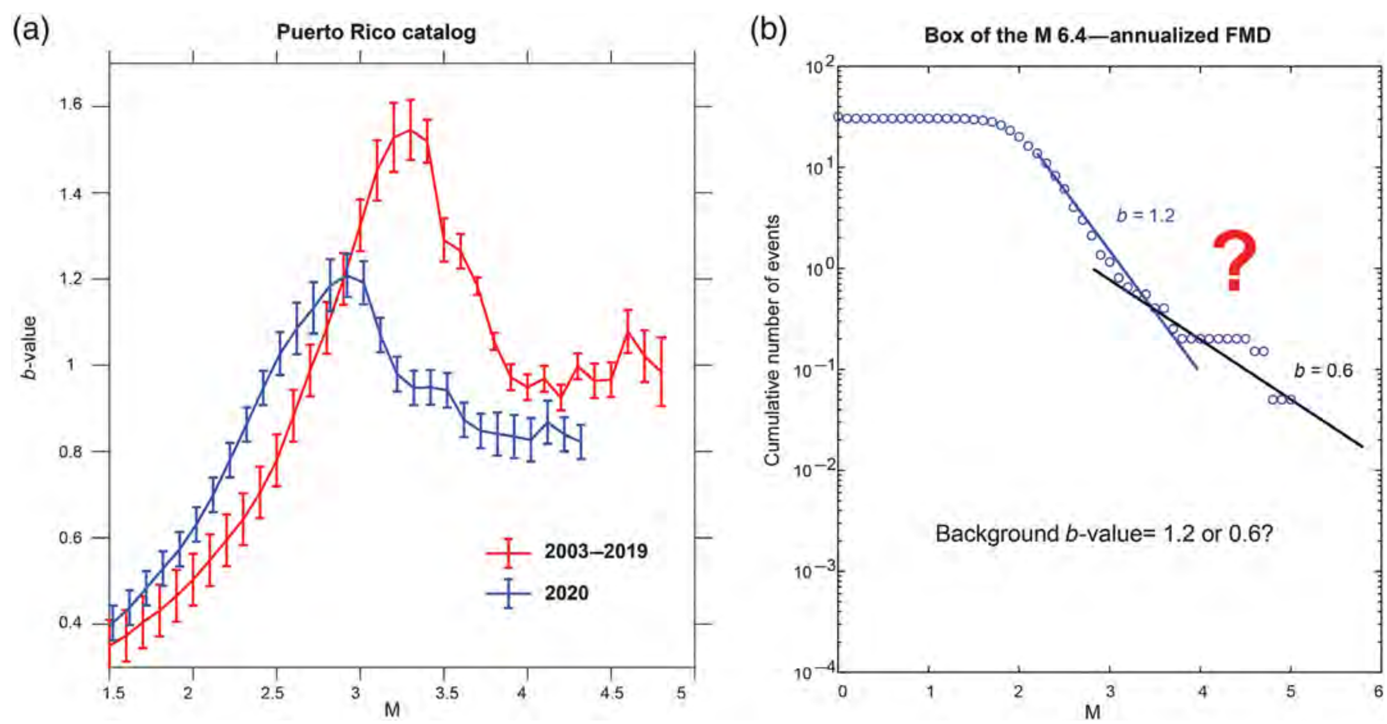


Figure 4. (a) b-value as a function of magnitude of completeness for the Puerto Rico catalog, for the periods 2003–2019 (red) and 2020 (blue). (b) Annualized FMD of the background (blue circles) at two different magnitude of completeness and relative b-values for the M_w 6.4 event dataset.

compliance with a linear power-law model. The substantial "kink" in the distribution around magnitude 3.0–3.5 leads to the aforementioned strong sensitivity of the background b-value on the choice of M_c. Figure 4 implies that a stable b-value analysis may only be possible from about magnitude 4.0, but then almost no data would be left for analysis.

The main difference between our analysis and the one by DC2020 lies in the b-values of the aftershock sequence, and it is ultimately related to the aforementioned data-quality issue. For the background, DC2020 compute a rather high b-value ($b = 1.2$) compared with our analysis ($b = 0.87$). This is not only a consequence of different sampling volumes (large circles vs. fault plane) but also a result of the "upward" bend of the FMD for events below magnitude 3. DC2020 use a much lower M_c here (about 2.0); we would use $M_c = 2.5$. DC2020 then compute an aftershock b-value of ~0.5–0.6 (their fig. 3). Our analysis, shown in Figure 3, results in $b = 1.1$. We cannot fully explain how DC2020 obtain such an unusual low b-value, and we note that their FMD does not fit the data for most of the range—too low for small magnitudes yet too high for larger ones (Fig. 4b).

We recognize that the dependence on the two free parameters of our analysis, the no-alert time and the magnitude of completeness, is potentially creating an arbitrariness in the analysis. To address this limitation, we introduced in Gulia *et al.* (2020) a systematic scan of the free parameter space to assess the robustness of the analysis. We repeated the analysis for the Puerto Rico case. If the M_c of the aftershocks is below completeness, then b-values are much too low, and an erroneous red alert is found. When M_c is high enough, and for all possible constellations of M_c and no-alert time, a green alert after the M_w 6.4 results.

Discussions and Conclusions

Testing earthquake forecasts in rigorous ways is highly important and the past 40 yr of research have seen a rather spotty record of the seismology community on testing (Jordan, 2006; Jackson, 1996; Kagan, 1999; Zechar *et al.*, 2016). One of the challenges is that often the models are a moving target. There is a broad consensus in the community (e.g., Jordan, 2006; Zechar *et al.*, 2011; Marzocchi *et al.*, 2015; Strader *et al.*, 2017; Schorlemmer *et al.*, 2018) that prospective and pseudoprospective testing in earthquake sciences (no different from medicine or other sciences) must follow strict rules, and community efforts such as Collaboratory for the Study of Earthquake Predictability (CSEP) have been created for this purpose (e.g., Gerstenberger and Rhoades, 2010; Werner *et al.*, 2010; Zechar *et al.*, 2010, 2013; Tsuruoka *et al.*, 2012).

One of the most fundamental rules for evaluating hypotheses in science is that the hypothesis to be tested cannot be changed arbitrarily; otherwise, biases (in favor or against a hypothesis) are likely to influence the test and endless discussion may occur. Another basic rule of science is the fact that quality limitation of the data must be accepted and respected even if we do not like them. Otherwise, the "garbage in, garbage out" criteria will almost inevitably apply.

The study by DC2020 has violated these two basic rules of hypothesis testing in several respects, biasing their analysis against the GW2019 hypothesis. We demonstrate here and in Gulia *et al.* (2020) that, when applied correctly, the Ridgecrest cases study would be fully in line with the FTLS hypothesis. DC2020 have deviated in at least six steps from the analysis; these are in parts major deviations, changing by 99% the data to be analyzed. As much as we appreciate that DC2020 evaluated our hypothesis, in our opinion, this test is meaningless or misleading because the method and data processing of DC2020 are substantially different. DC2020, therefore, test their own hypothesis, not ours, but they do not state so in their article. We consider this confusing for the community, and we even pointed out some of these shortcomings to the authors in a review before publication.

One might argue that a method or hypothesis should be robust enough to also work with somewhat modified parameters as a measure of robustness, a point raised by DC2020. We respond first of all that even before such a useful sensitivity analysis, one obviously needs to test the actual unmodified hypothesis and also document the changes transparently. However, much more important is in our view that the deviations applied by DC2020 are unjustified in several ways:

- The deviations violate the *physical framework* of GW2019. we consider it critically important and physically plausible to sample events in the immediate vicinity of the actual fault plane because stress changes caused by the mainshock are strongest here.
- The deviations violate the *statical framework* of GW2019 that aims to maximize the amount of data and hence robustness of the analysis. Instead, DC2020 use only a small fraction of the data available for no apparent reason (Figs. 1 and 2).
- The deviations violate the *principle of reproducibility* because they are not documented and possibly not intended modifications (e.g., all depth selected, M_c add on double counted).

The Puerto Rico case is more complex to interpret. Both groups agree that this case study does not represent a test of the GW2019 hypothesis in the first place. However, in addition, here DC2020 introduced inconsistencies in the analysis, in part possibly because of the same deviations stated here for Ridgecrest but even more so by ignoring the limitations of the data as well as the minimum required magnitude (M_w 6) to implement GW2019. Redoing the analysis using the original GW2019 approach with no modifications, we find also that this case would support the FTLS hypothesis (Fig. 3). Nevertheless, we argue that the offshore data quality is too poor, and the magnitude scale shows unexplained bends (Fig. 4) to allow for robust analysis. The automatic procedures for quality control in GW2019 would reject this case also.

Every forecast model has several free parameters. Some are obvious, first-order free parameters, such as the sample sizes used or the width of the volume sampled; these can be readily analyzed in a sensitivity analysis. Some are related to the automated quality analysis, such as the determination of M_c, and here the uncertainty in M_c determination can be used to estimate sensitivity. A third set of "free" parameters are resulting from expert choices based, for example, on data quality, such as the start time of the catalog or the fault plane used. In Gulia *et al.* (2020), we explore some of the free parameter space, confirming the robustness of the FTLS model to first-order free parameters, but a complete search of the free parameter space is difficult. It would require a logic-tree approach such as the ones used in probabilistic seismic hazard assessment, capturing aleatory and epistemic uncertainties. In forecast, the preferred method instead is to perform fully prospective test of models under controlled conditions and against predefined, authoritative data sources (e.g., Schorlemmer *et al.*, 2018). The GW2019 hypothesis may well fail such a test, but it deserves to be tested fairly. DC2020 did in our assessment, unfortunately, not conduct such a fair test of the actual hypothesis, nor did they perform a systematic sensitivity test.

Final Comment after Reading the Reply by the Authors

We carefully read the reply by Dascher-Cousineau *et al.* (2021) to our comment and thank the authors for the detailed discussion as well as the clarifications and corrections applied to their analysis. We still believe that all deviations we listed in our comment are correctly identified and justified. The aim of Gulia *et al.* (2020) was to implement the published FTLS *without any modifications*, and this is what we did.

As stated before in our comment, we welcome the independent evaluation of the FTLS by DC2020 and welcome their response to our criticisms raised. Details matter in science, and we are struck again how difficult it is in earthquake forecasting to not only ensure full reproducibility but to also write down a "recipe" that other qualified scientists can apply to new cases and reach the same conclusions. Cooking is a good analog: even a detailed recipe will not ensure the same outcome. For evaluating earthquake forecasting–related hypotheses, our experience documented in the article and replies also highlight the need for a collaborative and fully prospective testing environment such as the one provided by CSEP, with community-agreed rules and decoupling between modelers and evaluators.

Data and Resources

Data about Ridgecrest events are available from the Advanced National Seismic System (ANSS) Comprehensive Earthquake Catalog (ComCat) and Shelly (2020, SRL), and data about Puerto Rico events are available from the Puerto Rico Seismic Network. Data about European Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project are available at www.rise-eu.org. Both figures and calculations were performed by MATLAB (www.mathworks.com/products/matlab). All websites were last accessed in October 2019.

Declaration of Competing Interests

The authors declare no competing interests

Acknowledgments

The authors thank Editor-in-Chief Allison Bent, Associated Editor Anastasia Pratt, and the anonymous reviewer. This study was supported by the Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115.

References

Dascher-Cousineau, K., T. Lay, and E. E. Brodsky (2020). Two foreshock sequences post Gulia and Wiemer (2019), *Seismol. Res. Lett.* **91,** no. 5, 2843–2850, doi: 10.1785/0220200082.

Dascher-Cousineau, K., T. Lay, and E. E. Brodsky (2021). Reply to "Comment on 'Two Foreshock Sequences Post Gulia and Wiemer (2019)' by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky" by Laura Gulia and Stefan Wiemer, *Seismol. Res. Lett.* doi: 10.1785/0220210059.

Gardner, J. K., and L. Knopoff (1974). Is the sequence of earthquake in southern California, with aftershocks removed Poissonian? *Bull. Seismol. Soc. Am.* **66,** 1271–1302.

Gerstenberger, M. C., and D. A. Rhoades (2010). New Zealand Earthquake Forecast Testing Centre, *Pure Appl. Geophys.* **167,** nos. 8/9, 877–892, doi: 10.1007/s00024-010-0082-4.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** 193–199.

Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). The effect of a mainshock on the size distribution of the aftershocks, *Geophys. Res. Lett.* **45,** no. B1, 13,277–13,287, doi: 10.1029/2018GL080619.

Gulia, L., S. Wiemer, and G. Vannucci (2020). Pseudoprospective evaluation of the foreshock traffic-light system in Ridgecrest and implications for aftershock hazard assessment, *Seismol. Res. Lett.* **91,** no. 5, 2828–2842, doi: 10.1785/0220190307.

Jackson, D. D. (1996). Hypothesis testing and earthquake prediction, *Proc. Natl. Acad. Sci. Unit. States Am.* **93,** 3772–3775.

Jordan, T. H. (2006). Earthquake predictability, brick by brick, *Seismol. Res. Lett.* **77,** no. 1, 3–6.

Kagan, Y. Y. (1999). Is earthquake seismology a hard, quantitative science? *Pure Appl. Geophys.* **155,** nos. 2/4, 233–258.

Kanamori, H. (1977). The energy release in great earthquakes, *J. Geophys. Res.* **82,** no. 20, 2981–2987.

Marzocchi, W., T. H. Jordan, and G. Woo (2015). Operational earthquake forecasting and decision making, *Ann. Geophys.* **58,** no. 4, doi: 10.1007/978-3-642-12233-0_18.

Ross, Z. E., B. Idini, Z. Jia, O. L. Stephenson, M. Zhong, X. Wang, Z. Zhan, M. Simons, E. J. Fielding, S. Yun, *et al.* (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science* **366,** 346–351, doi: 10.1126/science.aaz0109.

Schorlemmer, D., M. J. Werner, W. Marzocchi, T. H. Jordan, Y. Ogata, D. D. Jackson, S. Mak, D. A. Rhoades, M. C. Gerstenberger, N. Hirata, *et al.* (2018). The Collaboratory for the Study of Earthquake Predictability: Achievements and priorities, *Seismol. Res. Lett.* **89,** no. 4, 1305–1313, doi: 10.1785/0220180053.

Shelly, D. R. (2020). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity, *Seismol. Res. Lett.* **91,** 1971–1978, doi: 10.1785/0220190309.

Strader, A., M. Schneider, and D. Schorlemmer (2017). Prospective and retrospective evaluation of five-year earthquake forecast models for California, *Geophys. J. Int.* **211,** no. 1, 239–251, doi: 10.1093/gji/ggx268.

Tormann, T., S. Wiemer, and A. Mignan (2014). Systematic survey of high-resolution *b*value imaging along Californian faults: Inference on asperities, *J. Geophys. Res.* **119,** 2029–2054, doi: 10.1002/2013JB010867.

Tsuruoka, H., N. Hirata, D. Schorlemmer, F. Euchner, K. Z. Nanjo, and T. H. Jordan (2012). CSEP testing center and the first results of the earthquake forecast testing experiment in Japan, *Earth Planet Space* **64,** 661–671, doi: 10.5047/eps.2012.06.007.

Werner, M. J., J. D. Zechar, W. Marzocchi, S. Wiemer, and C. S.-I. W. Grp (2010). Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, *Ann. Geophys.* **53,** no. 3, 11–30, doi: 10.4401/ag-4840.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan, *Bull. Seismol. Soc. Am.* **90,** no. 4, 859–869, doi: 10.1785/0119990114.

Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** 684–698.

Zechar, J. D., J. L. Hardebeck, A. J. Michael, M. Naylor, S. Steacy, S. Wiemer, J. C. Zhuang, and C. W. Grp (2011). Community Online Resource for Statistical Seismicity Analysis, *Seismol. Res. Lett.* **82,** no. 5, 686–690, doi: 10.1785/gssrl.82.5.686.

Zechar, J. D., W. Marzocchi, and S. Wiemer (2016). Operational earthquake forecasting in Europe: Progress, despite challenges, *Bull. Earthq. Eng.* **14,** no. 9, 2459–2469, doi: 10.1007/s10518-016-9930-7.

Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2010). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency and Computation: Practice and Experience* **22,** no. 12, 1836–1847, doi: 10.1002/cpe.1519.

Zechar, J. D., D. Schorlemmer, M. J. Werner, M. C. Gerstenberger, D. A. Rhoades, and T. H. Jordan (2013). Regional earthquake likelihood models I: First-order results, *Bull. Seismol. Soc. Am.* **103,** no. 2a, 787–798, doi: 10.1785/0120120186.

Manuscript received 17 November 2020
Published online 28 July 2021

Pseudoprospective Evaluation of the Foreshock Traffic-Light System in Ridgecrest and Implications for Aftershock Hazard Assessment

Laura Gulia*[1,2], Stefan Wiemer[1], and Gianfranco Vannucci[3]

Abstract

The M_w 7.1 Ridgecrest earthquake sequence in California in July 2019 offered an opportunity to evaluate in near-real time the temporal and spatial variations in the average earthquake size distribution (the *b*-value) and the performance of the newly introduced foreshock traffic-light system. In normally decaying aftershock sequences, in the past studies, the *b*-value of the aftershocks was found, on average, to be 10%–30% higher than the background *b*-value. A drop of 10% or more in "aftershock" *b*-values was postulated to indicate that the region is still highly stressed and that a subsequent larger event is likely. In this Ridgecrest case study, after analyzing the magnitude of completeness of the sequences, we find that the quality of the monitoring network is excellent, which allows us to determine reliable *b*-values over a large range of magnitudes within hours of the two mainshocks. We then find that in the hours after the first M_w 6.4 Ridgecrest event, the *b*-value drops by 23% on average, compared to the background value, triggering a red foreshock traffic light. Spatially mapping the changes in *b* values, we identify an area to the north of the rupture plane as the most likely location of a subsequent event. After the second, magnitude 7.1 mainshock, which did occur in that location as anticipated, the *b*-value increased by 26% over the background value, triggering a green traffic light. Finally, comparing the 2019 sequence with the M_w 5.8 sequence in 1995, in which no mainshock followed, we find a *b*-value increase of 29% after the mainshock. Our results suggest that the real-time monitoring of *b*-values is feasible in California and may add important information for aftershock hazard assessment.

Cite this article as Gulia, L., S. Wiemer, and G. Vannucci (2020). Pseudoprospective Evaluation of the Foreshock Traffic-Light System in Ridgecrest and Implications for Aftershock Hazard Assessment, *Seismol. Res. Lett.* **91**, 2828–2842, doi: 10.1785/0220190307.

Supplemental Material

Introduction

It is well known and almost universally observed that the stress changes caused by a major earthquake strongly affect seismic activity in the vicinity, and the rate of earthquakes increases near the mainshock rupture by several orders of magnitude (Okada, 1992; Stein, 1999; Ebel *et al.*, 2000). In most sequences, on average, this aftershock activity then decays exponentially back to the previous background rate (e.g., Reasenberg and Jones, 1990), a process first described by Omori (1895) and nowadays often described with reference to the concept of epidemic-type aftershock sequences (ETASs; Ogata, 1988). This systematic aftershock behavior can be satisfactorily explained and well modeled using models combining coulomb stress changes and rate and state friction (Dieterich *et al.*, 2000; Toda and Stein, 2003). It also constitutes the baseline of probabilistic assessments of aftershock probabilities (e.g., Reasenberg and Jones, 1990; Marzocchi *et al.*, 2017; Omi *et al.*, 2019). Today, the term operational earthquake forecasting (OEF) is often used when referring to aftershock forecasting in near-real time (Jordan *et al.*, 2014; Zechar *et al.*, 2016).

Far less well established and not currently used in OEF is the fact that the stress redistribution caused by a mainshock also systematically influences relative earthquake size distribution, the *b*-value of the Gutenberg and Richter relationship (Ishimoto and Iida, 1939; Gutenberg and Richter, 1944). Laboratory measurements taken since the 1960s have established that *b*-values are sensitive to stress (Scholz, 1968; Amitrano, 2003; Goebel *et al.*, 2013), and this inverse dependency of *b*-value

1. Swiss Seismological Service, ETH Zurich, Zurich, Switzerland; 2. Now at University of Bologna, Department of Physics and Astronomy, Bologna, Italy; 3. Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

*Corresponding author: lgulia@ethz.ch, laura.gulia@unibo.it

and the applied stress is fully consistent with a number of observed b-value variations with depth, faulting style, and the loading state of faults (e.g., Narteau *et al.*, 2009; Gulia and Wiemer, 2010; Scholz, 2015; Tormann *et al.*, 2015; Petruccelli, Gasperini, *et al.*, 2019; Petruccelli, Schorlemmer, *et al.*, 2019; Staudenmaier *et al.*, 2019;). Mainshock stress changes are therefore expected to systemically change b-values, as suggested by a number of case studies (Wiemer and Katsumata, 1999; Wiemer and Wyss, 2000; Enescu and Ito, 2003). Just recently, Gulia *et al.* (2018) confirmed this hypothesis in a systematic study. To establish generic b-value behaviors in aftershock sequences, they applied a stacking approach to 31 high-quality aftershock sequences from California, Japan, Italy, and Alaska and demonstrated that the b-values of those sequences generically increase by 20% after the mainshock. The higher b-value results suggest a far lower probability of a subsequent large event. Gulia *et al.* (2018) also presented a model based on coulomb stress changes that explains the observations and the observed dependencies on distance, magnitude, and faulting style.

Based on these findings, Gulia and Wiemer (2019) postulated the hypothesis that sequences in which the b-value of the aftershock decreased by 10% or more, instead of increasing as expected, would indicate that a bigger event was not yet to occur. The authors then extended their b-value analysis by successfully testing this hypothesis on three sequences in which a secondary larger mainshock occurred and proposed a foreshock traffic-light system (FTLS), which, taking b-value evolution over time as an indicator of the average stress condition of faults in a region, defines three alert (or concern) levels that can be used to determine in near-real time whether an ongoing sequence is likely. The lowest, "green" alert is triggered by a normally decaying aftershock sequence (b-value increases by 10% or more). The highest, "red" alert indicates a precursory sequence that is more likely to be followed by a larger event (b-value decreases by 10% or more). Sequences falling between these extremes trigger "orange" alerts. Gulia and Wiemer (2019) tested the FTLS on 58 sequences and found it to be more than 95% accurate. Differential b-value maps are proposed as an additional step to estimate the likely location of subsequent larger events. The FTLS is thus proposed as a tool for real-time discrimination between foreshocks and aftershocks, but the authors also point out that additional, ideally fully prospective tests are needed before FTLS can be used in OEF systems.

Key to the robustness of b-value-based forecast is a correct assessment of the completeness of reporting, M_c, for this variable fluctuates dramatically during aftershock sequences (Woessner and Wiemer, 2005; Helmstetter *et al.*, 2006; Hainzl, 2016). In the past, it often took weeks or even years to postprocess the rich catalogs of aftershock sequences to make them fully useful for statistical seismology. Consequently, another objective of our study is to investigate the reliability of assessed statistical parameters of aftershock sequences in the light of improved modern-day network-processing capabilities and automation. A further, related objective is to analyze whether high precision and more complete datasets based on cross correlation, provided by Shelly (2020), can improve the reliability and lower the latency of aftershock forecasting. We also investigate another potential limitation of near-real-time application, the availability of reliable focal mechanism data.

In many ways, the Ridgecrest sequence is an ideal case study for investigating the effects of mainshock on the size distribution of aftershocks, and our study is the first prospective evaluation of the FTLS as a purely data-driven decision support system. Finally, we discuss the implications of our analysis for aftershock hazard assessment.

The 2019 M_w 7.1 Ridgecrest Sequence

On the morning of 4 July (at 17:33 UTC time), an M_w 6.4 earthquake hit eastern California in the Mojave Desert (Ross *et al.*, 2019), injuring about 20 people and damaging numerous buildings in the Ridgecrest area (see Data and Resources). Over the past 40 years, this part of southern California has experienced several moderate earthquakes, the largest being the 20 September 1995 M_w 5.8 event, about 13 km away from the M_w 6.4 event.

The earthquakes following the M_w 6.4 quake outline two lineaments: one southwest–northeast and the other northwest–southeast, on an unmapped fault, exhibiting a distinctive "T" pattern created by the simultaneous activation of two or more faults (Hobbs, 2019; Ross *et al.*, 2019). During the hours after the mainshock, the U.S. Geological Survey (USGS) seismologists estimated in near-real-time probabilities of aftershocks and subsequent mainshocks, using in essence the Reasenberg and Jones (1990) approach (see Data and Resources). Immediately after the mainshocks, this model estimated the weekly probability of one quake being followed by a second mainshock of equal or larger magnitude at about 9% (Hardebeck *et al.*, 2019; Michael *et al.*, 2020). This figure was higher than the default value of 5% obtained when using the standard Reasenberg and Jones (1990) parameter, because of the higher than average aftershock productivity in the region (Hardebeck *et al.*, 2019). Just one day later, an M_w 7.1 earthquake struck (at 8:20 p.m. local time on 6 July or 03:20 UTC) at a distance of about 7 km.

The aforementioned probabilities of a subsequent larger earthquake occurring, as is common in California, were also cited in public. For example, after the second event Lucy Jones tweeted: "So the **M** 6.4 was a foreshock. This was a **M** 7.1 on the same fault as has been producing the Searles Valley sequence. This is part of the same sequence." This was followed by: "You know we say 1 in 20 chance that an earthquake will be followed by something bigger? This is that 1 in 20 time." And then: Yes, we estimate that there's about a 1 in 10 chance that Searles Valley will see another **M** 7. That is a 9 in 10 chance that tonight's **M** 7.1 was the largest."

Here, we monitor fluctuating b-values and apply the FTLS in a near real-time application, comparing the FTLS forecast with currently used aftershock probabilities for California. We then compare the FTLS's performance with preliminary, revised, and high-resolution datasets. A key aim in our research was to evaluate the feasibility of using b-value fluctuations for real-time hazard assessment.

Data and Method

To compute a reliable and detailed b-value time series, we used a window approach, moving a window with a fixed sample size, event by event. To provide a prospective evaluation, the method strictly adheres to the approach used by Gulia and Wiemer (2019; in review before the Ridgecrest mainshocks). The codes used can be downloaded from the ETH Zurich website (see Data and Resources). Here is a brief description of the approach and sequence-specific aspects. Using the quick focal mechanism of the M_w 6.4 (Global Centroid Moment Tensor [Global CMT], Dziewonski *et al.*, 1981; Ekström *et al.*, 2012) and the Wells and Coppersmith relationships (Wells and Coppersmith, 1994) corresponding to the tectonic style of the event strike slip, we built two possible fault planes, with a 1 km spaced grid. To decide quickly and automatically, which was the most likely fault plane, we selected all events recorded in the sequences within the first hour and within a radius of 3 km from each grid point of the fault plane (from now on, the box), then selected the plane in which most of the aftershocks occurred. Although more sophisticated rupture planes using multiple fault segments, among other things, are often available for larger events within days, we opted to apply a simple, quick, and robust approach that will facilitate independent testing as well as real-time application. We divided the dataset into two parts: a pre- and postinitiating event catalog. The start time of the precatalog depended on the quality and completeness of the local network: for the Californian seismicity, we downloaded from the Advanced National Seismic System Comprehensive Earthquake Catalog (ComCat) via the International Federation of Digital Seismograph Networks webservice (see Data and Resources); we started the analysis of the background seismicity from 1981, when the network was greatly improved. The data were first downloaded on 14 July 2019 and then updated week by week.

The computation of b-values critically depends on correct estimates of the magnitude of completeness (M_c) (e.g., Mignan and Woessner, 2012). A specific M_c was assessed for each window (250 event long) after a precutting level, established using the maximum curvature method with a correction factor of 0.2 (Wiemer and Wyss, 2000). A b-value was then calculated for each window, applying the maximum-likelihood method (Aki, 1965). We then defined a pre-event reference b-value, which was the median of all the single estimates preceding the M_w 6.4.

For the postevent catalog processing, we had to consider the temporal changes of the magnitude of completeness following a big event (Helmstetter *et al.*, 2006; Tormann *et al.*, 2013), which can easily mask or bias the space–time b-value fluctuations. During the first hours after a large event, M_c typically changes by two orders of magnitude, resulting in a somewhat heterogeneous dataset. Changes in completeness are not only network specific, but also depend on mainshock magnitude (Helmstetter *et al.*, 2006). Our analysis of Ridgecrest's completeness (Fig. 1) was fully consistent with previous experience, since M_c increased much more and over a longer time span after the M_w 7.1 than after the M_w 6.4 event. Specifically, after the M_w 6.4 M_c increased from the background value ($M_c = 1.2$) to about 1.8, before dropping back to a near-to-background value within 12 hr. After the M_w 7.1 event, it increased to between 3.3 and 3.5, then recovered within three days to near-to-background values.

Although we subsequently estimated M_c in each sample before computing a- and b-values, a common observation is that during periods of very strong gradients the M_c estimate is not conservative enough (e.g., Woessner and Wiemer, 2005), which potentially biases the analysis toward lower b-values. Based on our M_c analysis (see Fig. 1), typically in keeping with such an analysis (e.g., Gulia *et al.*, 2018), we therefore excluded from the dataset those events recorded during the initial, most heterogeneous period after the M_w 6.4 and M_w 7.1 events, and introduced a minimum cutoff magnitude. In the aftermath of the M_w 6.4, we excluded events occurring during the first 12 hr and precut the dataset at **M** 1.7. For the M_w 7.1, we removed events occurring during the first 48 hr and precut at **M** 1.2 (see the shaded areas in Fig. 1). This "no-alert-time" is certainly one of the limitations affecting the method's practical application: for the shorter this no-alert-time is, the more use FTLS decision support can be for practical mitigating actions. We subsequently tested the choice of these expert-selected parameters for sensitivity and confirmed that they did not critically influence our results. Subsequently, we also used an alternative, revised, and higher-resolution dataset (Shelly, 2020) to challenge and refine our analysis. Computing the percentage difference compared to the reference b-value was the final step. The values thus obtained allowed us to define the level of alert. If the percentage difference of the post-M_w 6.4 event was $\pm 10\%$, the alert was designated green or red, otherwise it was classified as orange.

Figure 2 schematically illustrates schematically the process of constructing b-value time series and FTLS values for the Ridgecrest earthquake sequence. This figure contains the b-value difference in percentage respect to the reference value to allow comparison between the two fault planes. After the occurrence of the first event with **M** greater or equal than 6, we calculate the b-value time series on its box, as explained in the previous lines, till the occurrence of a bigger event (step 1 in Fig. 2). Once a larger event occurs, we automatically refocus

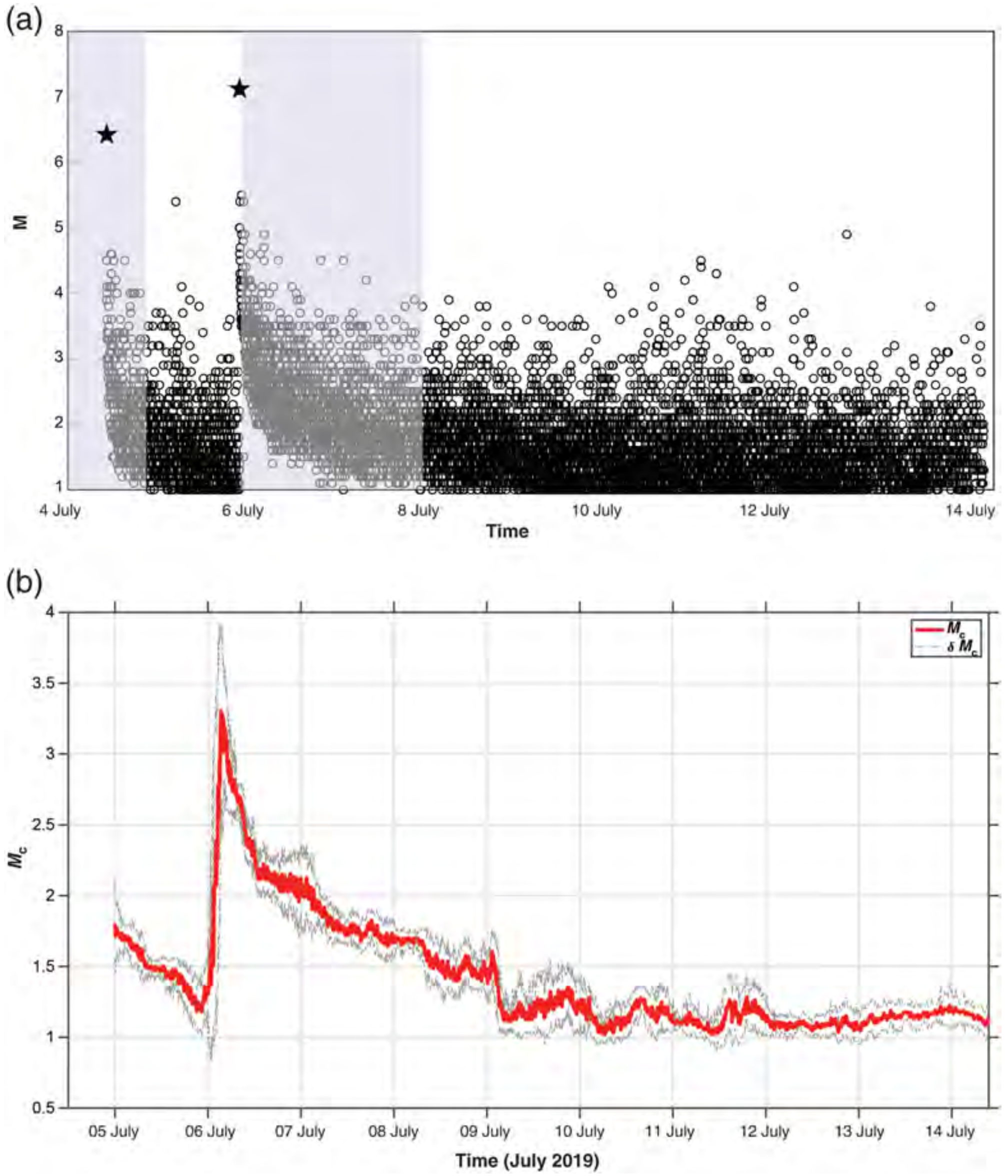


Figure 1. (a) Time-magnitude plot for the events following the 4 July M_w 6.4. Shaded areas indicate times when the dataset was least complete. (b) Time series of the magnitude of completeness (red lines) estimated using the maximum curvature method for samples containing 300 events, moved through the data in overlapping windows. Gray lines represent uncertainty estimates obtained by bootstrapping.

the analysis of b-value changes and FTLS on this new event, using the same procedures: We reselect the fault plane with the highest number of early aftershocks, reselect a new dataset, and finally rerun the code that estimates the background b-value (note from 1981 to the **M** 64, only, excluding the aftershocks and mainshocks of the first sequence) and aftershock b-values (step 2 in Fig. 2). We normalize always the b-values relative to the background value, allowing for comparisons between different sequences in one timeline (Fig. 2b) and refocusing on the new. Larger fault area is sensible, since the stress changes introduced by this event (larger and more recent) will dominate the changes in seismicity; this is now also the area of highest concern for larger events and the area with the most seismicity for analysis. In essence, all steps are automated and follow the procedure by Gulia and Wiemer (2019), the only "free" parameter is the starting date of the background b-value analysis (here, 1981).

Mapping of b-values to provide additional information on spatial changes was performed on a regular 1×1 km grid, selecting the closest 200 events within a maximum radius of 10 km. For the time series, we used the maximum curvature method (Wiemer and Wyss, 2000) for M_c, after precutting the dataset at the same M_c level already adopted for the pre and post time period. We plotted the percentage difference of the post $M_w \geq 6$ events with respect to the b-value map obtained for the background (i.e., the time span from 1981 up to the last event preceding the M_w 6.4).

The subcatalogs generated for each fault plane and for the three different catalogs are provided as text files in the supplemental material available to this article.

Results

Automatic fault selection

The 4 July M_w 6.4 Ridgecrest earthquake ruptured two conjugate strike-slip faults, which intersected to form a "T" shape. It took days before geodetic, seismic, and relocated seismicity data provided an overall view of this complex sequence (Hobbs, 2019; Ross *et al.*, 2019). By kinematically inverting for subevents using seismograms from the dense regional seismic network and global seismic stations, Ross *et al.* (2019) identified three simultaneous subevents and hypothesized that the rupture had been a cascading phenomenon, rather than a single continuous process. The three identified subevents coincided with at least three faults: the 6 km long northwest-trending fault that slipped first; then, the rupture propagated

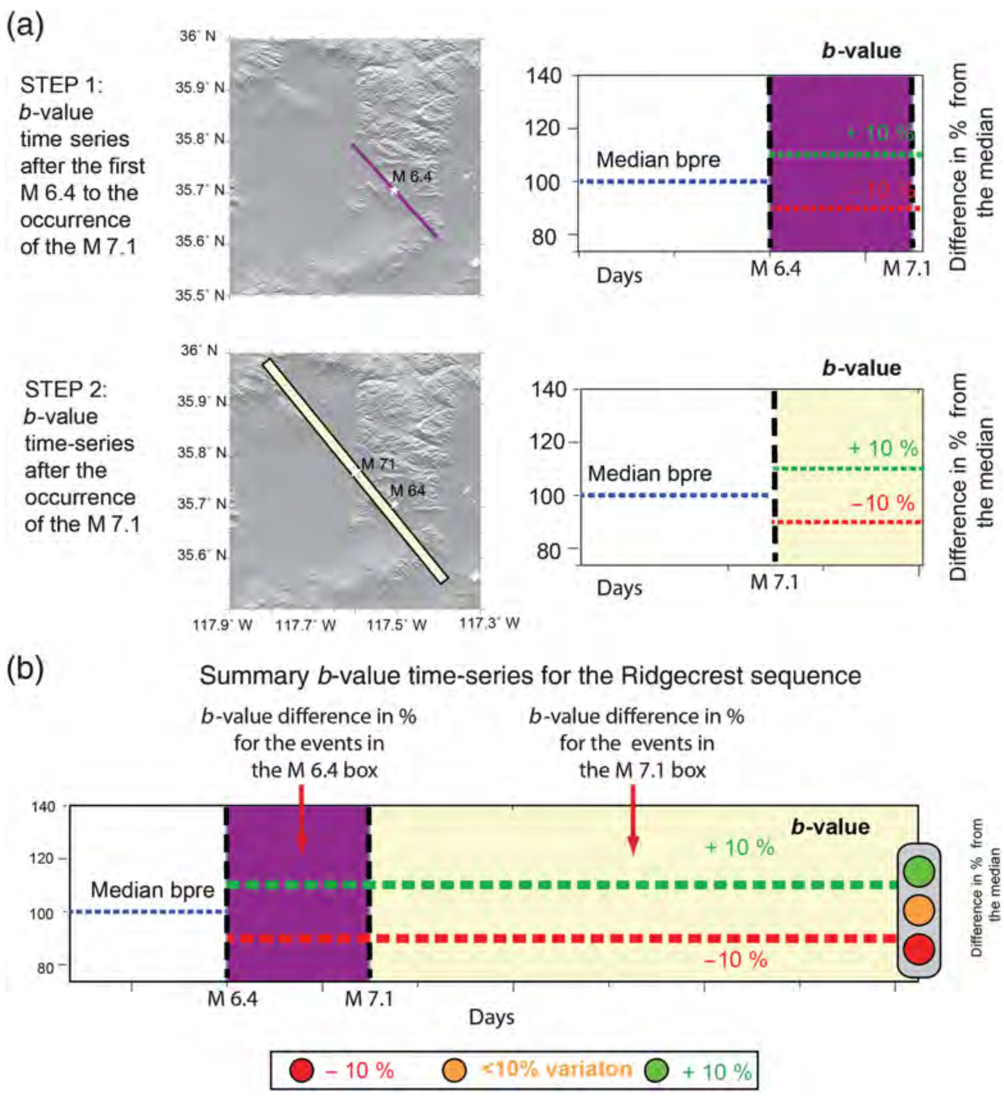


Figure 2. (a) Schematic representation of the single time series obtained on the **M** 64 and **M** 7.1 fault planes and (b) the summary one with the two fault planes in the near-real-time analysis of the Ridgecrest earthquake sequence.

over a short southwest-trending fault with only about 5 km of surface break, and finally the jump to a larger southwest-trending fault roughly 15 km long (Ross *et al.*, 2019).

The FTLS method was developed to be applied in near-real time, when both little other information apart from data from the focal mechanism and the automatically derived network catalog are known and publicly accessible. The seismic source used in our analysis is thus represented by a single plane. Following the method described by Gulia and Wiemer (2019), once the Global CMT provided the focal mechanism, the algorithm built the two fault planes, centered on the local hypocenter catalog (see Data and Resources). Between the two fault planes, the one with the largest number of early aftershocks within a 3 km radius was selected as the likely fault plane. For M_w 6.4, this purely statistical method chose the northwest-trending fault plane (Fig. 3) that represented the initial rupture, in the process described by Ross *et al.* (2019), and is the one aligned with the eventual M_w 7.1 hypocenter. The background or reference *b*-value for this box containing 1275 events above **M** 1 since 1981 is $b = 0.97$ (blue symbols in Fig. 4).

Seismicity preceding M_w 7.1

Figure 4a shows the *b*-value time series. All *b*-values after the M_w 6.4 event are substantially lower than the background *b*-value. A comparison of the frequency–magnitude distributions (FMDs) of events occurring between 4 and 6 July is in Figure 4b. During the time interval between the two big events, the *b*-value decreases from 0.97 to 0.75, a decrease by 23%, resulting in a red Foreshock Traffic Light System status (Fig. 4b). We also calculated the respective daily probability (Pr) commonly derived by extrapolating the observed FMD to an M_w 6.4 event or larger earthquake (Fig. 5c). These probabilities reached a peak value of 66% on 5 July, a value about one order of magnitude larger than the aforementioned ones derived by the USGS (see Data and Resources; 5% using default values, 9% using sequence-specific values according to Michael *et al.*, 2020 and Hardebeck *et al.*, 2019).

Next, we mapped the spatial distribution of the differential *b*-value (i.e., the background *b*-value map subtracted from the current episode map) to infer information on the likely nucleation region of a subsequent mainshock (Fig. 6a). The expectation described by Gulia and Wiemer (2019) is that a subsequent mainshock would nucleate near the strongest *b*-value decrease, in our conceptual model represented by high-stress asperities. In the case of the Ridgecrest sequence, this low *b*-value patch locates to the northwest of the M_w 6.4 epicenter and corresponds closely to the location of the subsequent 6 July M_w 7.1 (marked in Fig. 6b).

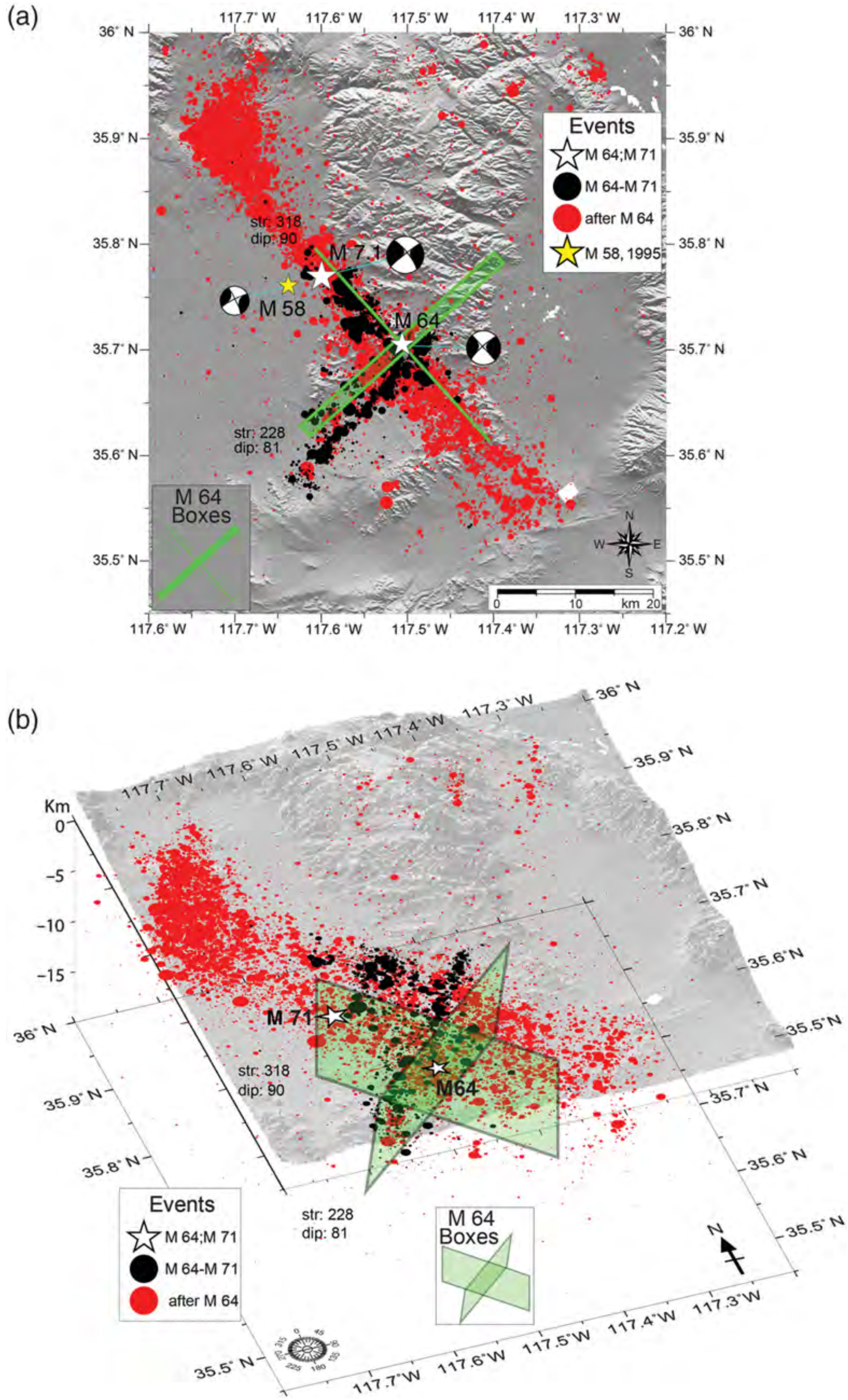


Figure 3. (a) Seismicity map with the events (white stars) on 4 July – M_w 6.4 (**M** 64), 6 July – M_w 7.1 (**M** 71) and subsequent events in black and red, respectively. The two green fault planes indicate the M_w 6.4 Global Centroid Moment Tensor (Global CMT) focal mechanism, with strike and dip directions. (b) 3D view of panel (a), from a 200° azimuth and 40° elevation.

Seismicity following the M_w 7.1

We then analyzed b-value evolution over time in the M_w 7.1 source volume, constructed following the same procedure as described previously for the M_w 6.4 event. We also determine a new background b-value of 0.87 for this much larger source volume, compared to the volume of M_w 6.4. The b-value time series, plotted in Figure 2 and starting two days after the M_w 7.1 earthquake, indicated a general increase from the normalized background value of more than 10%, reaching a peak of 26% within the first week (Fig. 2c). This qualified it for green FTLS status and suggested that the chance of a subsequent even larger event was lower than average. Figure 4c shows the FMDs of the background ($b = 0.87$) compared to the aftershocks ($b = 1.1$). We again calculated the probability of a subsequent event of equal or larger magnitude at 0.4% per day two days after the event and falling to 0.004% per day in subsequent weeks. These values were one order of magnitude lower than the USGS aftershock probabilities communicated during the sequence. The differential b-value map for events occurring in the first week with respect to their background (Fig. 6b) indicated a general rise in b-values throughout the region.

Revised and high-resolution datasets

Although this manuscript was under review, revised Global CMT and ComCat catalogs (downloaded on 21 January 2020) became available, so we repeated our analysis, to compare it with the performance of

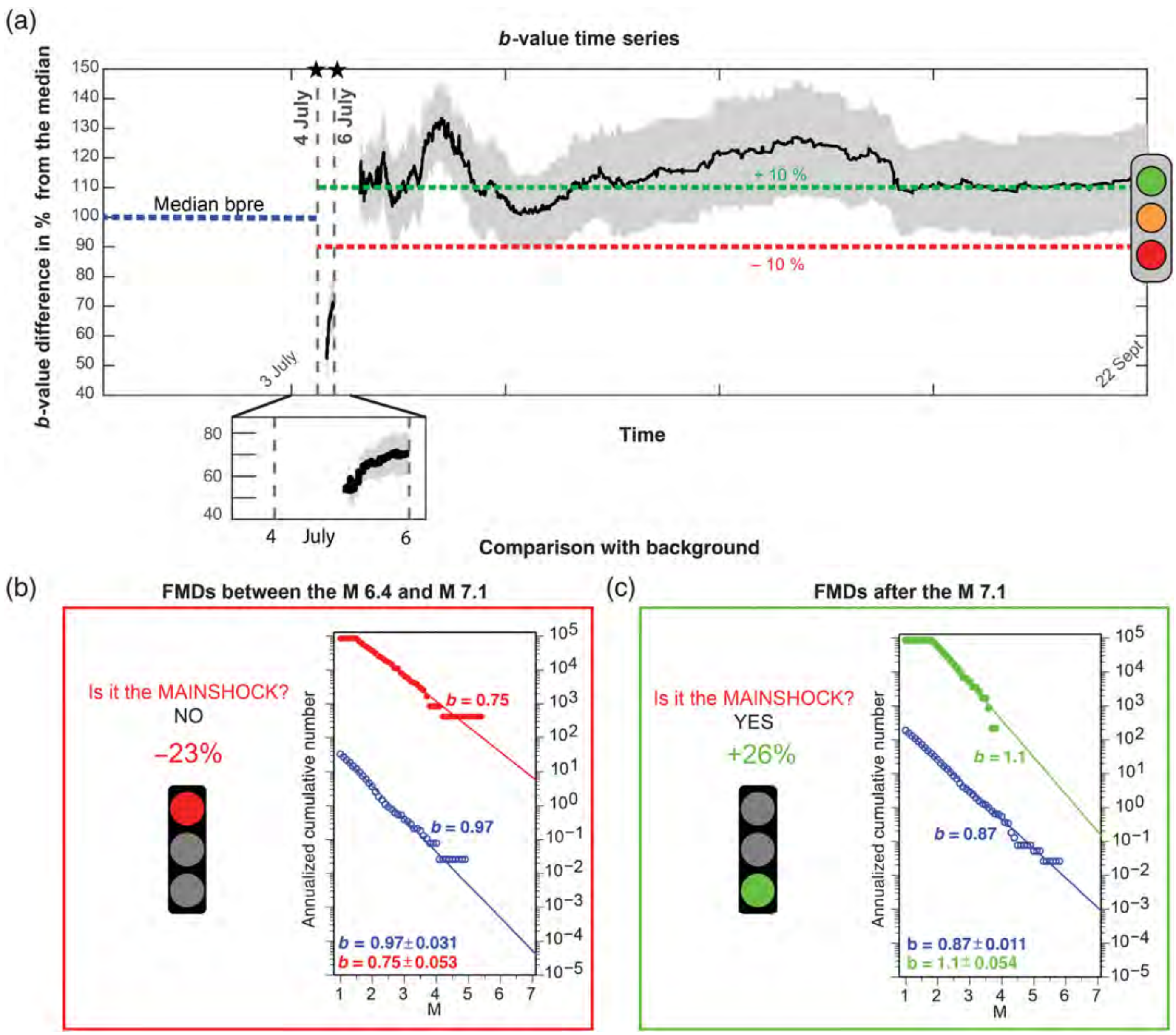


Figure 4. Performance of the foreshock traffic-light system (FTLS) in near-real time. (a) b-value time series for the M_w 7.1 sequence superimposed on the FTLS assessment (Gulia and Wiemer, 2019); blue dashed line is the reference b-value; black dashed vertical lines indicate M_w 6.4 and M_w 7.1, respectively. Black rectangle zooms in on the time series in the interval between the two **M** > 6 events. All the estimates are below the reference value. Gray indicates uncertainty (one standard deviation by Shi and Bolt, 1982). (b) Frequency-magnitude distributions (FMDs) for the source of the M_w 6.4 event for two time periods: background in blue and time between the two $M_w > 6$ events in red. (c) FMDs for the source of the M_w 7.1 event for two time periods: background in blue and maximum b-value reached in the first week of aftershocks.

FTLS using near-real-time data. The revised Global CMT focal mechanisms, available online since 8 November 2019, are very similar to their quick equivalents (Table 1), both in orientation and dip. We then recomputed the fault planes centered on the hypocenters of the two mainshocks (M_w 6.4 and M_w 7.1) for the revised ComCat catalog as well as for the high-resolution catalog compiled by Shelly (2020).

Minor displacement (by approximately 0.2 km) of the epicenter of the 4 July mainshock in the revised ComCat catalog makes the revised boxes imperceptibly different with respect to their quick counterparts (Table 2). The overall completeness of the catalogs remains largely unchanged. Consequently, the result showed the same almost imperceptible difference, with the overall b-value during the time interval between the two biggest events rising from 0.75 to 0.76, and the red alert from −23% to −22%. After the mainshock, we obtained the same b-values and the same green alert (+26%).

In addition, Shelly (2020) published a revised, higher-resolution catalog containing 34,000 events during the period 4–16 July for the Ridgecrest sequence, allowing us for the first time

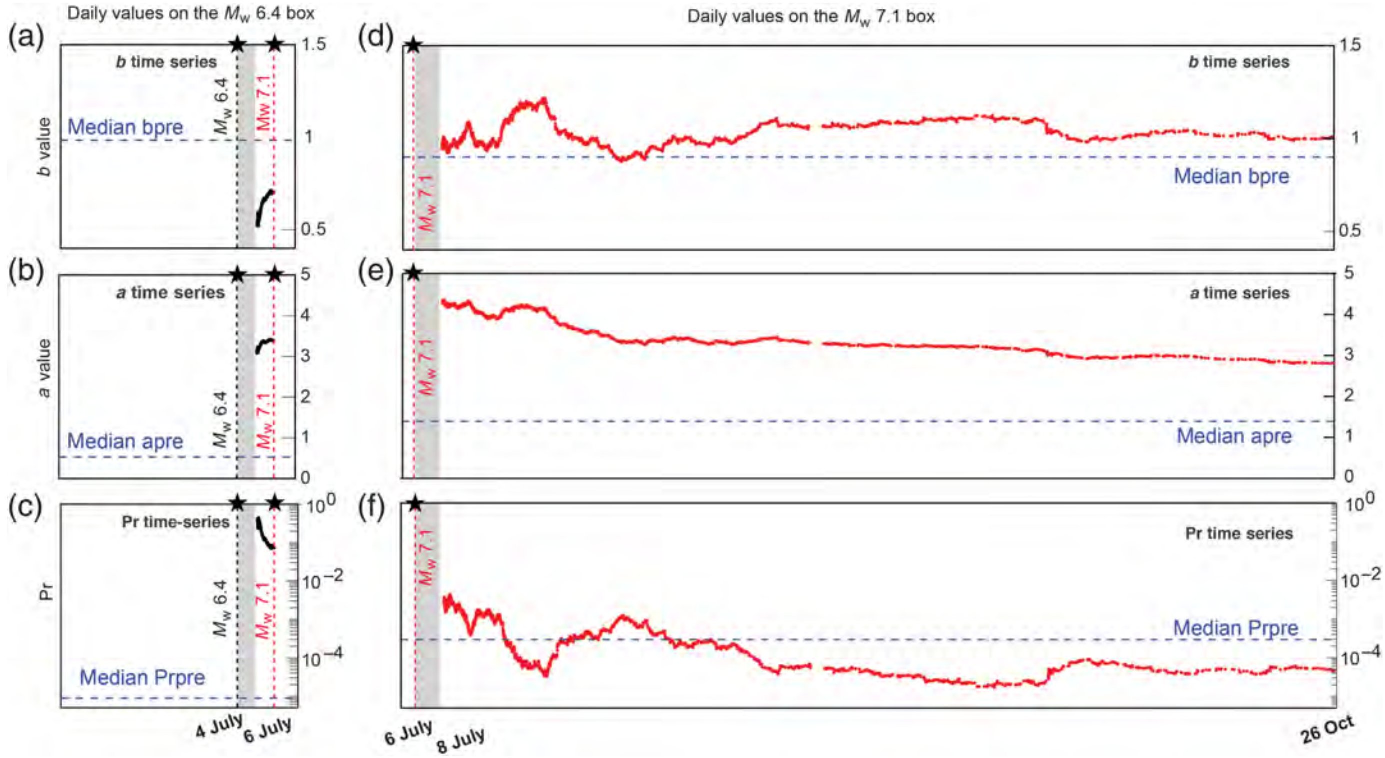


Figure 5. (a)–(f) Daily time series on the fault planes of the two major events. (a)–(c) Fault plane of the M_w 6.4 event: (a) b-value, (b) daily a-value, and (c) daily probability (Pr) of an M_w 6.4+. (d)–(f) Fault plane of the M_w 7.1 event: (d) b-value, (e) daily a-value, and (f) daily Pr of an M_w 6.4+. Blue dashed lines represent the mean value of all the background estimates.

to evaluate the b-value evolution and FTLS performance with a partially independently calculated and presumably higher quality dataset. This earthquake catalog is based on cross-correlation analysis of continuous waveforms and according to Shelly (2020) substantially more complete in magnitude, more consistent through time, and more precise in hypocenters. Shelly (2020) points out that cross correlation is not well suited for relocating **M** > 5 earthquakes, especially the two events with the highest magnitudes, because its waveforms are too dissimilar to those of smaller events. Indeed, in this dataset, the two epicenters roughly correspond to the location provided by USGS, albeit having different depths, with the M_w 6.4 deeper (from 10.5 to 15 km) and the M_w 7.1 shallower (from 8 to 3 km). For this reason, we use the same source volumes determined for the previous analysis (i.e., revised Global CMT moved to the ComCat hypocenter).

This catalog contains only 38 events preceding the M_w 6.4 quake, not enough to establish a reference b-value for the FTLS, so we used the revised ComCat catalog to estimate that value for the boxes of the M_w 6.4 and M_w 7.1 mainshocks. As shown in Shelly (2020), the cross-correlation analysis substantially lowers these events' overall magnitude of completeness, a finding supported by our $M_c(t)$ analysis (Fig. 7). The Shelly catalog reaches an M_c of about 0.7, roughly half-degree of magnitude lower than the standard ComCat catalog. However, the increase in M_c immediately after the mainshock is almost as high (rising to roughly $M_c = 3.0$–3.5), but completeness recovers faster and more systematically. Completeness for **M** 1.5 is reached 24 hrs earlier than using standard datasets (Fig. 7). This improvement is extremely important for our approach, but also for other real-time methods used to assess time-dependent earthquake probabilities.

Using the Shelly catalog, we repeated the b-value analysis using the same time windows but lower completeness and found almost identical results (−21% after the M_w 6.4 and +29% after the M_w 7.1), confirming that the results based on near-real-time data are in line with the more homogeneous, higher-quality catalog. To exploit the possible improvements of higher-quality data for aftershock hazard assessment, we then moved the start of our analysis closer to the mainshock origin time, thus shortening our no-alert time. After the M_w 6.4 earthquake, we were able to cut this no-alert time from 12 hr to just one, and after the M_w 7.1 from 48 to 24 hr (using M_c precuts of 1.5 in both cases). The time series of b-values is shown in Figure 8. The overall trend, the b-values themselves, and FTLS status all remain unchanged. However, it is worth noting that we can establish a low b-value after the **M** 6.4, with just 1 hr of no-alert time when high-quality data is available.

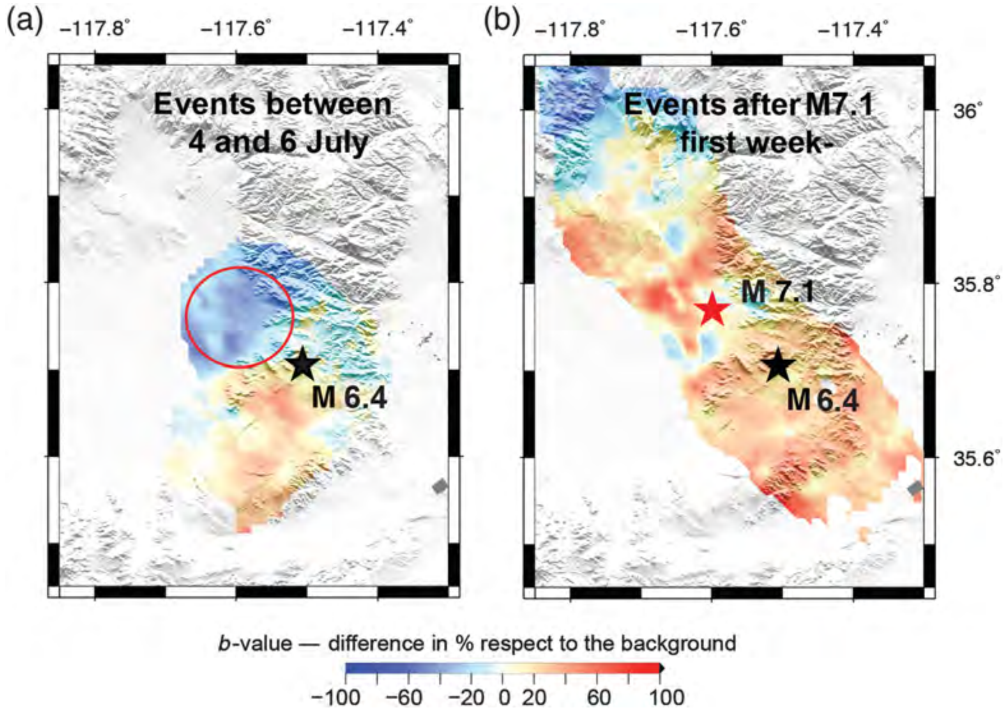


Figure 6. Mapped *b*-values with the difference in percentage with respect to the background for two different periods: (a) between M_w 6.4 and M_w 7.1; (b) the first week after M_w 7.1. The original maps were produced by ZMAP (Wiemer, 2001; Reyes and Wiemer, 2019) and postprocessed in the MATLAB using Generic Mapping Tools (GMT; see Data and Resources). Black star: **M** 6.4 epicenter; red star: **M** 7.1 epicenter.

Sensitivity analysis

Our method contains essentially three free parameters that we determined based on data analysis and expert choices: (1) the magnitude of completeness, (2) the no-alert time, and the (3) the sample size analyzed. The first two we have determined based on the completeness analysis (Figs. 1 and 6), and the last is a commonly used value in studies. We introduce a novel sensitivity analysis to evaluate the impact of the changes on the result of our study. We scan systematically the parameter space of the precut M_c and no-alert time parameters. The results shown in Figure 9 for the revised ComCat and the Shelly catalog are fully consistent with the previous interpretations: For all choices of M_c and no-alter times, there is a string decrease in *b*-value (red colors and red FTLS status) subsequent to the **M** 6.4. Following the **M** 7.1, the picture is somewhat different: for value at or below the estimated completeness (black dashed line in Fig. 9), there is decrease in *b*-value—an expected bias due to incompleteness. Above M_c, however, green colors indicate an increase in *b*-value and green FTLS status.

Seismic sequence in 1995

In 1995, an M_w 5.8 earthquake occurred in the same region, a few kilometers away from the M_w 7.1 (Fig. 3a). That event was not followed by a larger one. For comparison, we also applied the FTLS approach to this sequence, too. Figure 10 shows the FMDs and time series relative to the 1995 sequence, indicating a roughly 30% increase in the *b*-value, resulting in a correct green traffic-light classification. This result suggests that the FTLS approach can also be extended to events of smaller magnitude than the currently used $M_w \geq 6.0$ reference.

Discussion and Conclusion

Our analysis shows that the Ridgecrest earthquake sequence not only impacted the seismic activity rate, increasing the productivity of earthquakes near the fault by between 3 and 5 orders of magnitude, but also changed the relative size distributions and the *b*-values, in both space and time. This should

TABLE 1
Nodal Planes (np1 and 2) of the Quick and Revised Global Centroid Moment Tensor (Global CMT) Catalog for the Two Events on 4 July 2019, 17:33 UTC (Day 04) and 6 July 2019, 03:19 UTC (Day 06)

Global CMT	Day	Strike np1	Dip np1	Rake np1	Strike np2	Dip np2	Rake np2	Length (km)	Width (km)
Quick	04	228	81	0	318	90	−171	27.28	9.65
	06	322	78	−177	231	87	−12	61.9	13.79
Revised	04	227	86	3	137	87	176	26.84	9.58
	06	321	81	180	51	90	9	61.3	13.73

TABLE 2
Vertices of the Fault Planes (FP1 and FP2) Corresponding to the Nodal Planes in Table 1

Global CMT	Day	Lon FP1 (°)	Lat FP1 (°)	Depth FP1 (km)	Lon FP2 (°)	Lat FP2 (°)	Depth FP2 (km)
Quick	04	−117.3882	35.7822	5.9452	−117.4049	35.614	5.8858
		−117.6127	35.6181	5.9452	−117.6071	35.7963	5.8858
		−117.6238	35.6281	15.4748	−117.6071	35.7963	15.5342
		−117.3993	35.7923	15.4748	−117.4049	35.614	15.5342
	06	−117.4007	35.5422	1.2576	−117.3302	35.9421	1.1164
		−117.823	35.9809	1.2576	−117.8634	35.5918	1.1164
		−117.798	35.9968	14.7424	−117.8684	35.5969	14.8836
		−117.3756	35.5581	14.7424	−117.3353	35.9472	14.8836
Revised	04	−117.3926	35.7854	5.7213	−117.6032	35.7951	5.7162
		−117.61	35.6208	5.7213	−117.4004	35.6186	5.7162
		−117.6151	35.6252	15.2787	−117.4045	35.6155	15.2838
		−117.3977	35.7898	15.2787	−117.6072	35.7921	15.2838
	06	−117.3948	35.5492	1.2208	−117.8633	35.596	1.1363
		−117.8224	35.9776	1.2208	−117.3353	35.943	1.1363
		−117.8039	35.9898	14.7792	−117.3353	35.943	14.8637
		−117.3763	35.5614	14.7792	−117.8633	35.596	14.8637
Shelly (2020)	04	−117.3874	35.7885	10.2753	−117.598	35.7982	10.2702
		−117.6048	35.6238	10.2753	−117.3952	35.6216	10.2702
		−117.6098	35.6282	19.8327	−117.3993	35.6185	19.8378
		−117.3924	35.7929	19.8327	−117.602	35.7951	19.8378
	06	−117.3896	35.5515	−3.5382	−117.8582	35.5984	−3.6227
		−117.8172	35.9799	−3.5382	−117.3302	35.9453	−3.6227
		−117.7987	35.9921	10.0202	−117.3302	35.9453	10.1047
		−117.3711	35.5637	10.0202	−117.8582	35.5984	10.1047

See Table 1 for details of the symbols. Global CMT, Global Centroid Moment Tensor; Lat, latitude; Lon, longitude.

come as no surprise, since the size distribution is known to be sensitive to the applied shear stress on faults (e.g., Goebel *et al.*, 2013) and also to depend on location (e.g., Tormann *et al.*, 2015). Thus, b-values are not only linked to the seismotectonic context and evolution of events, but they also constitute an important factor influencing the probability of a subsequent larger event. The FTLS concept introduced by Gulia and Wiemer (2019) exploits the systematic differences in b-values observed between the majority of aftershock sequences that will normally decay over time and the small percentage of sequences that are followed by an even larger event. The FTLS method and codes were developed in the first half of 2019, but only published in October 2019 (Gulia and Wiemer, 2019). The Ridgecrest sequence, representing one of the best-monitored large mainshock–aftershock sequences, presented us with an ideal opportunity to test the FTLS hypothesis and developed software. The analysis presented is here is not yet a truly prospective, real-time application, because we were (and still are not) set up computationally and, to a certain extent, methodologically to conduct such an urgently needed but challenging test. However, it is meaningful in a pseudoprospective sense, an analysis that reproduces real-time condition. Our pseudoprospective study is, however, more rigors, and we would argue more meaningful than typical such studies, because the method and codes used to conduct the automatic analysis have been published before and were here used unchanged from the version of the method submitted for publication. In other words, they could not have been optimized to provide the best outcome for our hypothesis.

The results obtained and presented in this article support the FTLS hypothesis: seismicity following the M_w 6.4 event

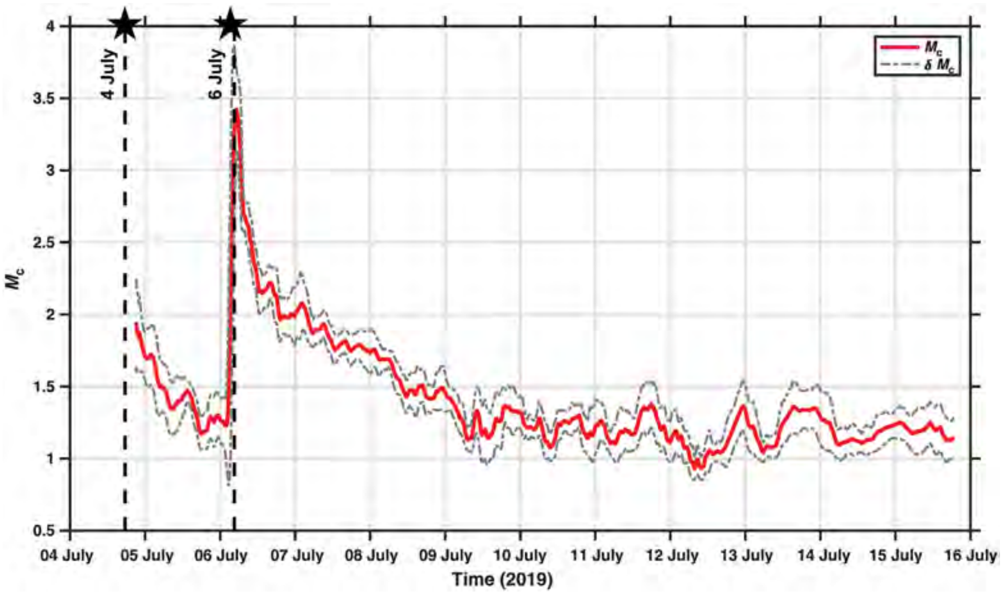


Figure 7. Time series of the magnitude of completeness (red lines) in the catalog by Shelly (2020) estimated using the maximum curvature method for samples containing 300 events, moved through the data in overlapping windows. Gray lines represent uncertainty estimates obtained by bootstrapping. The black dashed lines as well as the black stars represent the time of the **M** 6.4 and **M** 7.1 events.

knowing that a larger event is unlikely will be extremely valuable information. Indeed, after the **M** 7.1, we estimate about a factor 10 lower probability for a subsequent larger one that the standard USGS model.

Naturally, in principle, it would be great to extend the FTLS model to smaller mainshocks, because more data could be used to test the hypothesis. However, the data would have to be of very high quality, and their inclusion would probably increase the uncertainty of the analysis. The smaller size of the fault planes involved in such events (e.g., an**M** 5.5 source would be about 6 km long) would make it more challenging to identify the active fault. Because smaller mainshocks will generally result in fewer aftershocks, the spatiotemporal resolution of b-values is reduced, and the useful magnitude range between the largest events and M_c decreases, making it more difficult to establish reliable b-values. Probably scaling works in such a way that we would have to select events even closer to the mainshock fault only, which, in turn, makes pinpointing the location even more challenging. Also, sample sizes may be too small for robust analyses. Similarly, the relevant background (i.e., the reference level) would be even more local and thus harder to determine. In addition, the coulomb stress and failure modeling in Gulia *et al.* (2018) suggest that the amplitude of the b-value increase is magnitude dependent, so it is unclear whether b-value transients are scale invariant. Therefore, it needs to be explored whether the evaluation of the FTLS hypothesis can be extended to smaller events, but this will necessitate a very thorough analysis of any uncertainties and their influence on the stability of the analysis. An analysis of that kind is beyond the scope of this Ridgecrest case study.

The spatial patterns of changes in b-values have been proposed as additional information on the future location of subsequent larger events, and here too the Ridgecrest case study is well in line with this loosely formulated and as yet not formally tested hypothesis: the M_w 7.1 event occurred near the area of the steepest b-value decrease (Fig. 6). More research and testing are needed to integrate this spatial information into aftershock forecasting in an automate way; for now, we consider the information contained in b-value or earthquake probability maps additional information for experts to be considered.

showed a substantially lower b-value (a drop of 23%, Fig. 4), resulting in its correct red traffic-light designation. The b-value also rose by 26% after the M_w 7.1 quake, resulting in a correct green classification. This adds one correct positive and one correct negative to the confusion matrix analysis presented in Gulia and Wiemer (2019), increasing the accuracy assessment to above 96%. A correct green traffic light was also attributed after the 1995 M_w 5.8 earthquake. Because the FTLS hypothesis is proposed and evaluated for events with a magnitude of 6.0 and above, the M_w 5.8 results are not factored into the (retrospective) error matrix score.

The FTLS hypothesis itself needs to be further tested, and the error matrix approach needs to be carried out on future sequences in a fully prospective, independently conducted way. Such tests are now planned as part of the Collaboratory for the Study of Earthquake Predictability (Schorlemmer *et al.*, 2018), financed by the European Real-time earthquake rIsk reduction for a reSilient Europe project (see Data and Resources). In addition, the observed changes in b-values can and should also be directly converted into time-dependent earthquake probabilities, as shown in Tormann *et al.* (2016) and Gulia *et al.* (2016) for example. These probabilities are also reported for the Ridgecrest sequences (Fig. 5), which are very consistent with the FTLS results and will be tested in comparison to other models, such as the Reasenberg–Jones or ETAS models. The FTLS green alert may turn out to be the most important one in terms of its practical implications, for the vast majority (80%) of all sequences will fall into this category, and

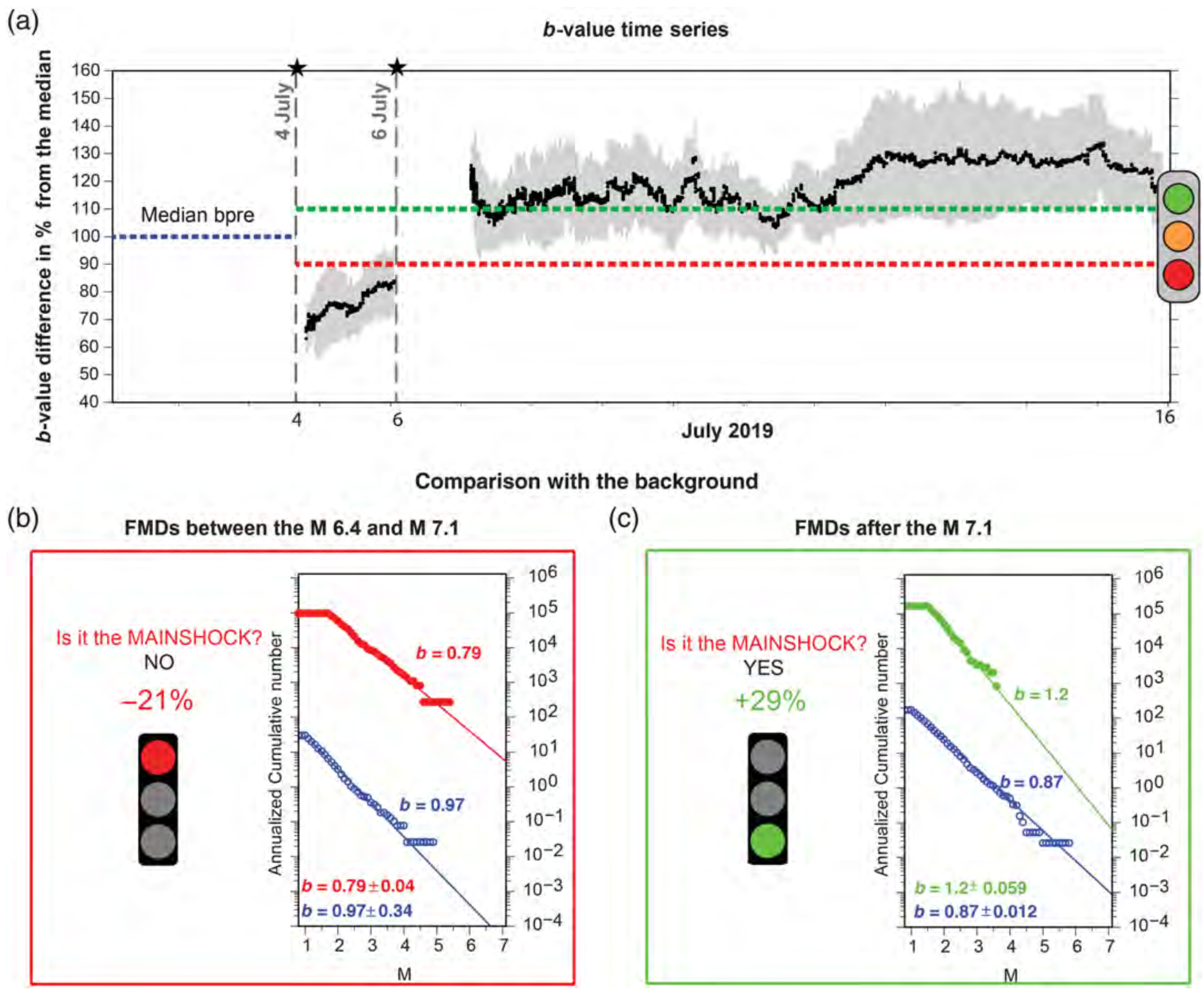
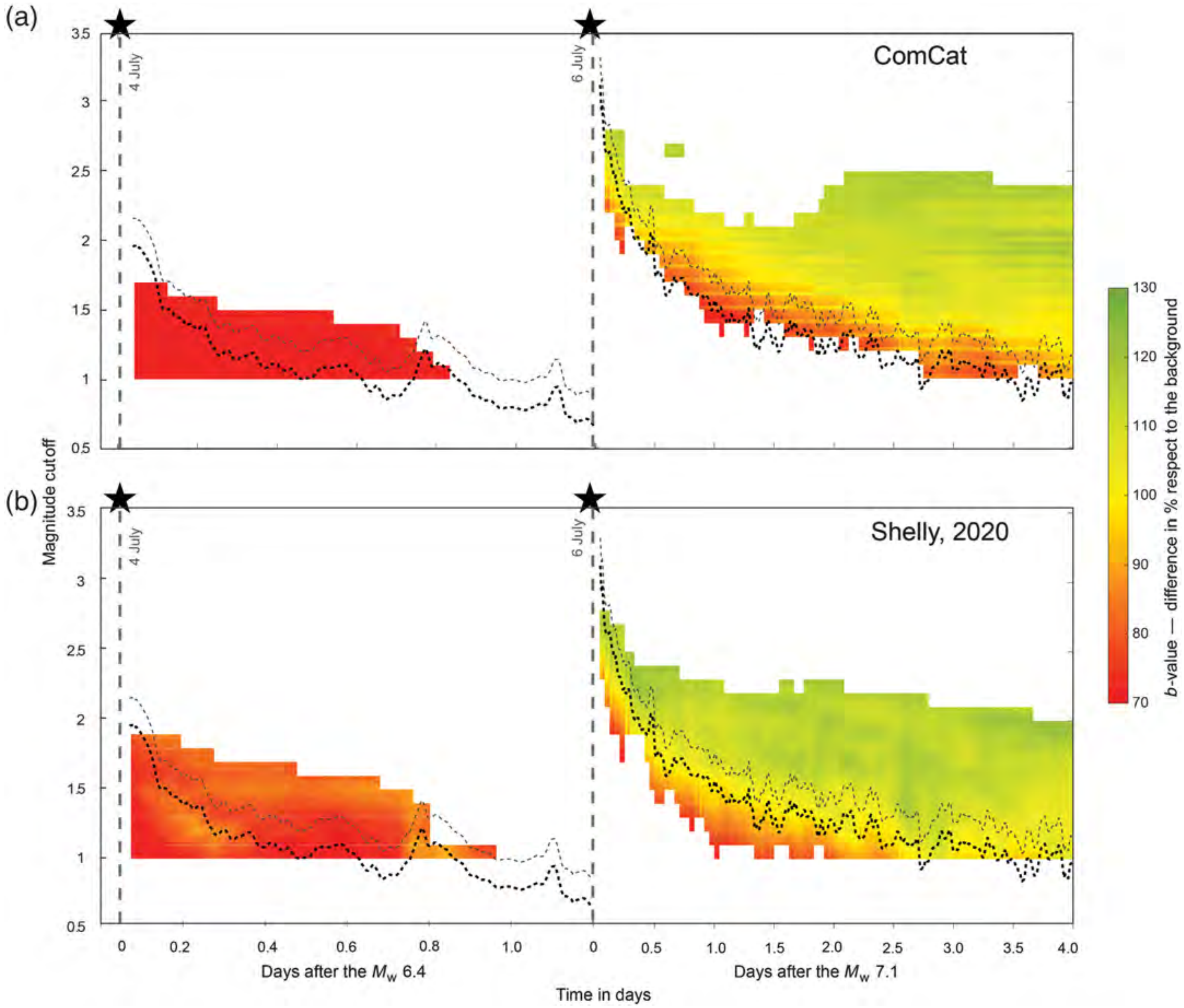


Figure 8. Performance of the FTLS with the high-resolution catalog by Shelly (2020): (a) b-value time series for the M_w 7.1 sequence superimposed on the FTLS assessment (Gulia and Wiemer, 2019); blue dashed line is the reference b-value; black dashed vertical lines indicate M_w 6.4 and M_w 7.1, respectively. Gray indicates uncertainty (one standard deviation by Shi and Bolt, 1982). (b) FMDs for the source of the M_w 6.4 event for two time periods: background in blue and time between the two $M_w > 6$ in red. (c) FMDs for the source of the M_w 7.1 event for two time periods: background in blue and maximum b-value reached in the first week of aftershocks.

Establishing with confidence a b-value time series critically hinges on the quality of the seismic network, and judging from our analysis the southern California network performed extremely well (Fig. 1) in near-real time (much of our analysis was in fact conducted within days of the M_w 6.4 event). The magnitude of completeness rapidly decreased (Fig. 1) and the FMD (Figs. 4 and 9) is among the best we have ever analyzed, closely following a linear Gutenberg–Richter distribution and leading within hours to reliable observations of b-value changes. Based on our experience, the differential b-value maps computed (Fig. 6) are also very reliable. Progress made in station coverage and automated network-processing approaches are clearly delivering very rapidly high-quality data that are useful for scientific analysis and risk assessment. Further improvements using advanced automated postprocessing methods may be feasible and desirable to decrease no-alert time. Our test using the higher-resolution catalog provided by Shelly (2020) supports this (Figs. 8 and 9). The catalog confirms every aspect of the results obtained using ComCat real-time data, so we consider the likelihood of data imperfection influencing our analysis to be very low. Equally importantly, the Shelly catalog allows us to reduce no-alert time to just 1 hr. Because the approach implemented by Shelly in principle reveals the real-time capabilities of seismic networks in the not-too-distant future, we suggest that it may be possible to produce an FTLS assessment within just one or a few hours. We also suggest that the sensitivity analysis to M_c and no-alter



time we introduce in Figure 9 are the powerful tools to quickly evaluate the robustness of an FTLS results. This may be also in real time a graphical representation a seismologist wants to consult in a crisis to ensure the results are not critically dependent on the choice of parameters.

The FTLS hypothesis is quite new, and although the successful Ridgecrest case provides additional support for it, in our view, it is too early to use it routinely for making decisions about civil protection or public communications. More extensive sensitivity and robustness studies are needed; the hypothesis should be independently evaluated by other research teams, and the hypothesis needs to be formally tested. There are plans for this, but it will take time. Simultaneously, numerical modeling may allow the formulation of a better physical understanding and maybe enhanced forecasting abilities. These efforts will take time, but, given the potential implications and greater understanding, we consider them highly worthwhile.

Figure 9. Sensitivity analysis on no-alert time and completeness. Color coded is the *b*-value difference in percentage with respect to the reference *b*-value as a function of magnitude cutoff and time after the M_w 6.4 (left) and M_w 7.1 (right). We always analyzed the first 300 events above this magnitude and after this time. (a) Black dashed line represents the estimated magnitude of completeness for the Comprehensive Earthquake Catalog (ComCat) reported in Figure 1; gray dashed line represents the same with the 0.2 correction factor, as adopted in our modeling; (b) the estimated magnitude of completeness for the high-resolution catalog by Shelly (2020) reported in Figure 7; gray dashed line represents the same with the 0.2 correction factor, as adopted in our modeling.

Data and Resources

The Comprehensive Earthquake Catalog (ComCat) by U.S. Geological Survey (USGS) was downloaded from the website https://earthquake.usgs.gov/fdsnws/event/1/catalogsand ZMAP (last accessed October 2019) (Reyes and Wiemer, 2019). Information about earthquake

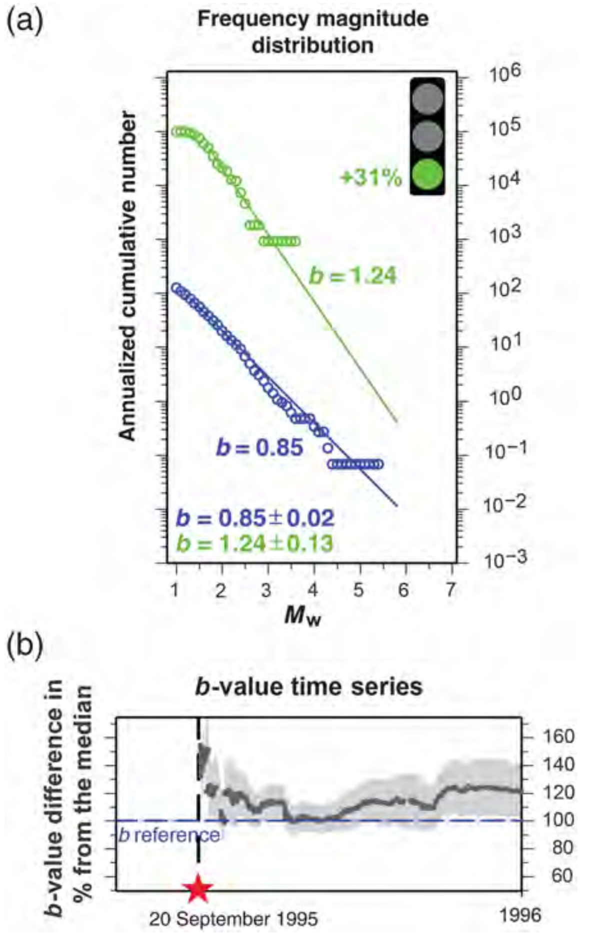


Figure 10. (a) FMDs for the M_w 5.8 sequence in 1995: in blue, the background *b*-value (1981–1995) and in green the highest *b*-value reached by the aftershocks during the first week after M_w 5.8; (b) *b*-value time series for the same sequence. Blue dashed line represents the reference *b*-value (see the Data and Method section).

hazard program is available at https://earthquake.usgs.gov/ (last accessed October 2019). Data about near-real-time probabilities of aftershocks and subsequent mainshocks estimated by USGS are available at https://earthquake.usgs.gov/earthquakes/eventpage/ci38457511/oaf/commentary (last accessed May 2020). The codes can be downloaded from ETH Zurich website (doi: 10.3929/ethz-b-000357449, last accessed May 2020). Information on International Federation of Digital Seismograph Networks (FDSN) webservice is available at https://earthquake.usgs.gov/fdsnws/event/1/ (last accessed May 2020). The network codes of FDSN are available at https://www.fdsn.org/networks/. Data about USGS are available at https://earthquake.usgs.gov/data/oaf/overview.php (last accessed May 2020). Information about Generic Mapping Tools is available at http://gmt.soest.hawaii.edu (last accessed May 2020) and the figures were made using website www.soest.hawaii.edu/gmt (last accessed May 2020). Data about European Real-time earthquake rIsk reduction for a reSilient Europe project are available at www.rise-eu.org (last accessed May 2020). The MATLAB is available at www.mathworks.com/products/matlab (last accessed November 2018). The supplemental material for this article includes the text files of the subcatalogs generated for each fault plane and for the three different catalogs.

Acknowledgments

Figures were made using Generic Mapping Tools (see Data and Resources; Wessel and Smith, 1998). The authors thank the editors, Allison Bent and Anastasia Pratt, Andy Michael, and one anonymous reviewer for helping in improving and clarifying the article. This study was supported by the Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115.

References

Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bull. Earthq. Res. Inst. Univ. Tokyo* **43,** 237–239.

Amitrano, D. (2003). Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the *b*-value, *J. Geophys. Res.* **108,** no. B1, 2044, doi: 10.1029/2001JB000680.

Dieterich, J. H., V. Cayol, and P. G. Okubo (2000). The use of earthquake rate changes as a stress meter at Kilauea Volcano, *Nature* **408,** no. 6811, 457–460, doi: 10.1038/35044054.

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.* **86,** 2825–2852, doi: 10.1029/JB086iB04p02825.

Ebel, J. E., K.-P. Bonjer, and M. C. Oncescu (2000). Paleoseismicity: Seismicity evidence for past large earthquakes, *Seismol. Res. Lett.* **71,** no. 2, 283–294, doi: 10.1785/gssrl.71.2.283.

Ekström, G., M. Nettles, and A. M. Dziewoński (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. In.* **200/201,** 1–9.

Enescu, B., and K. Ito (2003). Values of b and p: Their variations and relation to physical processes for earthquakes in Japan, *Annuals of Disas. Prev. Res. Inst.*, Kyoto Univ., No. 46 B.

Goebel, T. H. W., D. Schorlemmer, T. W. Becker, G. Dresen, and C. G. Sammis (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments, *Geophys. Res. Lett.* **40,** 2049–2054, doi: 10.1002/grl.50507.

Gulia, L., and S. Wiemer (2010). The influence of tectonic regimes on the earthquake size distribution: A case study for Italy, *Geophys. Res. Lett.* **37,** L10305 , doi: 10.1029/2010GL043066.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** 193–199.

Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). The effect of a mainshock on the size distribution of the aftershocks, *Geophys. Res. Lett.* **45**, no. B1, doi: 10.1029/2018GL080619.

Gulia, L., T. Tormann, S. Wiemer, M. Herrmann, and S. Seif (2016). Short-term probabilistic earthquake risk assessment considering time dependent b values, *Geophys. Res. Lett.* **43,** 1100–1108, doi: 10.1002/2015GL066686.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34,** 185–188.

Hainzl, S. (2016). Rate-dependent incompleteness of earthquake catalogs, *Seismol. Res. Lett.* **87,** no. 2A, 337–344.

Hardebeck, J. L., A. L. Llenos, A. J. Michael, M. T. Page, and N. van der Elst (2019). Updated California aftershock parameters, *Seismol. Res. Lett.* **90,** no. 1, 262–270.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-independent earthquake forecast models for southern California. *Bull. Seismol. Soc. Am.* **96,** no. 1, 90–106, doi: 10.1785/0120050067.

Hobbs, T. E. (2019). Can science tell when a large earthquake will be followed by an even larger one? *Temblor*, doi: 10.32858/temblor.056.

Ishimoto, M., and K. Iida (1939). Observations sur les séismes enregistrés par le microséismographe construit dernièrement, *Bull. Earthq. Res. Inst.* **17,** 443–478 (in French).

Jordan, T. H., W. Marzocchi, A. J. Michael, and M. C. Gerstenberger (2014). Operational earthquake forecasting can enhance earthquake preparedness, *Seismol. Res. Lett.* **85,** no. 5, 955–959.

Marzocchi, W., M. Taroni, and G. Falcone (2017). Earthquake forecasting during the complex Amatrice–Norcia seismic sequence, *Sci. Adv.* **3,** no. 9, e1701239.

Michael, A. J., S. K. McBride, J. L. Hardebeck, M. Barall, E. Martinez, M. T. Page, N. van der Elst, E. H. Field, K. R. Milner, and A. M. Wein (2020). Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 M_w 7.1 Anchorage, Alaska, Earthquake, *Seismol. Res. Lett.* **91,** no. 1, 153–173.

Mignan, A., and J. Woessner (2012). Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi: 10.5078/corssa-00180805.

Narteau, C., S. Byrdina, P. Shebalin, and D. Schorlemmer (2009). Common dependence on stress for the two fundamental laws of statistical seismology, *Nature* **462,** 642–645, doi: 10.1038/nature08553.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** no. 401, 9–27, doi: 10.1080/01621459.1988.10478560.

Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am.* **82,** 1018–1040.

Omi, T., Y. Ogata, K. Shiomi, B. Enescu, K. Sawazaki, and K. Aihara (2019). Implementation of a real-time system for automatic aftershock forecasting in Japan, *Seismol. Res. Lett.* **90,** no. 1, 242–250.

Omori, F. (1895). On the aftershocks of earthquakes, *J. Coll. Sc. Imperial Univ. Tokyo* **7,** 111–200.

Petruccelli, A., P. Gasperini, T. Tormann, D. Schorlemmer, A. P. Rinaldi, G. Vannucci, and S. Wiemer (2019). Simultaneous dependence of the earthquake-size distribution on faulting style and depth, *Geophys. Res. Lett.* **46,** no. 20, 11044–11053, doi: 10.1029/2019GL083997.

Petruccelli, A., D. Schorlemmer, T. Tormann, A. P. Rinaldi, S. Wiemer, P. Gasperini, and G. Vannucci (2019). The influence of faulting style on the size-distribution of global earthquakes, *Earth Planet. Sci. Lett.* **527**, doi: 10.1016/j.epsl.2019.115791.

Reasenberg, P. A., and L. M. Jones (1990). California aftershock hazard forecast, *Science* **247,** 345–346.

Reyes, C. G., and S. Wiemer (2019). *ZMAP7, a refreshed software package to analyze seismicity*, Geophysical Research Abstracts, Vol. 21, EGU2019-13153.

Ross, E. Z., B. Idini, Z. Jia, O. L. Stephenson, M. Zhong, X. Wang, Z. Zhan, M. Simons, E. J. Fielding, S. Yun, *et al.* (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science* **366,** 346–351, doi: 10.1126/science.aaz0109.

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. Seismol. Soc. Am.* **58,** 399–415.

Scholz, C. H. (2015). On the stress dependence of the earthquake *b* value, *Geophys. Res. Lett.* **42,** 1399–1402, doi: 10.1002/2014GL062863.

Schorlemmer, D., M. Werner, W. Marzocchi, T. Jordan, Y. Ogata, D. Jackson, S. Mak, D. Rhoades, M. Gerstenberger, N. Hirata, *et al.* (2018). The collaboratory for the study of earthquake predictability: Achievements and priorities, *Seismol. Res. Lett.* **89,** 1305–1313, doi: 10.1785/0220180053.

Shelly, D. R. (2020). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity, *Seismol. Res. Lett.* 1–8, doi: 10.1785/0220190309.

Shi, Y., and B. Bolt (1982). The standard error of the magnitude-frequency *b* value, *Bull. Seismol. Soc. Am.* **72,** 1677–1687.

Staudenmaier, N., T. Tormann, B. Edwards, A. Mignan, and S. Wiemer (2019). The frequency-size scaling of non-volcanic tremors beneath the San Andreas fault at Parkfield: Possible implications for seismic energy release, *Earth Planet. Sci. Lett.* **516,** 77–107, doi: 10.1016/j.epsl.2019.04.006.

Stein, R. S. (1999). The role of stress transfer in earthquake occurrence, *Nature* **402,** no. 6762, 605–609, doi: 10.1038/45144.

Toda, S., and R. Stein (2003). Toggling of seismicity by the 1997 Kagoshima earthquake couplet: A demonstration of time-dependent stress transfer, *J. Geophys. Res.* **108,** no. B12, 2567, doi: 10.1029/2003JB002527.

Tormann, T., B. Enescu, J. Woessner, and S. Wiemer (2015). Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake, *Nature Geosci.* **8,** 152–158.

Tormann, T., S. Wiemer, B. Enescu, and J. Woessner (2016). Normalized rupture potential for small and large earthquake along the Pacific Plate off Japan, *Geophys. Res. Lett.* **43,** 7468–7477.

Tormann, T., S. Wiemer, S. Metzger, A. Michael, and J. L. Hardebeck (2013). Size distribution of Parkfield's microearthquakes reflects changes in surface creep rate, *Geophys. J. Int.* **193,** no. 3, 1474–1478, doi: 10.1093/gji/ggt093.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.* **84,** 974–1002.

Wessel, P., and W. H. F. Smith (1998). New version of the generic mapping tools released, *Eos. Trans.* **76,** 329.

Wiemer, S. (2001). A software package to analyze seismicity: ZMAP, *Seismol. Res. Lett.* **72,** no. 3, 373–382, doi: 10.1785/gssrl.72.3.373.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan, *Bull. Seismol. Soc. Am.* **90,** no. 4, 859–869, doi: 10.1785/0119990114.

Wiemer, S., and K. Katsumata (1999). Spatial variability of seismicity parameters in aftershock zones, *J. Geophys. Res.* **104,** 13,135–13,151, doi: 10.1029/1999JB900032.

Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** 684–698.

Zechar, J. D., W. Marzocchi, and S. Wiemer (2016). Operational earthquake forecasting in Europe: Progress, despite challenges, *Bull. Earthq. Eng.* **14,** 2459–2469.

Manuscript received 9 October 2019
Published online 8 July 2020

Application of the EEPAS seismic forecasting method to Italy

E. Biondini[1], D. Rhoades[2], P. Gasperini[1,3]

[1]Dipartimento di Fisica e Astronomia, Università di Bologna, Italy

[2] GNS Science P.O. Box 30-368, Lower Hutt 5040, New Zealand

[3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy

emanuele.biondini2@unibo.it, d.rhoades@gns.cri.nz, paolo.gasperini@unibo.it

Summary

The EEPAS (Every Earthquake a Precursor According to Scale) forecasting model is a space–time point-process model based on the precursory scale increase (ψ) phenomenon and associated predictive scaling relations. It has been previously applied to New Zealand, California and Japan earthquakes with target magnitude thresholds varying from about 5 to 7. In all previous application, computations were done using the computer code implemented in Fortran language by the model authors. In this work we applied it to Italy using a suite of computing codes completely rewritten in Matlab and Python. We first compared the two software codes to ensure the convergence and adequate coincidence between the estimated model parameters for a simple region capable of being analyzed by both software codes, then using the rewritten codes we optimized the parameters for a different and more complex polygon of analysis using the catalog data from 1990 to 2011 then we perform a retrospective (pseudo-prospective) forecasting experiment of Italian earthquakes from 2012 to 2021 with Mw≥5.0 and compare the forecasting skill of EEPAS with other forecasting models.

Abbreviated title:

Keywords: Earthquake interaction, forecasting, and prediction, Statistical seismology

Introduction

EEPAS is a seismic forecasting method based on the statistical analysis of seismicity (Rhoades and Evison, 2004). Its basic assumption is that magnitudes and rates of minor seismicity increase before a strong shock. This phenomenon (called $\psi - phenomenon$) was described by Evison and Rhoades (2004) for some region of the world in which high quality seismic catalogs are available. They analyzed 47 earthquakes with magnitude ranging between 5.8 and 8.2 to derive three different empirical relations scale relations, for time, magnitude, and area. These relate the magnitude of mainshock (m_a) with the precursor magnitude (m_m), the precursor time (t_P) and the precursor area (A_p). Such empirical scale relations show that in general the magnitude of precursor events is smaller than the magnitude of mainshock of at least one magnitude unit. EEPAS model consider each earthquake as an individual precursor according to their appropriate magnitude scale, rather than as a possible member of a ψ phenomenon.

The details of the EEPAS method are described in a number of papers (e.g. Rhoades and Evison, 2004, Evison and Rhoades, 2005, Rhoades, 2007, 2011, Rhoades et al., 2020) some of which contains typos that makes the formulation not perfectly identical in all of them. For such reason in Appendix A of the present paper we describe again the method as well as some assumptions made without explicit mentions in previous papers.

We implemented such formulations in a suite of Matlab and Python codes that we first compared with the code EEPSOF (Rhoades, 2021) used in all previous applications of EEPAS methods. The results of this comparison are described in Appendix B and indicate a reasonable agreement between the values of the parameters estimated by the two computational approaches, even if not a perfect coincidence due to some differences in the calculation methods adopted, and in particular in the numerical algorithms used by the two codes for spatial integration.

For comparison purposes we also consider other forecasting models and in particular the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1989, 1998) and two time-independent forecasting models: the Spatially Uniform Poisson (SUP) and the Spatially Variable Poisson (SVP) models

(Console et al., 2006). The description of our implementations of such models is reported in Appendix C.

The fitting of the free parameters of various models is carried out by maximizing the log-likelihood function of an inhomogeneous Poisson point process, which is given by:

$$\ln L = \sum_{t_i \in (t_a, t_b); m_i \geq m_T; (x_i, y_i) \in R} \ln \lambda(t_i, m_i, x_i, y_i) - \int_{t_a}^{t_b} \int_{m_T}^{m_u} \iint_R \lambda(t, m, x, y)\, dy\, dx\, dm\, dt \tag{1}$$

Where $\lambda(t, m, x, y)$ is the rate density function for PPE (eq. A9), EEPAS (eq. A7), ETAS (eqs. C7a, C7b), SUP and SVP (eq. C8) models. (t_a, t_b) is the time interval of the fitting period, (m_T, m_u) is the magnitude range of target earthquakes and R is the spatial region of analysis.

Application to Italy

We chose as target shocks threshold $m_T = 5.0$ because in Italy such earthquakes potentially cause damage to buildings and threat the health and the life of inhabitants. Such choice is also consistent with most of the applications of EEPAS model to other regions of the World (Rhoades and Evison 2004, Evison and Rhoades 2005, Rhoades 2007, 2011).

We chose the learning time interval from 1990 to 2011 for fitting the EEPAS model as the accuracy and completeness of the Italian catalog improved significantly since 1990 (Gasperini et al., 2013) and use the ten years interval from 2012 to 2021 for the retrospective testing of the model.

As application region R, we consider a regular tessellation of the Italian territory made of square cells with side $L = 30\sqrt{2}$ km from 7°E to 19°E in longitude and from 36°N to 47°N in latitude. The choice of the side size is made, for compatibility with our previous works by Gasperini et al., (2021), so that each square cell is (almost) perfectly inscribed in a circular cell with radius of 30 km like those used in the latter paper. Because the completeness of the seismic catalogue is poor in offshore areas, according to Gasperini et al., (2021), we consider only the cells within which at least one earthquake with $M \geq 4.0$ occurred inland from 1600 to 1959 according to the CPTI15 catalogue (Rovida et al., 2020) and from 1960 to 2021 according to the HOmogenized instrRUmental Seismic (HORUS) catalog (Lolli et al., 2020). We also excluded the cells that are not contiguous to the main analysis polygon (such as insulated cells on islands). At the end, the square cells that constitute the region of analysis R are in all 177 (Fig. 1).

For model parameters fitting, a seismic catalog, characterized by a completeness magnitude (m_c) of at least two units lower than the target magnitude (m_T) is desirable (Rhoades and Evison, 2004). For Italy a seismic catalog with homogeneous magnitudes and high resolution is the HORUS catalog (Lolli et al., 2020) reporting earthquakes from 1960 to present. According to Lolli et al. (2020), HORUS can be considered complete within the Italian mainland for $m \geq 4.0$ since 1960, for $m \geq 3.0$ since 1981, for $m \geq 2.5$ since 1990, for $m \geq 2.1$ since 2003 and for $m \geq 1.8$ since 2005. As dataset for this work, we used only shallow earthquakes with depth $Z \leq 40\ km$. To avoid edge effects in the fitting of model parameters, the contribution of earthquakes in the neighborhood of the region R must also be considered (Rhoades and Evison, 2004). We assume as neighborhood region the area included in the CPTI15 polygonal (Fig. 2) according to Rovida et al. (2020).

To account for the limited accuracy of magnitude data, we binned all magnitudes to the nearest tenth of units:

$$m_{binned} = \frac{\mathrm{int}(m_{raw} \times 10 + 0.5)}{10} \tag{2}$$

This also means that all magnitude lower thresholds (e.g. $m_T \geq 5.00$) have to be intended 0.05 units smaller ($m_T \geq 4.95$).

The HORUS catalog reports 27 target shocks with Mw ≥ 5.0 from1990 to 2011 and 27 from 2012 to 2021. This indicates that the rate of target shocks in the testing period is about twice than in the learning period. Hence, the forecasting of the correct number of earthquakes by all forecasting methods will be hard.

After the first target shock (“mainshock”) of a seismic sequence, the forecasting of successive target shock (“aftershocks”) is easier, owing to the presence of small aftershocks (Gasperini et al., 2021). Hence, we also consider a declustered set of target shocks where all the target shocks occurred within

50 km and one year after any other target shock are removed. This reduces the number of considered target shocks for the declustered dataset to 12 and 9 for the learning and testing time intervals respectively.

Estimation of EEPAS model parameters

Considering the high number of free parameters to be determined for EEPAS model (in principle about 20), the maximization of the log-likelihood function (1) would be very time consuming and subject to numerical instability. However, according to Rhoades and Evison (2004), simultaneous optimization of all parameters is not necessary because some of them, such as the b-value of the Gutenberg and Richter (1944) relation and the parameters of the aftershock epidemic decay model (p, k, c, ν), can be, in fact, separately fitted or even be simply assigned based on previous works in the same area.

The $b-$value of the Gutenberg and Richter (1944) relation is chosen so that to be representative of the behavior of the frequency magnitude distribution of target events in the fitting time interval. For the undeclustered and declustered sets the values $b = 1.084$ and $b = 1.176$ respectively make the number of predicted target events as close as possible to the observed numbers (27 and 12 respectively). These values are kept fixed even for other forecasting models, computed for comparison.

The parameters of the aftershock model are not particularly critical for the EEPAS model, however they are necessary to determine the weight w (eq. A18) of the contribution of each earthquake ($M \geq m_c$) by defining the probability with which an earthquake can be defined an aftershock of a previous seismic event. The parameters $p = 1.2$ and $c = 0.03$ of equation (A15) were chosen as typical parameters of Omori's law (Ogata 1983). The two parameters ν and k in equation (A13) were fitted by maximizing the likelihood with earthquakes with $m \geq m_T$ occurred within R in the period 1990-2011.

Finally, the parameter $\sigma_U = 0.006$ of equation (A17) is chosen so that to be consistent with the mean value of the cluster diffusion parameter for Italy (Musmeci and Vere-Jones, 1992). The parameter $\delta = 0.7$ of equation (A16) is taken from previous works for New Zealand, California and Japan (Evison and Rhoades 2005, Rhoades and Evison, 2004, Rhoades, 2007, 2011). The parameters of the PPE model (13) a, d, and s are fitted simultaneously using the maximum likelihood method.

Regarding EEPAS parameters, the fit is made in three successive iterations. The parameter b_M is fixed to 1 for all three iterations that means the perfect scaling between precursor and target magnitudes (Rhoades and Evison, 2004).

In the first iteration, the parameters b_T and b_A are fixed to 0.40 and 0.35, respectively, based on analyses conducted on scaling relationships obtained from the analysis of individual earthquakes. The parameters σ_M and σ_T are also fixed to 0.32 and 0.23, respectively. Such values correspond to the residual standard deviation for the magnitude and time scaling relations (Rhoades and Evison, 2004). Finally, parameters a_T, a_M, σ_A and μ are computed by maximum likelihood estimation.

In the second iteration, the previously fitted parameters a_T, a_M, σ_A are kept fixed at the obtained values and the parameters b_T, b_A, σ_M,σ_T and μ are computed instead by the maximum likelihood.

In the third and last iteration a final computation is made of all parameters (a_T, a_M, σ_A, b_T, b_A, σ_M,σ_T and μ) simultaneously providing the optimizer with starting values of the parameters as obtained in previous optimizations. The parameter μ, responsible for mixing the two models PPE and EEPAS is the only parameter fitted in all three iterations of optimization.

The parameters of the PPE and EEPAS (unweighted and weighted) obtained by maximizing the likelihood are reported in Table 1 and 2 for the undeclustered (mainshocks+aftershocks) and declustered (mainshocks only) target sets respectively. In the same tables we also report the parameters of the other forecasting models (SUP, SVP, ETAS-SUP and ETAS-SVP) computed for comparison.

In table 3 and 4 we report the goodness of fit estimators of various model for the undeclustered and declustered target sets respectively. We can note how both the ETAS models have better scores (higher loglikelihood, information gain per event and G and lower AIC) than EEPAS and other models. Both EEPAS models also have lower loglikelihoods than SVP and for the declustered set higher AIC (worse) than SUP. Such scores are not particularly significant because only represent the goodness of the fit of the various forecasting models with the learning dataset and then might include some degree of overfitting of the learning dataset.

Retrospective comparison of forecasting models with the testing dataset

We apply the suite of tests defined by the Collaboratory for the Study of Earthquake Predictability (CSEP, Jordan, 2006, Zechar et al., 2010) and particularly the new ones described by Bayona et al., (2022).

Such tests assess the consistency of observed earthquakes with a forecast model by i) the conditional loglikelihood (cL-test) ii) the observed number of earthquakes (N-test), iii) their spatial distribution (S-test) and iv) their magnitude distribution (M-Test). However, we do not report the results for the latter, because all forecasting models assume a Gutenberg–Richter frequency–magnitude distribution and then all pass the M-test.

Traditional CSEP tests are based on a likelihood function that approximates earthquakes in individual cells or bins as independent and Poisson distributed (Schorlemmer et al., 2007, 2010, Zechar et al., 2010). However, the Poisson distribution insufficiently captures the spatiotemporal variability of earthquakes, especially in the presence of clusters of seismicity (Werner and Sornette, 2008, Lombardi and Marzocchi, 2010, Nandan et al., 2019). Hence, the new CSEP tests are based on the negative binomial distribution (NBD) that reduces the sensitivity of CSEP evaluations to clustering

$$p\big((\omega|\tau,\nu)\big) = \frac{\Gamma(\tau+\omega)}{\Gamma(\tau)\omega!}\nu^{\omega}(1-\nu)^{\omega} \qquad (3)$$

Where $\omega = 1,2,\ldots$ is the number of events, $\tau > 0$ and $0 \leq \nu \leq 1$ are parameters and Γ is the Gamma function. The mean and the variance of NBD are given by

$$\mu = \tau\frac{1-\nu}{\nu};\ \sigma^2 = \frac{1-\nu}{\nu^2} \tag{4}$$

If the confidence level goes below a given threshold, the model fails to describe satisfactorily the observed data even if it might be considered for comparison with other models if all models fail some consistency test.

The N-test compares the number of predicted earthquakes in all (time–space–magnitude) bins with the number of target earthquakes observed.

The binary cL-test compares the joint binary log-likelihood of the forecasting model with the really observed seismicity, with the distribution of joint binary log-likelihoods obtained by the simulation of random catalogs consistent with the forecasted one. If the former lies in the lower tail of the random binary log-likelihood distribution, the forecasting model does not reproduce well the real seismicity pattern and then the test fails. The binary log-likelihood is obtained by calculating the probability of an earthquake in a forecast bin rather than that of observing one or more earthquakes. The probability of observing $\omega = 0$ events, given an expected events number or rate λ, is $P_0 = \exp(-\lambda)$, while the probability of observing more than zero events is $P_1 = 1 - P_0$ (Bayona et al., 2022). The binary log-likelihood for each bin is thus given by

$$BLL = X_i \ln(1 - \exp(-\lambda)) + (1 - X_i)\ln(\exp(-\lambda)) \tag{5}$$

Where X_i represents the contribution to the log-likelihood score if the $i-th$ bin contains one or more events. The value of $X_i = 1$ if the $i-th$ bin contains at least one event, on the contrary its value is 0. The observed binomial joint log-likelihood is given by the summation of the BLL over all space-magnitude-time bins:

$$JBLL = \sum_{l=1}^{s}\sum_{j=1}^{m}\sum_{k=1}^{t} X(l,j,k)\ln\big[1 - \exp\big(-\lambda(l,j,k)\big)\big] + [1 - X(l,j,k)\ln(\exp(-\lambda(l,j,k)] \tag{6}$$

The S-test evaluates the consistency of the spatial occurrence of target earthquakes regardless of their magnitudes. For the S-test the joint binary log-likelihood is calculated by first normalizing the forecast rate to the number of active cells.

For the cL, N and S test, the computed statistics is the quantile score that is the fraction of simulated likelihoods less or equal to the likelihood observed by the model. A small value, lower than the Bonferroni-adjusted significance level $\alpha = 0.05/2 = 0.025$, means that the model inadequately describes the seismicity pattern.

To evaluate the relative skill of the forecasting models, we use the information gain per earthquake (IGPE, Rhoades et al., 2011), or per active bin (IGPA, Bayona et al., 2022), which are based on the likelihood difference with respect to a reference (baseline) forecasting model divided by the number of earthquakes or by the number of active bins (the bins in which the likelihood contribution is not zero) respectively. The IGPA is thus given by

$$IGPA = \frac{N_{base} - N_{mod}}{M} + \frac{1}{M}\sum_{m=1}^{M}[X_{mod}(m) - X_{base}(m)] \tag{7}$$

Where N_{base} and N_{mod} are the total number of earthquakes expected by the baseline and the model respectively, M is the number of active bin, and $X_{mod}(m)$ and $X_{base}(m)$ are the joint log-likelihood score obtained in the bin with the $m-th$ target earthquake by the model and the reference baseline model respectively. According to Rhoades at al. (2011) the variance of $X_{mod}(m) - X_{base}(m)$ is given by

$$s^2 = \frac{1}{M-1}\sum_{m=1}^{M}(X_{mod}(m) - X_{base}(m))^2 - \frac{1}{M^2 - M}\left[\sum_{m=1}^{M} X_{mod}(m) - X_{base}(m)\right]^2. \tag{8}$$

The IGPA error is estimated as $\pm ts\sqrt{M}$, where t is the 95th percentile of the Student's t inverse cumulative distribution with $M-1$ degrees of freedom.

As baseline model we take the SUP, which is the simpler one. We do not use the correction for the number of free parameters as proposed by Rhoades et al., (2014), because the fitting of models is independent on the testing set targets being made using the learning set. In addition, we do not use the parimutuel gambling score (PGS) by Zhuang (2010) and Zechar and Zhuang (2014), because Serafini et al. (2022) recently demonstrated that PGS is improper when the number of forecasting methods being tested is greater than two.

In Fig. 3 and Table 5 we report the numbers of undeclustered targets (mainshock+aftershocks) predicted by various models using different time intervals (3 months, 6 months, 1 year, 5 years, and 10 years) of prediction. All models definitely underestimate the total number of targets (27) actually occurred. The reason is that the average rate of targets in the testing set (about 2.7 per year) is more than twice than that in learning set (about 1.2 per year).

This strong difference influences the results of the cL-Test (Fig. 4 and Table S1) and N-Test (Fig. 5 and Table S2) that rejects all models for all time intervals. Conversely, the S-Test (Fig. 6 and Table S3) is passed by all models for all time intervals.

The results of the IGPA (T-test) for undeclustered targets in Fig. 7 and Table S4 indicate that most preferable models are the ETAS-SVP and ETAS-SUP for the shortest prediction interval of 3 months and the EEPAS-NW and EEPAS-W for longer prediction intervals. Looking also at the confidence intervals, the preference appears statistically significant only for time intervals of 5 or 10 years.

In Fig. 8 and Table 6 we report the numbers of declustered targets (mainshock only) predicted by various models using different time intervals of prediction. All models still underestimate the total number of targets (9) actually occurred as even in this case the average rate of targets in the testing set (0.9 per year) is smaller than in the learning set (0.5 per year).

The cL-Test of consistency (Fig. 9 and Table S5) is not passed by any model for any time interval whereas the N-Test (Fig. 10 and Table S6) is passed by most model for most time intervals (only excluding ETAS-SUP and ETAS-SVP for longer time intervals). The S-Test (Fig. 11 and Table S7) is passed by all models for all time intervals.

The results of the IGPA (T-test) for declustered targets in Fig. 12 and Table S8 confirm that the most preferable models are the ETAS-SVP and ETAS-SUP for the shortest prediction interval of 3 months and the EEPAS-NW and EEPAS-W for longer prediction intervals. In this case such preferences appear statistically not significant at any time intervals.

Conclusions

We applied the EEPAS seismic forecasting model to Italy similarly to previous application in other seismic regions of the world (e.g., Rhoades and Evison, 2004, Evison and Rhoades, 2005, Rhoades, 2007, 2011, Rhoades et al., 2020) using a suite of computing codes completely rewritten in Matlab and Python and implementing both the EEPAS formulations with the not weighted (EEPAS-NW) and weighted (EEPAS-W) seismicity. We calibrated and fitted the model parameters using earthquakes of HORUS seismic catalogue of Italy (Lolli et al., 2020) for the learning period 1990-2011. The EEPAS model was then applied to forecast all earthquakes (mainshocks + aftershocks) of the same seismic catalogue with $M \geq 5.0$ and only the mainshocks occurred within the polygon of analysis for the test period 2012-2021. We compared the forecasting skill of EEPAS with the ones obtained by other time dependent (ETAS-SUP and ETAS-SVP) and time independent (SUP, SVP and PPE) models implemented on the same dataset. We used a set of new CSEP consistency test based on a binary likelihood function as described in Bayona et al., (2022). This latter reduces the sensitivity of spatial log-likelihood scores to the occurrence of seismic events (Bayona et al., 2022) with respect to previous versions of the tests based on a Poisson distribution assumption. The number of expected target earthquakes by each model compared tends to decrease as the forecasting interval increase. The highest expected number of earthquakes is for a window 3 months. However, all models tend to underestimate the numbers the total expected events actually occurred, 27 and 9 for the mainshock + aftershock and mainshock only datasets, respectively. This is due to the different average rate of target events in the learning and testing period. Such difference affects the performance of consistency tests, in particular for the not declustered data set. In fact, for this latter, there is no consistency between the log-likelihood and the numbers of expected events obtained by the model with the observed ones. For this reason, the cL and N tests failed for all models and for all forecasting intervals. On the contrary, all models passe the binary S-test and describes correctly the seismicity spatial pattern of target events. The difference in the seismicity rate between the learning and the testing period is less pronounced for the mainshock only dataset. This allows to pass all consistency test (cL, N, and S test) by almost all models compared for all forecasting intervals. The only exceptions are the ETAS-SUP and the ETAS-SVP that for forecasting interval larger than 3-6 months fail the cL and the binary S tests, respectively. We also assess the relative forecasting skill of various model using the IGPA (Rhoades et al., 2011, Bayona et al., 2022) considering as baseline reference model the SUP. For both mainshocks+aftershock and mainshocks only datasets, the most preferable model is the ETAS-SUP and ETAS-SVP for the shortest forecasting interval of 3 months and the EEPAS-NW and EEPAS-W for the longer prediction intervals. These results confirm the different characteristics of the models ETAS and EEPAS. This latter rather than the ETAS models, is in fact more appropriate to make forecasts analyzing the long-term seismicity.

Acknowledgements

This paper benefitted from funding provided by the European Union within the ambit of the H2020 project RISE (No. 821115), which in particular fully financed the PhD grant of one of the authors (E.B.)

References

Antonioli, A., Belardinelli, M.E. & Cocco, M,. 2004. Modelling Dynamic Stress Changes Caused by an Extended Rupture in an Elastic Stratified Half-Space, *Geophys. J. Int.,* 157 (1): 229–44.

Båth, M., 1965. Lateral Inhomogeneities of the Upper Mantle. *Tectonophysics* 2 (6): 483–514, doi: 10.1016/0040-1951(65)90003-X.

Bayona, J. A., Savran, W. H., Rhoades, D. A., & Werner, M. J., 2022. Prospective evaluation of multiplicative hybrid earthquake forecasting models in California. *Geophs. J. Int.*, *229*(3), 1736–1753, doi: 10.1093/gji/ggac018.

Byrd, R., Hribar, M. E., & Nocedal, J., 1999. An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM Journal on Optimization*, *9*(4), 877–900.

Console, R., & Murru, M., 2001. A simple and testable model for earthquake clustering. *J. Geophys. R.*, *106*(B5), 8699–8711, doi: 10.1029/2000jb900269.

Console, R., Murru, M., & Catalli, F., 2006. Physical and stochastic models of earthquake clustering. *Tectonophysics*, *417*(1–2), 141–153, doi: 10.1016/j.tecto.2005.05.052.

Evison, F., & Rhoades, D., 2004. Long-term seismogenesis and self-organized criticality. *Earth, Planets and Space*, *56*(8), 749–760, doi: 10.1186/BF03353083.

Evison, F., & Rhoades, D., 2005. Multiple - mainshock events and long - term seismogenesis in Italy and New Zealand. *New Zealand Journal of Geology & Geophysics*, *48*, 523–536, doi: 0028-8306/05/4803-0523.

Frankel, A., 1995. Mapping seismic hazard in the central and eastern United States. *Seismol. Res. Lett.*, *66*(4), 8–21. Doi: 10.1785/gssrl.66.4.8.

Gasperini, P, Biondini, E., Lolli, B., Petruccelli, A., & Vannucci, G., 2021. Retrospective short-term forecasting experiment in Italy based on the occurrence of strong (fore) shocks. *Geophys. J. Int.*, 1192–1206, doi: 10.1093/gji/ggaa592.

Gasperini, P., Lolli, B., & Vannucci, G., 2013. Empirical calibration of local magnitude data sets versus moment magnitude in Italy. *Bull. Seism. Soc. Am.*, *103*(4), 2227–2246, doi: 10.1785/0120120356.

Gutenberg, R., & Richter, C., 1944. *Frequency of earthquakes in California.* Bull. Seism. Soc. Am. 34(4), 185-188, doi:10.1785/BSSA0340040185.

Jackson, D. D., & Kagan, Y. Y., 1999. Testable Earthquake Forecasts for 1999. *Seismol. Res. Lett.*, *70*(4), 393–403, doi: 10.1785/gssrl.70.4.393.

Jordan, T. H., 2006. Earthquake Predictability, Brick by Brick. *Seismol. Res. Lett.*, *77*(1), 3–6, doi: 10.1785/gssrl.77.1.3.

Lolli, B., Randazzo, D., Vannucci, G., & Gasperini, P., 2020. The homogenized instrumental seismic catalog (HORUS) of Italy from 1960 to present. *Seismol. Res. Lett.*, *91*(6), 3208–3222, doi: 10.1785/0220200148.

Lombardi, A. M., & Marzocchi, W., 2010. The ETAS model for daily forecasting of italian seismicity in the CSEP experiment. *Ann. Geophys., 53*(3), 155–164, doi: 10.4401/ag-4848.

Musmeci, F., & Vere-Jones, D., 1992. A space-time clustering model for historical earthquakes. *Ann. Inst. Stat. Math.*, *44*(1), 1–11, doi: 10.1007/BF00048666.

Nandan, S., Ouillon, G., & Sornette, D., 2019. Magnitude of Earthquakes Controls the Size Distribution of Their Triggered Events. *J. Geophys. Res.*, *124*(3), 2762–2780 doi: 10.1029/2018JB017118.

Nelder, J. A., & Mead, R., 1965. A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313, doi: 10.1093/comjnl/7.4.308.

Ogata, Y., 1983. Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. *Journal of Physics of the Earth*, *31*(2), 115–124, doi: 10.4294/jpe1952.31.115.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.*, *83*(401), 9–27, doi: 10.1080/01621459.1988.10478560.

Ogata, Y., 1989. Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, *169*(1–3), 159–174, doi: 10.1016/0040-1951(89)90191-1.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences. *Ann. Inst. Stat. Math.*, *50*(2), 379–402, doi: 10.1023/A:1003403601725.

Rhoades, D. A., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., Schorlemmer, D., Werner, M. J., & Jordan, T. H., 2014. Regional earthquake likelihood models II: Information gains of multiplicative hybrids. *Bull. Seism. Soc. Am.*, *104*(6), 3072–3083, doi: 10.1785/0120140035.

Rhoades, D.A. (2021). MANUAL FOR EEPAS SOFTWARE (EEPSOF) VERSION 2.3w. *GNS Science Internal Report*, 49pp.

Rhoades, D. A., 2011. Application of a long-range forecasting model to earthquakes in the Japan mainland testing region. *Earth, Planets and Space*, *63*(3), 197–206, doi: 10.5047/eps.2010.08.002.

Rhoades, D. A., & Evison, F. F., 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale. *Pure Appl. Geophys.*, *161*(1), 47–72, doi: 10.1007/s00024-003-2434-9.

Rhoades, D. A., Rastin, S. J., & Christophersen, A., 2020. The effect of catalogue lead time on medium-term earthquake forecasting with application to New Zealand data. *Entropy*, *22*(11), 1–15, doi: 10.3390/e22111264.

Rhoades, D. A., 2007. Application of the EEPAS Model to forecasting Earthquakes of moderate magnitude in Southern California. *Seismol. Res. Lett.*, *78*(1), 110–115, doi: 10.1785/gssrl.78.1.110.

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., & Imoto, M., 2011. Efficient Testing of Earthquake Forecasting Models. *Acta Geophysica*, *59*(4), 728–747, doi: 10.2478/s11600-011-0013-5.

Rhoades, D. A., Rastin, S. J., & Christophersen, A., 2020. The effect of catalogue lead time on medium-term earthquake forecasting with application to New Zealand data. *Entrophy,* 22(11), 1-15, doi: 10.3390/e22111264.

Rovida, A., Locati, M., Camassi, R., & Lolli, B., 2020. The Italian earthquake catalogue CPTI15. In *Bull. Earth. Eng., 18, 2953-2984,* doi: 10.1007/s10518-020-00818-y.

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A., 2007. Earthquake likelihood model testing. *Seismol. Res. Lett.*, *78*(1), 17–29, doi:10.1785/gssrl.78.1.17.

Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., & Jordan, T. H., 2010. First Results of the Regional Earthquake Likelihood Models Experiment. *Pure Appl. Geophys.*, *167*(8–9), 859–876, doi: 10.1007/s00024-010-0081-5.

Serafini, F., Naylor, M., Lindgren, F., Werner, M. J., & Main, I., 2022. Ranking earthquake forecasts using proper scoring rules: binary events in a low probability environment. *Geophys. J. I.*, *230*(2), 1419–1440, doi: 10.1093/gji/ggac124.

Utsu, T., 1961. A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, *30*, 521–605.

Werner, M. J., & Sornette, D., 2008. Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments. *J. Geophys. Res.*, *113*(8), 1–20, doi: 10.1029/2007JB005427.

Zechar, J. D., Gerstenberger, M. C., & Rhoades, D. A., 2010. Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts. *Bull. Seismol. Soc. Am.*, *100*(3), 1184–1195, doi: 10.1785/0120090192.

Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., & Jordan, T. H., 2010. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. Concur. Comp. Pract. Exp, *22*(12), 1836–1847, doi: 10.1002/cpe.1519.

Zechar, J. D., & Zhuang, J., 2014. A parimutuel gambling perspective to compare probabilistic seismicity forecasts. *Geophys. J. I.*, *199*(1), 60–68, doi: 10.1093/gji/ggu137.

Zhuang, J., Ogata, Y., & Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc*, *97*(458), 369–380, doi: 10.1198/016214502760046925.

Zhuang, J., 2010. Gambling scores for earthquake predictions and forecasts. *Geophys. J. I.*, *181*(1), 382–390, doi: 10.1111/j.1365-246X.2010.04496.x.

Appendix A - Formulation of PPE and EEPAS forecasting models

In the EEPAS model each i-th earthquake, occurred at time t_i with magnitude m_i and located at (x_i, y_i), is assumed to contribute to the transient increment of the rate density $\lambda(t, m, x, y)$ of future seismicity (defined as the derivative of the expected number of earthquakes with respect to time, magnitude and location coordinates) by the term

$$\lambda_i(t, m, x, y) = w_i f_{1i}(t) g_{1i}(m) h_{1i}(x, y) \tag{A1}$$

Where w_i is a weighting factor which depends on other earthquakes in its proximity (see below). $f_{1i}(t)$, $g_{1i}(m)$ and $h_{1i}(x, y)$ are the probability density functions of time, magnitude and location respectively. The assumed forms for these distributionsare defined consistently with the ψ scaling relations by Rhoades and Evison (2004). The time distribution is assumed to be Log-Normal with the form

$$f_{1i}(t) = \frac{H(t - t_i)}{(t - t_i) \ln(10)\, \sigma_T \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log(t - t_i) - a_T - b_T m_i}{\sigma_T}\right)^2\right] \tag{A2}$$

where $H(t - t_i)$ is the Heaviside step function, which value is 1 if $t - t_i > 0$, or 0 otherwise. This means that at the time t, the rate density function is contributed only by earthquakes occurred before t. a_T, b_T and σ_T are free parameters to be determined.

The magnitude distribution $g_{1i}(m)$ is assumed to be Normal with the form:

$$g_{1i}(m) = \frac{1}{\sigma_M \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m - a_M - b_M m_i}{\sigma_M}\right)^2\right] \tag{A3}$$

where a_M, b_M and σ_M are free parameters.

The space distribution is assumed to be bivariate Normal with circular symmetry with the form

$$h_{1i}(x, y) = \frac{1}{2\pi \sigma_A^2 10^{b_A m_i}} \exp\left[-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma_A^2 10^{b_A m_i}}\right] \tag{A4}$$

where σ_A and b_A are free parameters.

An adjustment is necessary because of the missing contribution of earthquakes below the minimum completeness magnitude m_c which underestimates on average the rate density at magnitude m by a fraction $\Delta(m)$ of its real value given by

$$\Delta(m) = \phi\left(\frac{m - a_M - b_M m_c - \sigma_M^2 \beta}{\sigma_M}\right) \tag{A5}$$

where ϕ is the Normal distribution integral. Then $\Delta(m)$ can also be written as

$$\Delta(m) = \frac{1}{2}\operatorname{erf}\left[\left(\frac{m - a_M - b_M m_c - \sigma_M^2 \beta}{\sigma_M \sqrt{2}}\right) + 1\right] \tag{A6}$$

where erf is the Error function.

Hence, to compensate the lack of earthquakes with magnitude lower than the completeness magnitude m_c, $\lambda_i(t, m, x, y)$ is then inflated by a factor $\frac{1}{\Delta(m)}$.

The total rate density is obtained by summing the contribution of all past earthquakes and also adding a background term that allows for the possibility that an earthquake can occur without an appreciable scale increase of precursory shocks:

$$\lambda(t, m, x, y) = \mu\lambda_0(t, m, x, y) \quad + \sum_{t_i > t_0; m_i > m_c}^{t - delay} \eta(m_i)\lambda_i(t, m, x, y) \tag{A7}$$

where $\lambda_0(t, m, x, y)$ is the background rate density, t_0 is the time of the beginning of the catalogue, μ is the mixing parameter and can be interpreted as the proportion of earthquakes that occur without precursory shocks. η is a normalizing function which is defined by

$$\eta(m_i) = \frac{b_M(1 - \mu)}{E(w)} \exp\left[-\beta\left(a_M + (b_M - 1)m_i + \frac{\sigma_M^2 \beta}{2}\right)\right] \tag{A8}$$

where $E(w)$ is the mean weight of earthquakes in the catalogue, a_M, b_M and σ_M are the parameters defined in eq. (A3) and $\beta = b \ln 10$, with b being the slope of the frequency-magnitude distribution of Gutenberg and Richter (1944). $\eta(m_i)$ ensures that the number of earthquakes expected by the model approximatively matches the actual number of target earthquakes. The *delay* term in equation (A7) avoids that the fit of the parameters is influenced by the short-term clustering of earthquakes (such as aftershocks and swarms). The EEPAS model, in fact, is focused on the long-term clustering detected by the precursory scale increase phenomenon and its associated scale relationships. For this reason, such delay (usually assumed of 50 days) after the time of occurrence of each earthquake is applied and no earthquake from the input catalogue is considered before such time interval elapsed after a target shock.

The background rate density $\lambda_0(t, m, x, y)$ depends on the proximity of the site (*x*, y) with respect to previous seismicity. It is described by a quasi-time-invariant smoothed seismicity model, described by Rhoades and Evison (2004), which is similar to the forecasting model proposed by Jackson and Kagan (1999) and is called PPE (Proximity to Past Earthquakes). It takes the form

$$\lambda_0(t, m, x, y) = f_{0i}(t) g_{0i}(m) h_{0i}(x, y) \tag{A9}$$

where $f_{0i}(t)$ is the time density function, $g_{0i}(m)$ is the magnitude density function and $h_{0i}(x, y)$ is the space density function. The time density function takes the form

$$f_{0i}(t) = \frac{1}{t - t_0} \tag{A10}$$

This ensures that at any time the estimated rate of earthquakes with $m \geq m_T$ within the region R is similar to the past rate.

The magnitude density function is that implied by the frequency magnitude law of Gutenberg and Richter (1944):

$$g_{0i}(m) = \beta \exp[\beta(m - m_c)] \tag{A11}$$

Finally, $h_{0i}(x, y)$ is the sum over all earthquakes with $m_i \geq m_T$ from time t_0 up to, but not including time t of smoothing kernels with the form

$$h_{0i}(x_i, y_i) = \sum_{t_i > t_0; m_i > m_T}^{t-delay} a(m_i - m_T) \frac{1}{\pi} \left(\frac{1}{d^2 - r_i^2} \right) + s \tag{A12}$$

where r_i is the distance in km between (x, y) and the epicenter (x_i, y_i); a is a normalizing parameter, d is a smoothing distance and s is a small term that includes the contribution of earthquakes occurred far from past epicenters.

The rate density $\lambda_0(t, m, x, y)$ of the PPE model decreases gradually with time elapsed after an earthquake occurrence and increases when a new earthquake occurs. The function $h_{0i}(x, y)$ takes into account the earthquake location and the function $f_{0i}(t)$ the passage of time.

The purpose of the weighting factor w_i in eq. (A1) is to give more weight to earthquakes that are more likely to be part of a long-term clustering, thus giving less weight to events that are aftershocks of previous earthquakes. Two different weighting strategies were applied in the past application of EEPAS. The simplest one is giving the same weight $w_i = 1$ to each earthquake in the catalogue. With this strategy aftershocks triggered by previous earthquakes have the same weight of any other shock. The other strategy is to assign a lower weight to any earthquake which is likely to be an aftershock of a previous earthquake. Therefore, the total rate density is mostly given by earthquakes that are part of long-term clustering.

This latter strategy requires estimating the rate density λ' for aftershock occurrence, incorporating epidemic-type aftershock behavior (Ogata, 1988, 1998, Console and Murru, 2001). The aftershock model adopted for EEPAS takes the form:

$$\lambda'(t, m, x, y) = \nu\lambda_0(t, m, x, u) + k \sum_{t_i \geq t_0} \lambda_i'(t, m, x, y) \tag{A13}$$

Where λ_0 is the rate density given by PPE model, ν is the proportion of earthquake that are not aftershocks, k is a normalization constant and $\lambda'(t, m, x, y)$ describes the aftershocks occurrence with the form:

$$\lambda_i'(t, m, x, y) = f_{2i}(t) g_{2i}(m) h_{2i}(x, y) \tag{A14}$$

Where $f_{2i}(t)$, $g_{2i}(m)$ and $h_{2i}(x, y)$ are respectively the density rate functions for time, magnitude, and locations of the aftershocks of the i-th earthquake. The time distribution is given by the modified Omori law (Utsu, 1961, Ogata, 1983):

$$f_{2i}(t) = H(t - t_i) \frac{p - 1}{(t - t_i + c)^p} \tag{A15}$$

Where t_i is the time of the i-th earthquake, c and p are the Omori law parameters.

The magnitude distribution follows the Gutenberg and Richter (1944) law, and it is assumed that the magnitude of an aftershock is smaller than its mainshock by at least δ units

$$g_{2i}(m) = H(m_i - \delta - m)\beta \exp[\beta(m - m_i)] \tag{A16}$$

The addition of the parameter δ is based on the so-called Bath's law (Båth, 1965), according to which the largest aftershock typically has a magnitude about 1.2 units smaller than the mainshock. Finally, the spatial distribution is assumed to be Normal bivariate with circular symmetry:

$$h_{2i}(x, y) = \frac{1}{2\pi\sigma_U^2 10^{m_i}} \exp\left[-\frac{(\mathrm{x} - \mathrm{x_i})^2 + (\mathrm{y} - \mathrm{y_i})^2}{2\sigma_\mathrm{U}^2 10^{\mathrm{m_i}}}\right] \tag{A17}$$

Where σ_U is a free parameter. The weighting factor is then computed as

$$w_i = \frac{\nu\lambda_0(t_i, m_i, x_i, y_i)}{\lambda'(t_i, m_i, x_i, y_i)} \tag{A18}$$

In this way if an earthquake has the characteristics of an aftershock will have a weight close to 0, on the contrary if an earthquake that in no way resembles an aftershock will have a weight close to 1.

Appendix B Implementation of Matlab and Python codes and comparison with EEPSOF Version 2.3w

We developed a suite of codes in Matlab and Python languages reproducing the formulation described in the Appendix A. We tested them vs the EEPSOF code (Version 2.3w) developed by D. A. Rhoades (Rhoades, 2021) and provided as binary Linux executable file compiled by Fortran77. To make the comparison, we adopted a simplified spatial geometry as the EEPSOFT hardly manage the complex shape made by a tessellation of the Italian area as described in the main text. The purpose of the comparison is to ensure that the optimized parameters value and the relative maximum log-likelihoods are satisfactorily similar.

One difference between the Matlab implementation and EEPSOF is the treatment of spatial data. While the EEPSOF code computes itself the kilometric distances directly from geographical coordinates, for the Matlab implementation we have chosen to preliminary convert all coordinates from the WGS84 geographic reference to kilometric coordinates in the RDN2008 Italy Zone (E-N) EPSG: 7794 by the QGIS software.

We applied both codes to the dataset of target earthquakes with magnitude $\mathrm{M} \geq 5.0$ occurred from 1990 to 2020 within the analysis polygon. The latter is a rectangle with sides of 576 km eastward and of 745 km northward (Fig. A1). The vertices of the polygon for EEPSOF were converted from kilometric coordinate in the RDN2008 Italy Zone (E-N) system to the WGS84 coordinate reference system. For fitting the EEPAS model, we used the earthquakes from the HORUS catalogue with $M \geq 2.5$ and $Z \leq 50\ km$ occurred inside the polygon from 1960 to 2020. To avoid edge effects in the fitting of model parameters, the contribution of earthquakes in the neighborhood up to 200 km from the polygon were also considered (Fig. A1). The used dataset contains 38086 events, of 24816 which are within the analysis polygon. For both software codes, the log-likelihood optimization is carried out using the downhill simplex method (Nelder and Mead, 1965) as described in Rhoades and Evison (2004).

For the comparison, the fit of the EEPAS parameters is made in five iterations, one for the parameters a_T, a_M and ϑ_A and the others adding one at a time the parameters $(\vartheta_T, \vartheta_M, b_A, b_T)$ to notice the onset of possible deviations. In the first iteration the parameters $\vartheta_T, \vartheta_M, b_A, b_T$ were set to 0.23, 0.32, 0.35 and 0.40 respectively, based on analyses conducted on scaling relationships obtained from the analysis of individual earthquakes (Rhoades and Evison, 2004). With these parameters set, a_T, a_M and σ_A were fit by the maximum likelihood estimation using as starting values 1.5, 1.4 and 3.3, respectively. The fit procedure continued by adding one parameter at a time and considering the previously obtained values as initial values. The parameters, log-likelihoods values and the expected numbers of earthquakes are reported in Table A1.

The optimized parameter values for the first iteration are unequivocally similar for the two codes as the differences are less than 0.90%. In the second and third step the estimates of parameters a_M and ϑ_T begins to slightly deviate with maximum percentual differences up to about 3.0%. With the introduction in the fit of the spatial parameters, the differences of the other parameters also increase. In the fifth and final iteration, the differences are more pronounced, particularly for the parameters a_T, σ_A and b_T, where their values become 12.5%, 10.2% and 24.0%, respectively. However, the differences in log-likelihoods and expected numbers of earthquakes remain small. Such differences mainly regard spatial parameters and are probably related to the different way in which distances are handled by the two different software codes as well as to the different methodology with which integration over space is made.

Appendix C - Implementation of the ETAS, SUP and SVP models

In the literature we can find several implementations of the Epidemic Type Aftershock Sequence (ETAS) model to earthquake forecasting in Italy (e.g. Console et al., 2006, Lombardi and Marzocchi, 2010). In all of them the time dependence is modeled as a sum of Omori decays starting at the times of each occurred shock

$$f(t) = \sum_{i=1}^{n} \frac{H(t - t_i)K}{(t - t_i + c)^p} \tag{C1}$$

where *K*, p and *c* are free parameters and $H(t - t_i)$ is the Heaviside step function which is 1 if $t - t_i > 0$ and is 0 otherwise.

The productivity of each occurred shock of magnitude M_i is described by

$$r = e^{\alpha(M_i - M_c)} \tag{C1}$$

where α is a free parameter and M_c is the minimum magnitude of completeness.

The decay of the productivity with the distance from the epicenter (x_i, y_i) of each occurred shock is described by (e.g. Console et al., 2006, Lombardi and Marzocchi, 2010)

$$g(x,y) = \frac{(q-1)\left[d^{2(q-1)}\right]}{\pi[(x - x_i)^2 + (y - y_i)^2 - d^2]^q} \tag{C3}$$

where *q* and *d* are free parameters and (x_i, y_i) are the epicenters of the occurred shock.

The space decay can also be given in exponential form (Zhuang et al., 2002)

$$g(x,y) = \frac{e^{-\frac{1}{2}\frac{(x-x_i)^2+(y-y_i)^2}{de^{\alpha(M_i-M_c)}}}}{2\pi d e^{\alpha(M_i-M_c)}} \tag{C4}$$

Finally, the frequency magnitude distribution of shocks is given by the Gutenberg and Richter (1944) law

$$h(m) = \beta e^{-\beta(m - M_c)} \tag{C5}$$

Where $\beta = b\ ln10$ is a free parameter.

Combining all the previous terms together, and adding a time invariant background seismicity term $\lambda_0(x, y, m)$, the rate density of ETAS models using the two space decay formulations (C3 and C4 respectively) is given by

$$\lambda(t, x, y, m) = \nu\lambda_0(x, y, m) + [f(t)rg(x, y)h(m)] \tag{C6}$$

$$\nu\lambda_0(x, y, m) + \left\{\sum_{i=1}^{n} \frac{H(t - t_i)K}{(t - t_i + c)^p} e^{\alpha(M_i - M_c)} \frac{(q-1)\left[d^{2(q-1)}\right]}{\pi[(x - x_i)^2 + (y - y_i)^2 - d^2]^q} \beta e^{-\beta(m - M_c)}\right\} \tag{C7a}$$

or

$$\nu\lambda_0(x, y, m) + \left\{\sum_{i=1}^{n} \frac{H(t - t_i)K}{(t - t_i + c)^p} e^{\alpha(M_i - M_c)} \frac{e^{-\frac{1}{2}\frac{(x - x_i)^2 + (y - y_i)^2}{de^{\alpha(M_i - M_c)}}}}{2\pi de^{\alpha(M_i - M_c)}} \beta e^{-\beta(m - M_c)}\right\} \tag{C7b}$$

In both formulations ν represents the ratio between the expected number of independent events and the total number of events. For the comparisons in this work, we only used the formulation (C7a) according to Lombardi and Marzocchi (2010).

The time invariant models of seismicity consist of stationary Poisson processes, which average shock rate may be spatially uniform (Spatially Uniform Poisson, SUP) or variable (Spatially Variable Poisson, SVP). SUP and SVP can also be seen as independent models of seismicity occurrence to compare with other forecasting models (Console et al., 2006).

Their rate density is given by:

$$\lambda_0(x, y, m) = \mu_0(x, y)\beta \exp[-\beta(m - M_c)] \tag{C8}$$

where $\mu_0(x, y)$ is the space rate density of earthquakes with magnitudes equal or larger than M_c. In the SUP model the space-density is assumed to be uniform and independent of the location (x, y). μ_0 is obtained by dividing the number of earthquakes with magnitude above M_c over the whole analysis region R by the total surface area considered.

In the SVP model, the space density $\mu_0(x, y)$ is considered as a continuous smooth function of the geographical location (x, y). To estimate it as a space varying function it is necessary to divide the polygon in squared cells of suitable size. The number of earthquakes N_k with magnitude equal or larger than M_c in each cell is estimated. Each N_k value, representative of a single cell is then smoothed by a Gaussian filter with correlation distance d_c and normalized so that to preserve the total number of events as described in Frankel (1995). For each cell, the smoothed N_k is given by

$$\widetilde{N}_k = \frac{\Sigma_l N_k \exp\left(-\Delta_{kl}^2/{d_c}^2\right)}{\Sigma_l \exp\left(-\Delta_{kl}^2/{d_c}^2\right)} \tag{C9}$$

where Δ_{kl} is the distance between the center of the $k_{\mathrm{th}}$and the l_{th} cells. To obtain $\widetilde{N}_k$ in terms of number of events per unit of time and area, it must be divided by the total duration of the seismic catalogue and by the area of the cell. The value of $\mu_0(x,y)$ in each point of the space is computed by the weighted mean of the four nearest cells that surround the point. To determine d_c we follow the procedure suggested by Console and Murru, (2001): the catalogue is divided in two sub-catalogues of about same temporal length and d_c is chosen as the value that maximize the log-likelihood of a sub-catalog using the smoothed seismicity obtained from the other sub-catalog (Fig A2). The analysis for the optimal d_c is conducted for both sub-catalogs and the obtained value for d_c are respectively $d_{c1} = 16.0$ and $d_{c2} = 13.0$. The optimal correlation distance $d_c = 14.5$ is obtained by the mean of such the two estimations. Once optimized the value d_c the space density of earthquakes $\mu_0(x,y)$ of the SVP background model can be assessed for each cell and for each point in the space (Figure 2). The *b*-value of the Gutenberg and Richter (1944) distribution is the same computed for the EEPAS model as described in the main text.

The parameter q is set to 1.5 according to physical investigation showing that the static stress changes decrease with epicentral distance as r^{-3} (Lombardi and Marzocchi 2010, . The other parameters $(k, p, c, \alpha, d, \nu)$ are fitted by the maximization of the likelihood function (eq. 20 of main text) using the interior point method .

Figures

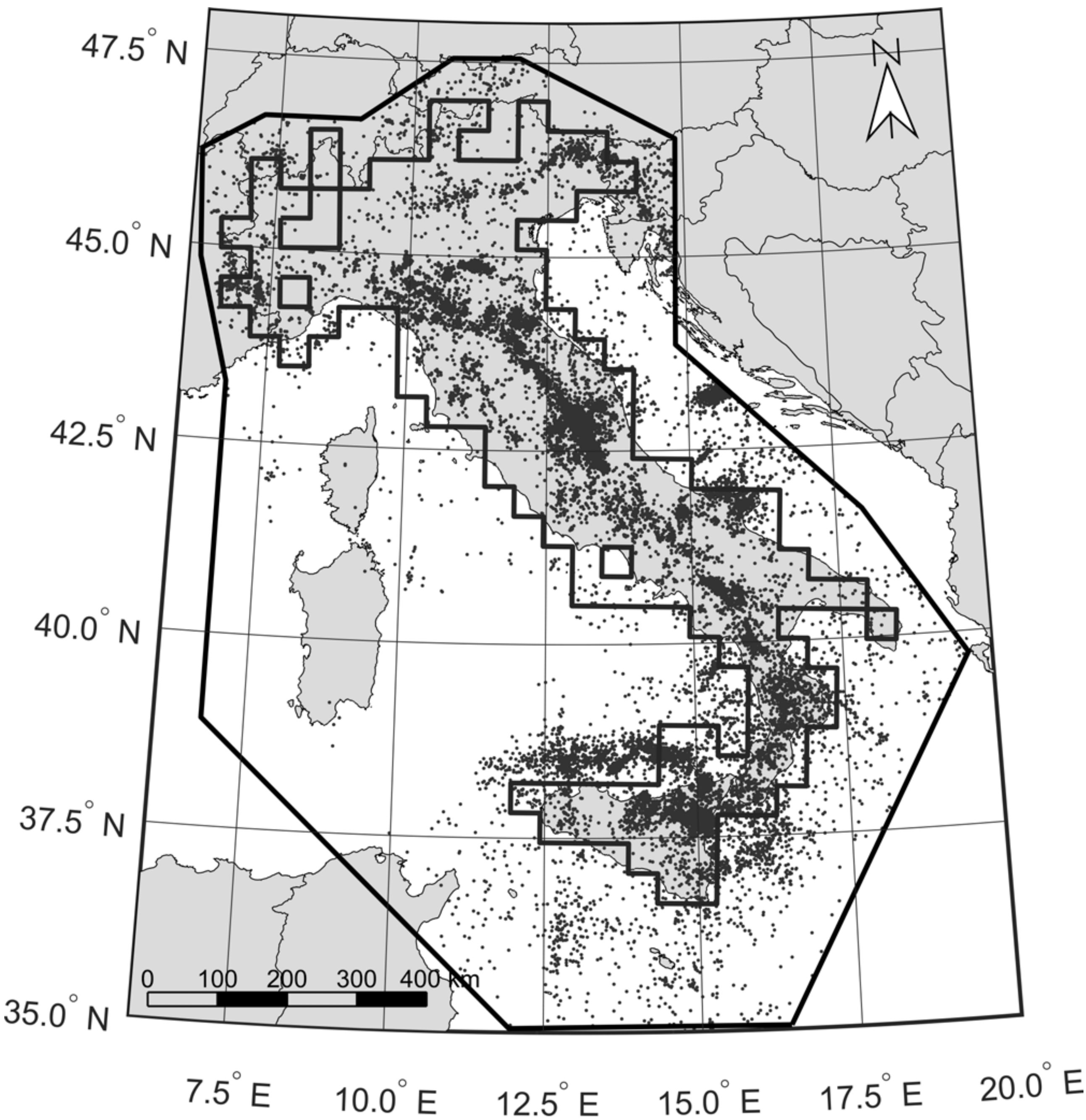


Figure 1: Epicenters of earthquakes with magnitude ≥ 2.5 occurred within the CPTI15 polygon (outer thick polygonal) between 1990 and 2021. The inner thick polygonal represents the forecasting area *R*.

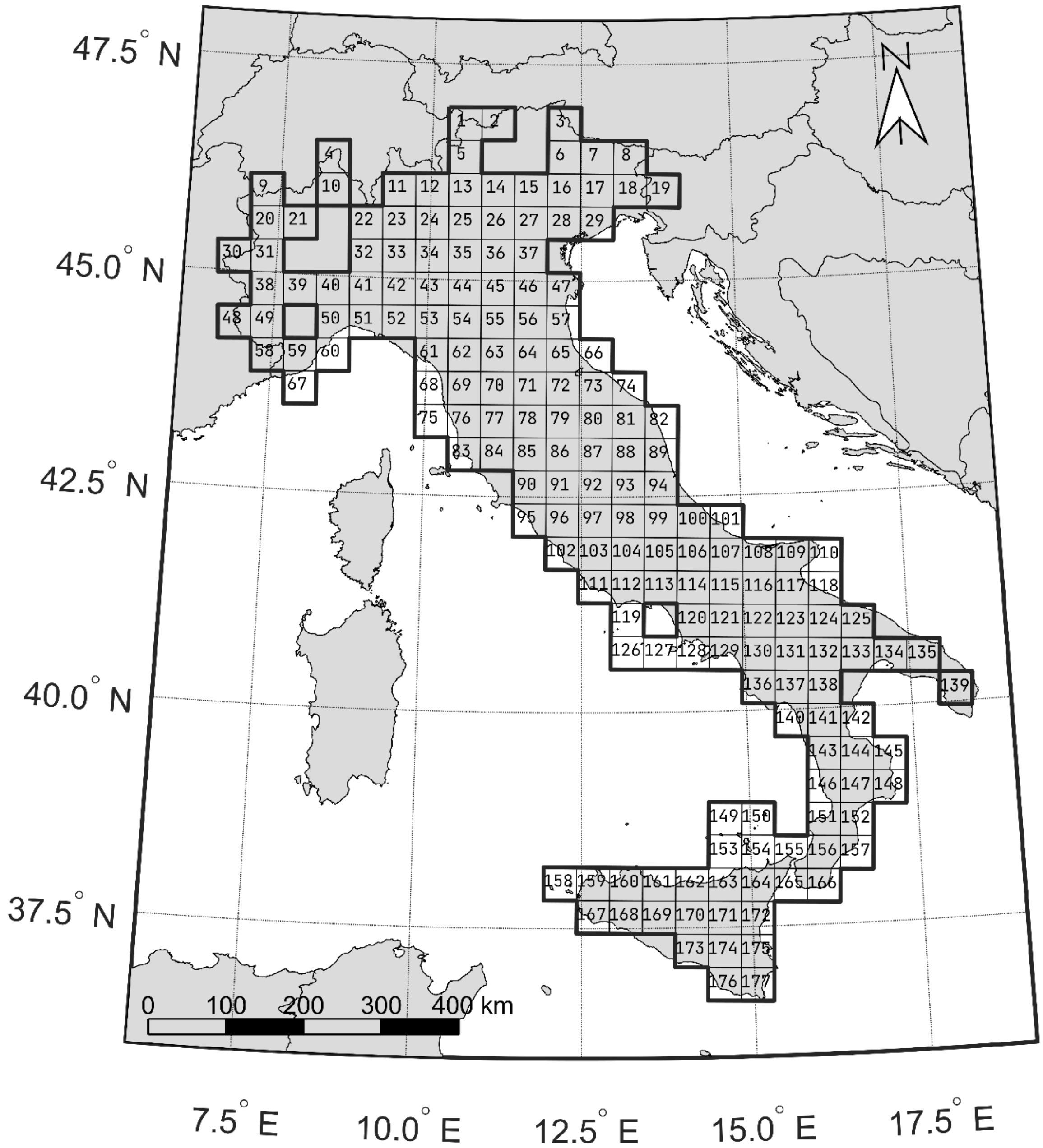


Figure 2. Tessellation of the Italian territory region used for the fitting of parameters and for the retrospective experiment. The thick black line delimits the analysis region R. The cells that compose are only those within which at least one earthquake with $M \geq 4.0$ from 1600 to 2021 have occured according to CPTI15 catalogue (Rovida et al. 2020) and have $30\sqrt{2}$ km of side.

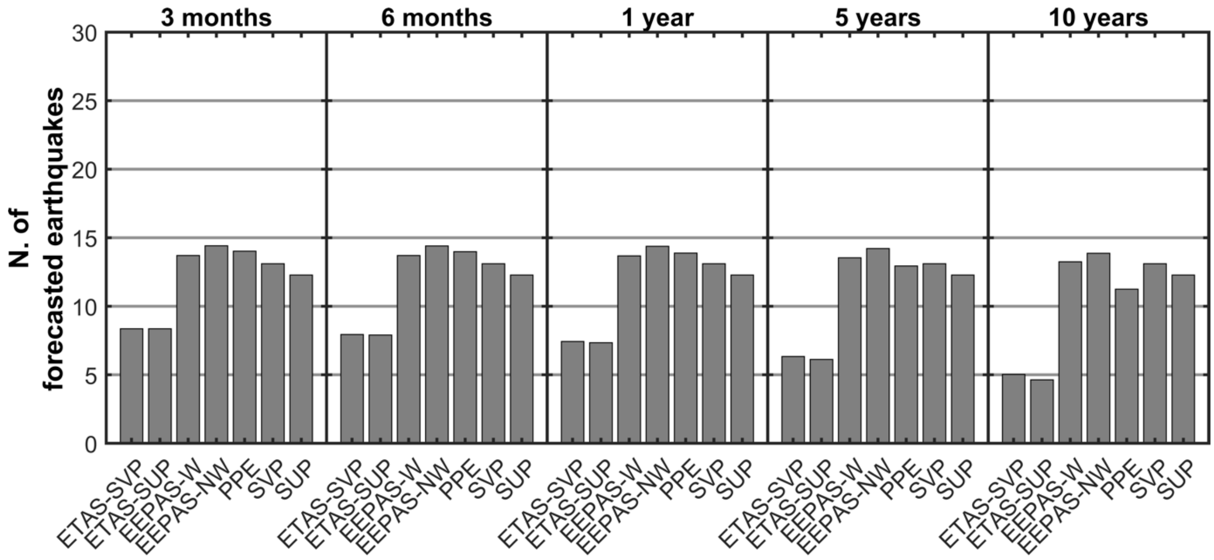


Figure 3 - Numbers of targets (mainshocks+aftershocks) in the testing set (2012-2021) predicted by various models using different time intervals. The effective total number of targets is 27.

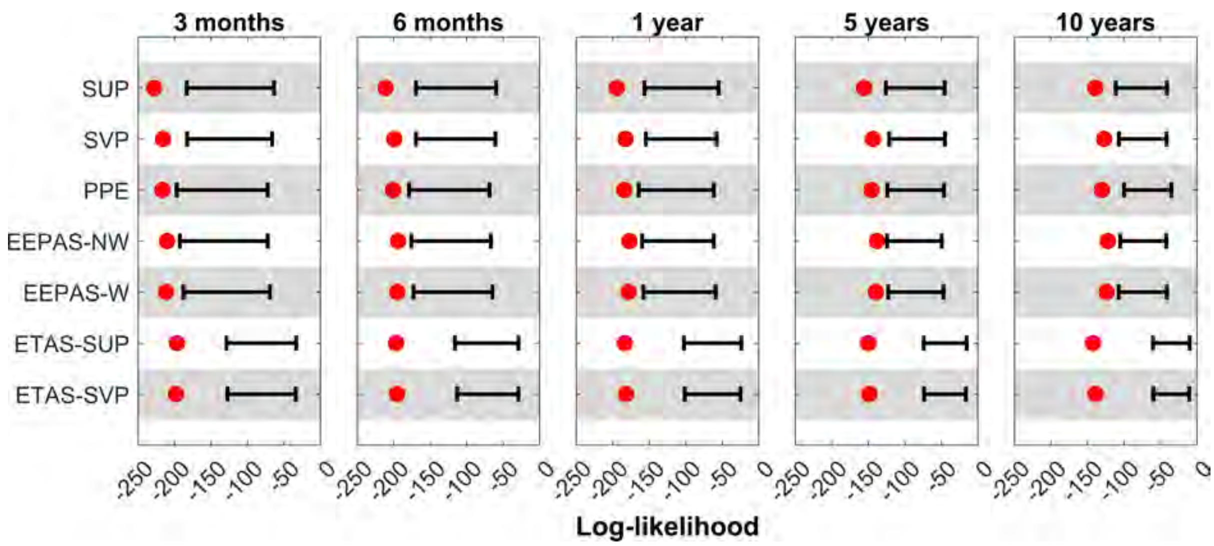


Figure 4 – Results of conditional likelihood consistency test (cL-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks+aftershocks). Black bars indicate 95% confidence limits.

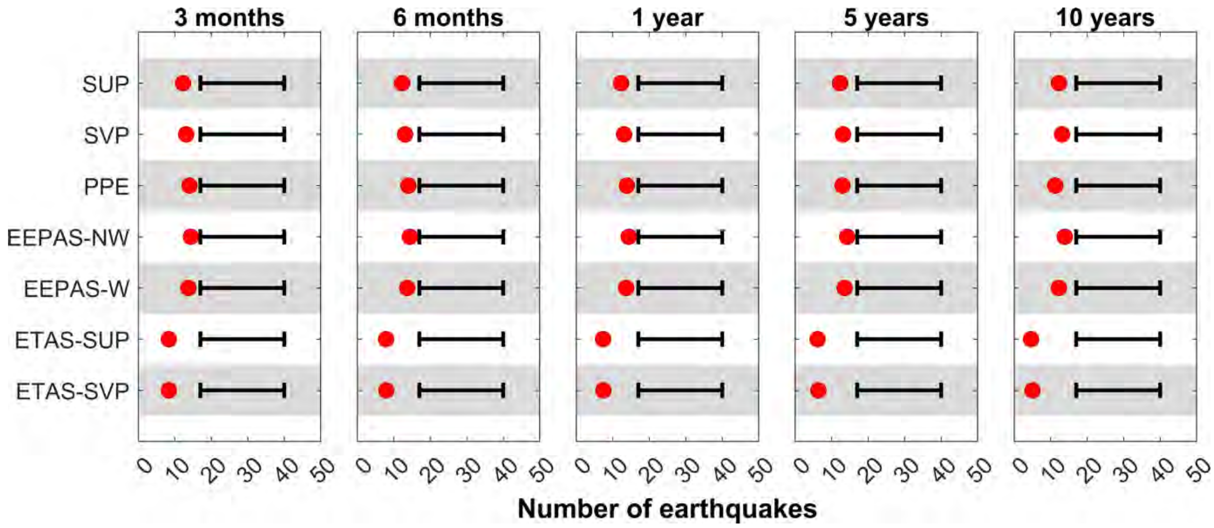


Figure 5 – Results of number consistency test (N-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks+aftershocks). Black bars indicate 95% confidence limits.

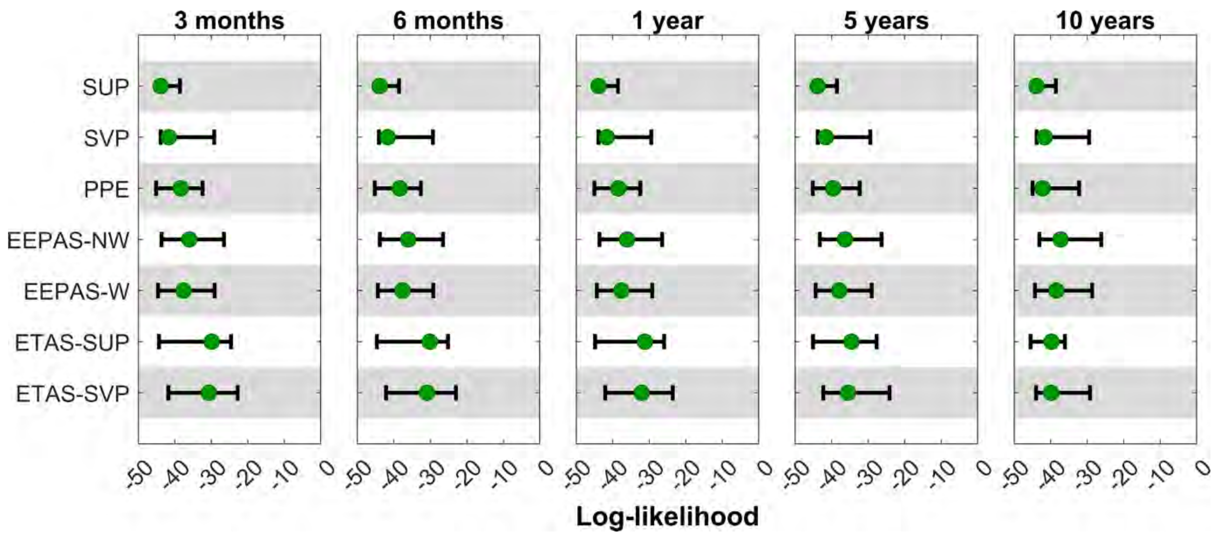


Figure 6 – Results of spatial consistency test (S-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks+aftershocks). Black bars indicate 95% confidence limits.

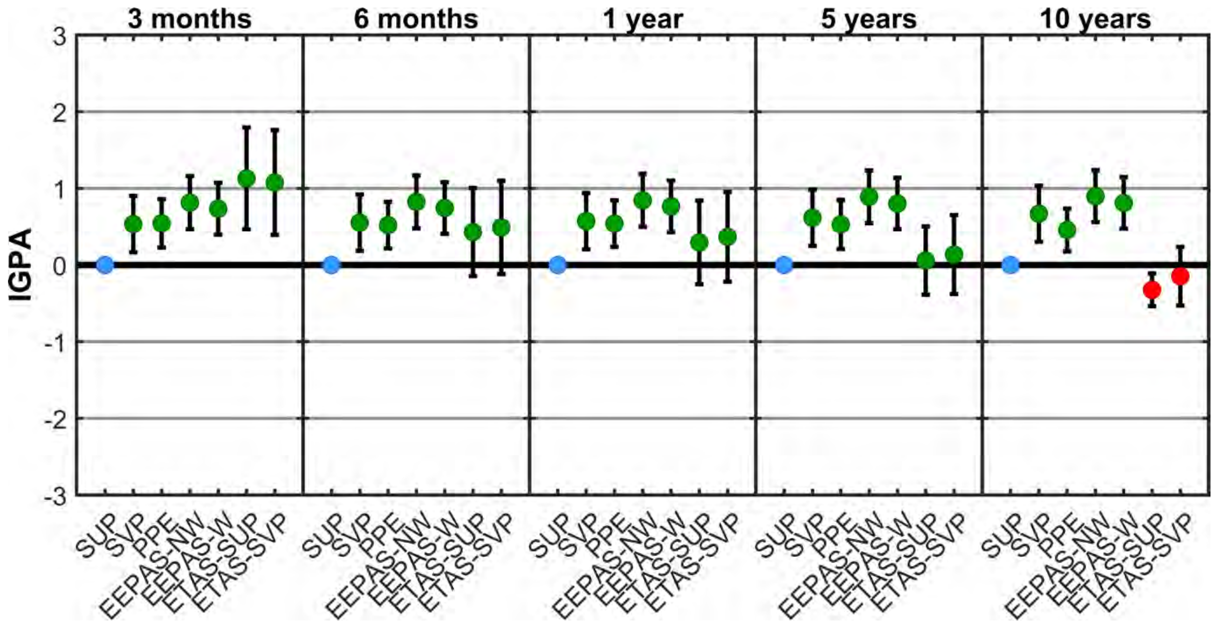


Figure 7 – Comparison) between various models in different time intervals (mainshocks+aftershocks) in the testing set (2012-2021) by the IGPA (T-test). Black bars indicate 95% confidence limits.

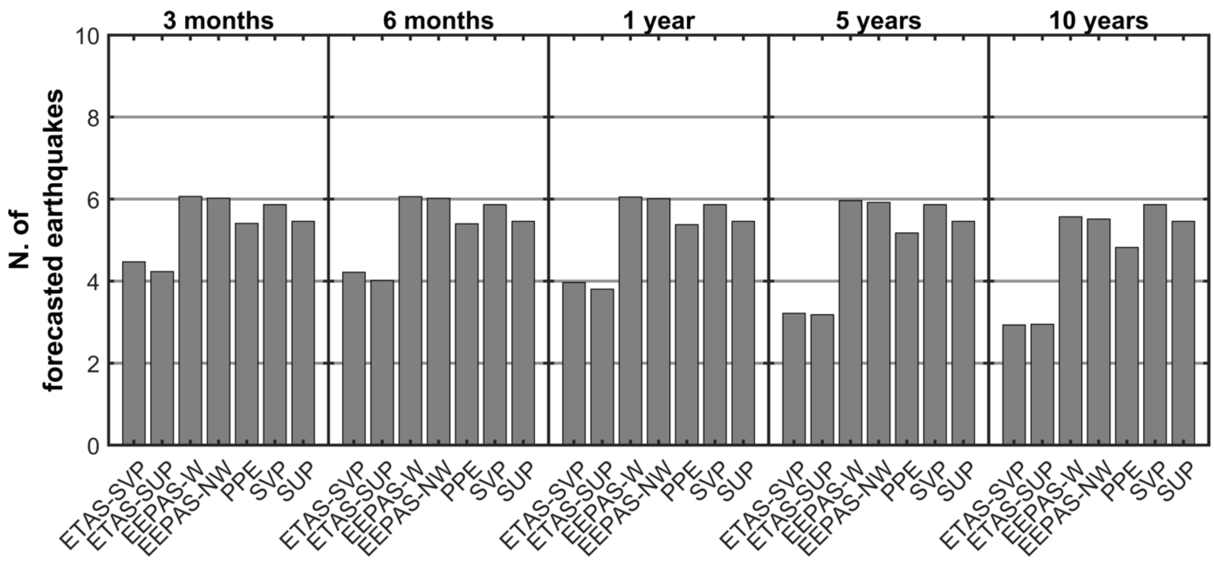


Figure 8 – Numbers of targets (mainshocks only) in the testing set (2012-2021) predicted by various models and time intervals. The effective total number of targets is 9.

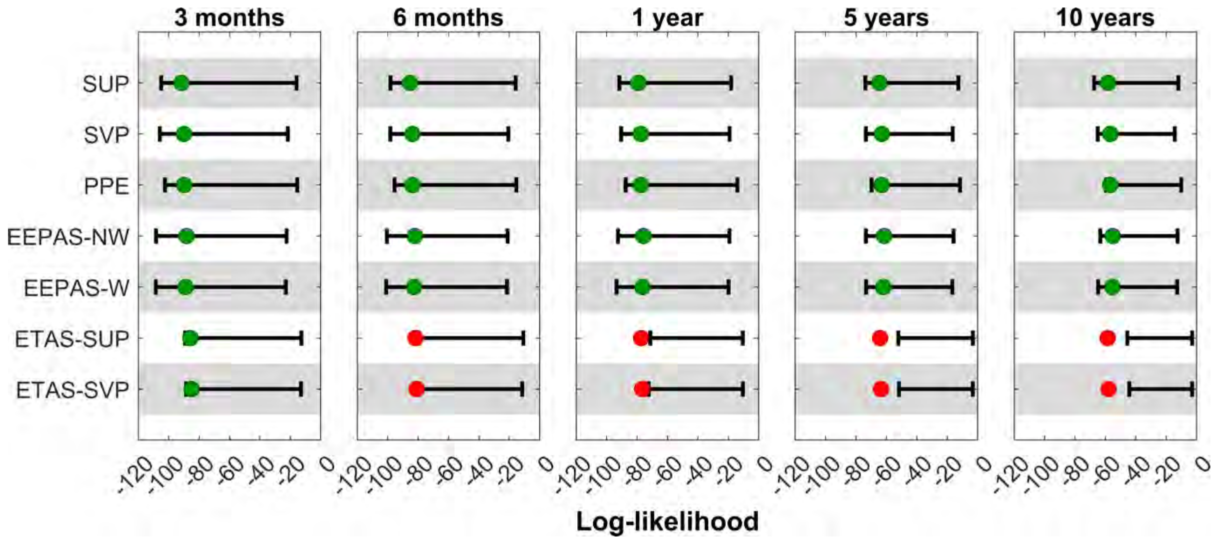


Figure 9 – Results of conditional likelihood consistency test (cL-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks only). Black bars indicate 95% confidence limits.

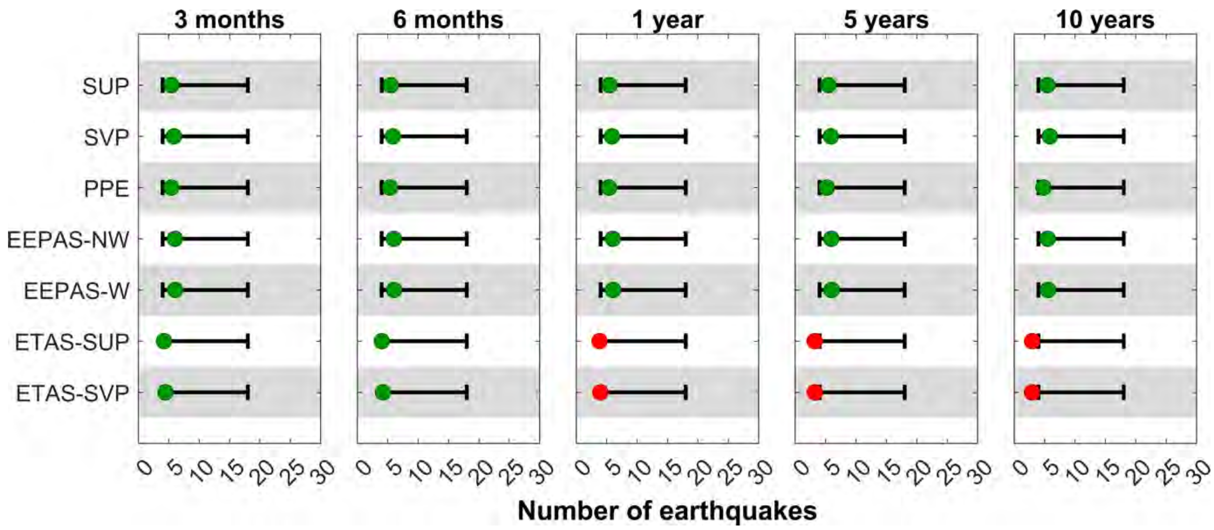


Figure 10 – Results of number consistency test (N-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks only). Black bars indicate 95% confidence limits.

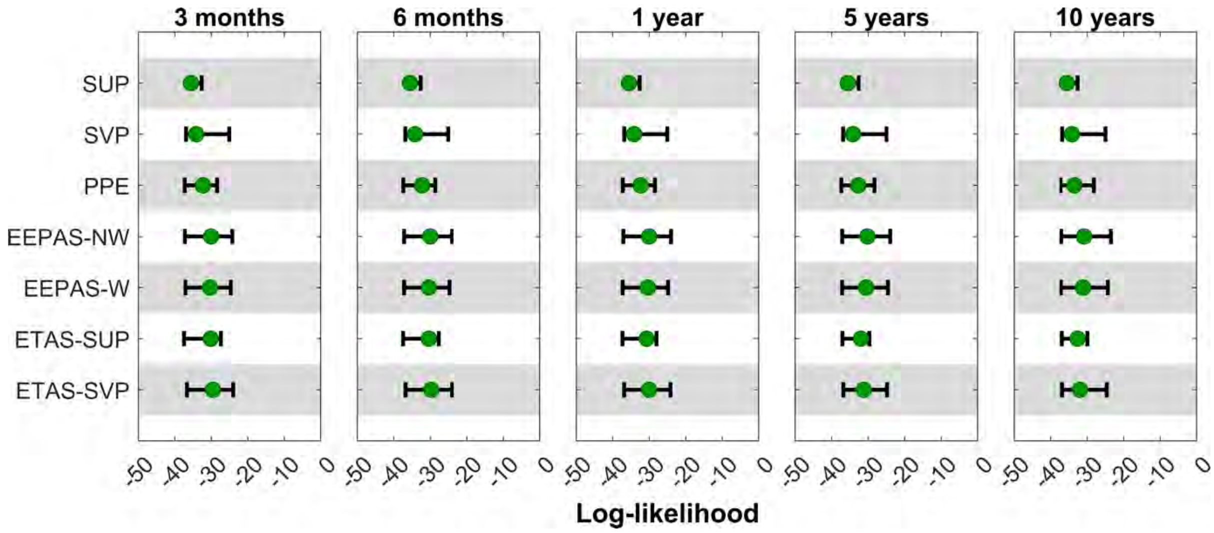


Figure 11 – Results of spatial consistency test (S-test) in the testing set (2012-2021) for various models using different time intervals (mainshocks only). Black bars indicate 95% confidence limits.

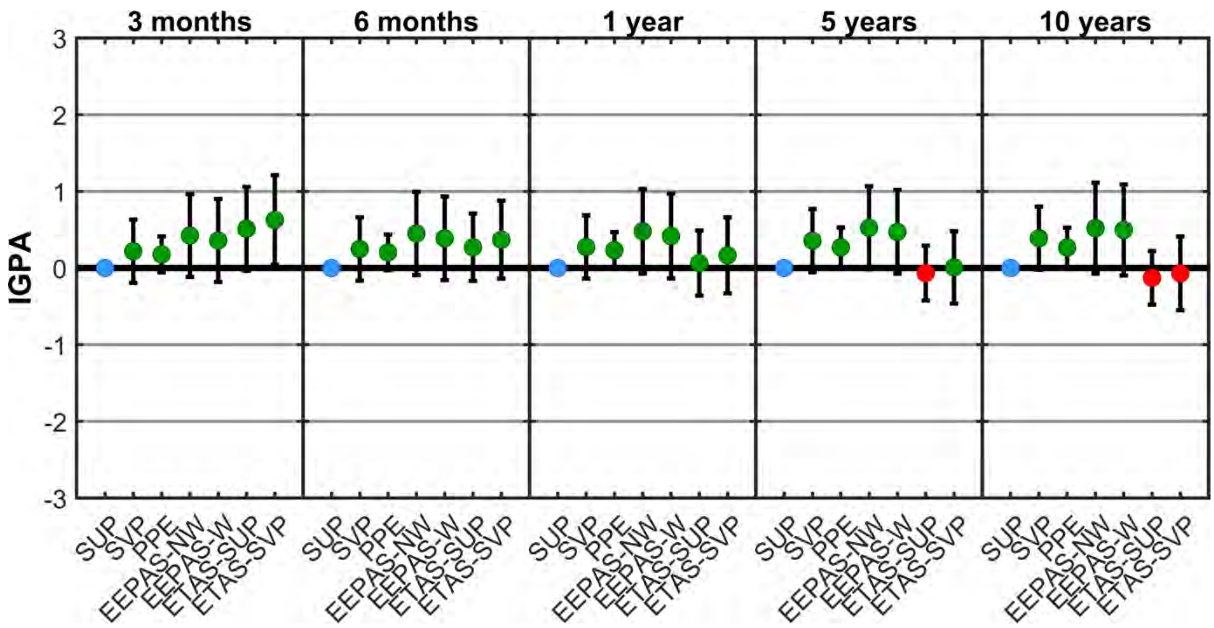


Figure 12 – Comparison between various models in different time intervals (mainshocks only) in the testing set (2012-2021) by the IGPA (T-test). Black bars indicate 95% confidence limits. Black bars indicate 95% confidence limits.

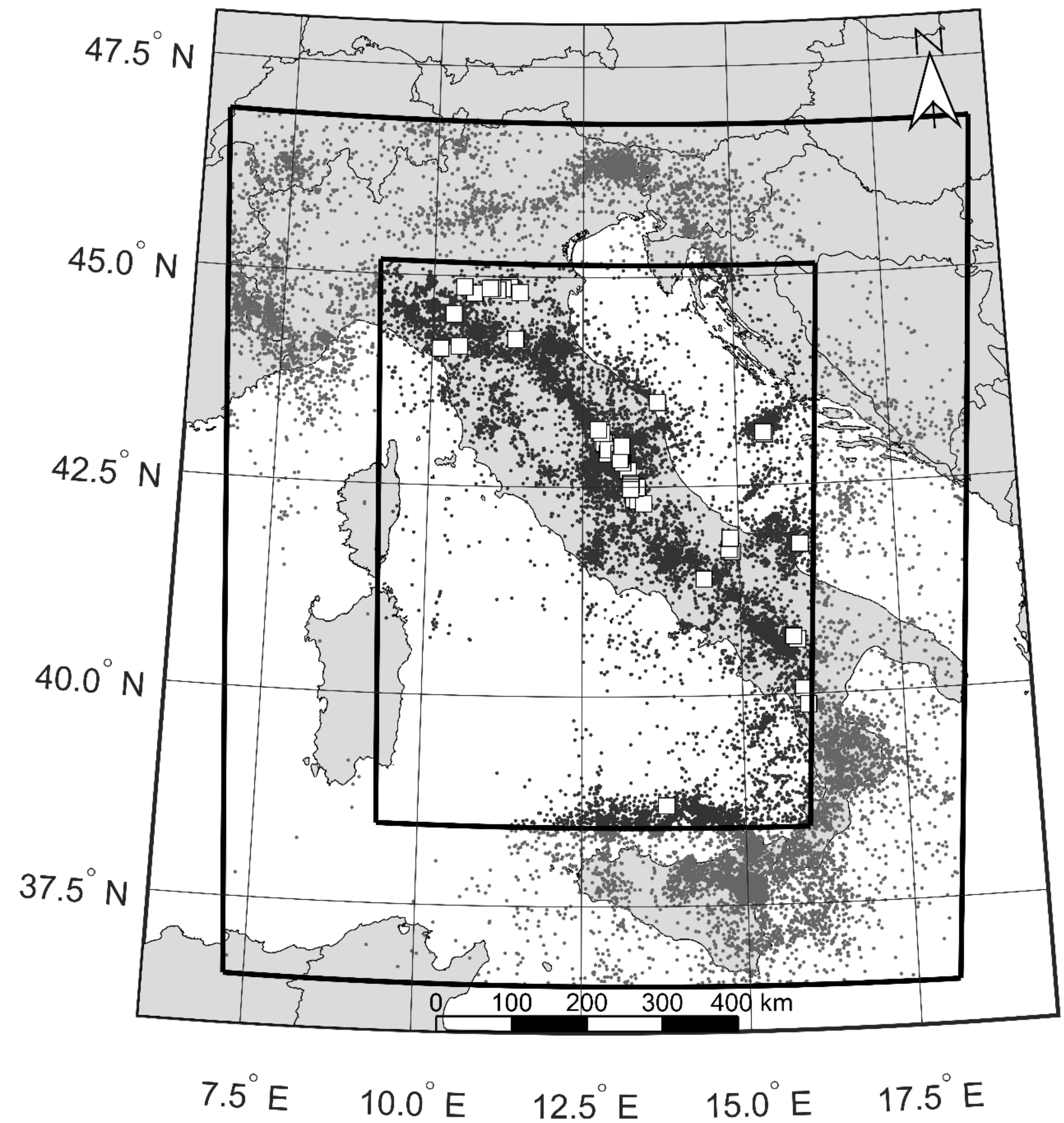


Figure A1: Map of epicenters of earthquakes with $M \geq 2.5$ and $Z \leq 50km$ occurred from 1990 to 2020 within the region involved for the Software codes comparison. The interior rectangular area represents the Analysis polygon for which the EEPAS model is applied. The black point represents the epicenters of earthquake occurred within the Analysis polygon. The external Rectangular represents the influence area for which earthquake indicated by the gray points are also considered for the parameters estimation to avoid edge effects. The white squares represents target earthquake with $M \geq 5.0$ occurred within the analysis polygon in the period 1990-2020.

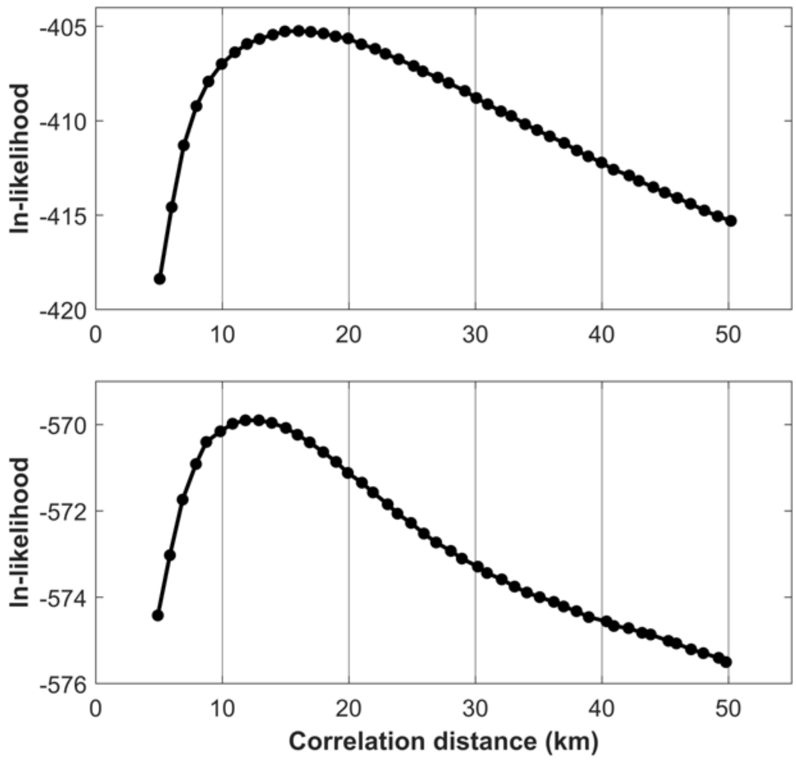


Figure A2: On the top diagram the log-likelihood of the sub-catalog of earthquakes occurred in the period 1990- April 2009 under the time-independent SVP model obtained by the seismicity from April 2009 to 2021. On the bottom as the top figure but with the two time period for the sub-catalogues inverted.

Tables

Table 1- Estimated parameters for various models (mainshock + aftershocks)

SUP		SVP		PPE		EEPAS-NW		EEPAS-W		ETAS-SUP		ETAS-SVP	
b	1.084	b	1.084	b	1.084	b	1.084	b	1.084	b	1.084	b	1.084
		d_c	14.50	a	0.615	a	0.615	a	0.615	K	0.028	K	0.031
				d	29.60	d	29.60	d	29.60	c	0.002	c	0.002
				s	9.0×10^{-13}	s	9.0×10^{-13}	s	9.0×10^{-13}	p	1.005	p	1.022
						a_M	1.222	a_M	1.222	d	1.261	d	1.204
						b_M	1*	b_M	1*	α	1.106	α	0.928
						σ_M	0.246	a_M	0.243	q	1.5*	q	1.5*
						a_T	2.553	a_T	2.720	f_r	0.264	f_r	0.294
						b_T	0.352	b_T	0.315			d_c	14.50
						σ_T	0.150	σ_T	0.150				
						b_A	0.523	b_A	0.508				
						σ_A	1*	σ_A	1*				
						μ	0.177	μ	0.159				

*Fixed or estimated independently.

Table 2 – Performance estimators of various models in the learning time interval (1990-2011) (mainshock + aftershocks)

	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SVP	ETAS-SUP
E	27	27.22	27	27.67	27.73	27.49	27.52
lnL	-524.63	-465.47	-514.11	-500.39	-496.06	-361.60	-363.60
IGPE	0.00	2.19	0.39	0.90	1.06	6.04	5.96
AIC	1051.3	934.9	1036.2	1026.8	1018.1	739.2	743.2
ΔAIC	0.00	2.15	0.28	0.45	0.61	5.78	5.70
G	1.00	8.95	1.48	2.45	2.88	419.1	389.13

Table 3- Estimated parameters for various models (mainshocks)

SUP		SVP		PPE		EEPAS-NW		EEPAS-W		ETAS-SUP		ETAS-SVP	
b	1.176	b	1.176	b	1.176	b	1.176	b	1.176	b	1.176	b	1.176
		d_c	14.50	a	0.390	a	0.390	a	0.390	K	0.029	K	0.033
				d	32.73	d	32.73	d	32.73	c	0.002	c	0.003
				s	1.6×10^{-7}	s	1.6×10^{-7}	s	1.6×10^{-7}	p	1.000	p	0.988

a_M	1.336	a_M	1.336	d	1.227	d	1.749
b_M	1*	b_M	1*	α	0.783	α	0.766
σ_M	0.200	a_M	0.201	q	1.5*	q	1.5*
a_T	1.350	a_T	1.357	f_r	0.380	f_r	0.316
b_T	0.600	b_T	0.599			d_c	14.50
σ_T	0.153	σ_T	0.150				
b_A	0.452	b_A	0.483				
σ_A	1.636	σ_A	1.453				
μ	0.343	μ	0.357				

*Fixed or estimated independently.

Table 4 – Performance estimators of various models in the learning time interval (1990-2011) (mainshocks)

	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SVP	ETAS-SUP
E	12.00	12.19	11.99	14.26	14.75	12.01	11.97
lnL	-246.15	-237.68	-243.52	-239.92	-239.79	-209.75	-208.28
IGPE	0.00	0.75	0.22	0.52	0.54	3.03	3.16
AIC	494.30	479.37	495.04	505.84	505.40	435.49	432.57
ΔAIC	0.00	0.62	-0.03	-0.48	-0.46	2.45	2.57
G	1.00	2.02	1.24	1.68	1.71	20.77	23.46

Table 5 - Numbers of earthquakes predicted by various models in the testing time interval (2012-2021) (mainshocks + aftershocks)

Time interval	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	12.3	13.1	14.0	14.4	13.7	8.4	8.4
6 Months	12.3	13.1	14.0	14.4	13.7	7.9	7.9
1 Year	12.3	13.1	13.9	14.4	13.7	7.3	7.4
5 Years	12.3	13.1	12.9	14.2	13.5	6.1	6.3
10 Years	12.3	13.1	11.3	13.9	13.2	4.6	5.0

Table 6 - Numbers of earthquakes predicted by various models in the testing time interval (2012-2021) (mainshocks)

Time interval	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	5.46	5.86	5.41	6.02	6.06	4.23	4.47
6 Months	5.46	5.86	5.40	6.02	6.06	4.02	4.22
1 Year	5.46	5.86	5.38	6.01	6.05	3.81	3.97
5 Years	5.46	5.86	5.18	5.92	5.96	3.18	3.22
10 Years	5.46	5.86	4.82	5.51	5.57	2.95	2.93

Table A1: Estimated parameters, expected number of target earthquake and log-likelihood values for each iteration step.

EEPSOF code									
1st step		**2nd step (σ_T)**		**3rd step (σ_M)**		**4th step (b_A)**		**5th step (b_T)**	
a_T	2.2264	a_T	2.2109	a_T	2.2204	a_T	2.2187	a_T	2.2173
a_M	1.2295	a_M	1.2289	a_M	1.0165	a_M	1.0339	a_M	1.0141
σ_A	2.3893	σ_A	2.3739	σ_A	2.3360	σ_A	1.2387	σ_A	1.1125
		σ_T	0.2757	σ_T	0.2696	σ_T	0.2686	σ_T	0.2683
				σ_M	0.4845	σ_M	0.4820	σ_M	0.4918
						b_a	0.4899	b_a	0.5181
								b_T	0.4021
L	-961.64	L	-961.491	L	-960.876	L	-960.632	L	-960.594
$\bar{E}$	40.8569	$\bar{E}$	40.0899	$\bar{E}$	39.7087	$\bar{E}$	39.8445	$\bar{E}$	39.5826
MATLAB code									
1st step		**2nd step (σ_T)**		**3rd step (σ_M)**		**4th step (b_A)**		**5th step (b_T)**	
a_T	2.2190	a_T	2.2074	a_T	2.2118	a_T	2.2075	a_T	2.4950
a_M	1.2188	a_M	1.2208	a_M	1.0477	a_M	1.0049	a_M	1.0025
σ_A	2.3929	σ_A	2.3769	σ_A	2.3456	σ_A	1.0015	σ_A	1.0097
		σ_T	0.2670	σ_T	0.2603	σ_T	0.2607	σ_T	0.2652
				σ_M	0.4493	σ_M	0.4771	σ_M	0.4642
						b_a	0.5413	b_a	0.5399
								b_T	0.3236
L	-961.791	L	-961.719	L	-961.264	L	-960.945	L	-960.755
$\bar{E}$	41.222	$\bar{E}$	40.6459	$\bar{E}$	40.5468	$\bar{E}$	40.6171	$\bar{E}$	40.9646

Supporting information

Table S1 – Binary cL test in the testing time interval (2012-2021) (mainshocks + aftershocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.001	0.002	0.007	0.008	0.005	0.000	0.000
6 Months	0.001	0.002	0.005	0.007	0.004	0.000	0.000
1 Year	0.001	0.002	0.004	0.005	0.004	0.000	0.000
5 Years	0.001	0.002	0.002	0.006	0.004	0.000	0.000
10 Years	0.000	0.002	0.000	0.003	0.003	0.000	0.000

Table S2 – Binary N test in the testing time interval (2012-2021) (mainshocks + aftershocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.0002	0.0005	0.0013	0.0019	0.0010	2.4×10^{-7}	2.5×10^{-7}
6 Months	0.0002	0.0005	0.0013	0.0019	0.0010	8.0×10^{-8}	8.9×10^{-8}
1 Year	0.0002	0.0005	0.0012	0.0019	0.0009	1.9×10^{-8}	2.4×10^{-8}
5 Years	0.0002	0.0005	0.0004	0.0016	0.0008	4.4×10^{-10}	9.5×10^{-10}
10 Years	0.0002	0.0005	0.0000	0.0011	0.0006	1.0×10^{-12}	6.3×10^{-12}

Table S3 – Binary S test in the testing time interval (2012-2021) (mainshocks + aftershocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.095	0.084	0.598	0.415	0.443	0.821	0.593
6 Months	0.613	0.090	0.589	0.416	0.452	0.837	0.602
1 Year	0.277	0.091	0.581	0.415	0.441	0.818	0.549
5 Years	0.083	0.085	0.434	0.358	0.393	0.680	0.297
10 Years	0.092	0.009	0.148	0.283	0.344	0.785	0.190

Table S4 – Information Gain per active bin in the testing time interval (2012-2021) (mainshocks + aftershocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.00(7)	0.53(6)	0.54(5)	0.81(3)	0.73(4)	1.13(1)	1.08(2)
6 Months	0.00(7)	0.55(3)	0.52(4)	0.82(1)	0.74(2)	0.43(6)	0.49(5)
1 Year	0.00(7)	0.57(3)	0.54(4)	0.84(1)	0.76(2)	0.23(6)	0.37(5)
5 Years	0.00(7)	0.62(3)	0.53(4)	0.89(1)	0.80(2)	0.06(6)	0.14(5)
10 Years	0.00(7)	0.67(3)	0.46(4)	0.90(1)	0.81(2)	-0.32(6)	-0.15(5)

Table S5 – Binary cL test in the testing time interval (2012-2021) (mainshocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.071	0.087	0.070	0.114	0.112	0.033	0.033
6 Months	0.070	0.085	0.068	0.105	0.110	0.022	0.021
1 Year	0.072	0.077	0.070	0.112	0.108	0.013	0.016
5 Years	0.072	0.080	0.053	0.921	0.099	0.004	0.004
10 Years	0.075	0.077	0.037	0.073	0.079	0.002	0.002

Table S6 – Binary N test in the testing time interval (2012-2021) (mainshocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.102	0.139	0.098	0.155	0.159	0.029	0.039
6 Months	0.102	0.139	0.097	0.154	0.159	0.022	0.029
1 Year	0.102	0.139	0.096	0.154	0.158	0.016	0.020
5 Years	0.102	0.139	0.080	0.145	0.149	0.005	0.006
10 Years	0.102	0.139	0.057	0.107	0.111	0.003	0.003

Table S7 – Binary S test in the testing time interval (2012-2021) (mainshocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.568	0.159	0.710	0.631	0.641	0.877	0.637
6 Months	0.765	0.160	0.711	0.632	0.641	0.881	0.627
1 Year	0.709	0.158	0.695	0.628	0.640	0.869	0.587
5 Years	0.662	0.161	0.623	0.605	0.611	0.860	0.497
10 Years	0.567	0.162	0.484	0.509	0.536	0.824	0.389

Table S8 – Information Gain per active bin in the testing time interval (2012-2021) (mainshocks)

Time intvl	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 Months	0.00(7)	0.22(5)	0.18(6)	0.42(3)	0.36(4)	0.51(2)	0.63(1)
6 Months	0.00(7)	0.25(5)	0.20(6)	0.45(1)	0.39(2)	0.27(4)	0.37(3)
1 Year	0.00(7)	0.28(3)	0.23(4)	0.75(1)	0.42(2)	0.06(6)	0.17(5)
5 Years	0.00(7)	0.36(3)	0.27(4)	0.52(1)	0.47(2)	-0.07(6)	0.01(5)
10 Years	0.00(7)	0.39(3)	0.27(4)	0.52(1)	0.50(2)	-0.13(6)	-0.07(5)

Inconsistencies and Lurking Pitfalls in the Magnitude–Frequency Distribution of High-Resolution Earthquake Catalogs

Marcus Herrmann*[1] and Warner Marzocchi[1]

Abstract

Earthquake catalogs describe the distribution of earthquakes in space, time, and magnitude, which is essential information for earthquake forecasting and the assessment of seismic hazard and risk. Available high-resolution (HR) catalogs raise the expectation that their abundance of small earthquakes will help better characterize the fundamental scaling laws of statistical seismology. Here, we investigate whether the ubiquitous exponential-like scaling relation for magnitudes (Gutenberg–Richter [GR], or its tapered version) can be straightforwardly extrapolated to the magnitude–frequency distribution (MFD) of HR catalogs. For several HR catalogs such as of the 2019 Ridgecrest sequence, the 2009 L'Aquila sequence, the 1992 Landers sequence, and entire southern California, we determine if the MFD agrees with an exponential-like distribution using a statistical goodness-of-fit test. We find that HR catalogs usually do not preserve the exponential-like MFD toward low magnitudes and depart from it. Surprisingly, HR catalogs that are based on advanced detection methods depart from an exponential-like MFD at a similar magnitude level as network-based HR catalogs. These departures are mostly due to an improper mixing of different magnitude types, spatiotemporal inhomogeneous completeness, or biased data recording or processing. Remarkably, common-practice methods to find the completeness magnitude do not recognize these departures and lead to severe bias in the *b*-value estimation. We conclude that extrapolating the exponential-like GR relation to lower magnitudes cannot be taken for granted, and that HR catalogs pose subtle new challenges and lurking pitfalls that may hamper their proper use. The simplest solution to preserve the exponential-like distribution toward low magnitudes may be to estimate a moment magnitude for each earthquake.

Cite this article as Herrmann, M., and W. Marzocchi (2020). Inconsistencies and Lurking Pitfalls in the Magnitude–Frequency Distribution of High-Resolution Earthquake Catalogs, *Seismol. Res. Lett.* **92**, 909–922, doi: 10.1785/0220200337.

Supplemental Material

Introduction

Enriching seismic catalogs with smaller earthquakes (i.e., below M 2.0) offers many benefits due to the higher spatiotemporal resolution of seismicity (Ebel, 2008; Brodsky, 2019). For example, such microearthquakes help identify the location and extent of active faults (e.g., Fischer and Horálek, 2003; Waldhauser *et al.*, 2004; Chiaraluce *et al.*, 2007; Piccinini *et al.*, 2009; Improta *et al.*, 2019), allow faster and/or new inferences about seismotectonic processes (e.g., Hatzfeld *et al.*, 2000; Bohnhoff *et al.*, 2006; Bulut *et al.*, 2009; Valoroso *et al.*, 2013; Marzorati *et al.*, 2014; Hainzl *et al.*, 2016; Meng and Peng, 2016; Shelly *et al.*, 2016; Ross *et al.*, 2020), and provide potentially better conditions for seismicity and hazard analyses with their application to forecasting models (e.g., Wiemer and Schorlemmer, 2007; Werner *et al.*, 2011; Mignan, 2014; Tormann *et al.*, 2014; Gulia and Wiemer, 2019).

With the increasing availability of high-resolution (HR) earthquake catalogs, these expectations might easily be taken for granted. Their refined locations suggest an improved resolution, whereas their abundance of smaller events suggests an improved completeness. Scientists may be tempted to blindly assume that the popular Gutenberg–Richter (GR) scaling relation, or its tapered version (TGR), observed in ordinary earthquake catalogs (i.e., older network-based catalogs that span a limited range of magnitudes), holds also in the low-magnitude range of HR catalogs.

Many seismicity studies use the magnitude–frequency distribution (MFD) to estimate seismicity rates and the *b*-value

1. Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli 'Federico II', Naples, Italy

*Corresponding author: marcus.herrmann@unina.it

(the slope of the GR relation), from which the occurrence probability of larger events and eventually the seismic hazard and risk can be inferred. The earthquake magnitude is usually expected to follow an exponential-like distribution according to the unbounded, TGR, and truncated GR relation when the maximum (corner) magnitude is about $\geq M_c + 3$ (Marzocchi *et al.*, 2020), in which M_c is the lower magnitude cutoff, or magnitude of completeness. (In the following, we refer to "exponential-like" simply as "exponential.") An exponential distribution above M_c is a necessary and sufficient condition to calculate the b-value (Marzocchi *et al.*, 2020)—otherwise, the physical meaning of the b-value becomes questionable. Various methods exist to determine M_c (e.g., Wiemer and Wyss, 2000; Wössner and Wiemer, 2005; Amorèse, 2007; Schorlemmer and Woessner, 2008; Mignan and Wössner, 2012) or to model the full MFD including the incomplete part (Kijko and Smit, 2017; Martinsson and Jonsson, 2018; Mignan, 2019), but the exponential property of the MFD is rarely verified through canonical statistical tests.

Several factors can significantly alter MFDs and produce artifacts that bias any inferred estimate. It is generally well known that artificial changes in the reporting of magnitudes (e.g., due to recalibrations or changing seismic networks) may affect the homogeneity of catalogs (Habermann, 1987; Zúñiga and Wyss, 1995; Tormann *et al.*, 2010; Kamer and Hiemer, 2015) and that major earthquakes lead to marked under-reporting of small events shortly after (Kagan, 2004; Helmstetter *et al.*, 2006; Hainzl, 2016; de Arcangelis *et al.*, 2018). Moreover, a single uniform magnitude scale is not always possible for several practical reasons (Kanamori, 1983), so that HR catalogs often mix different kinds of magnitudes, which may have different nonexponential MFDs. For instance, the local magnitude, M_L, has an exponential scaling only in a limited magnitude range: it begins to saturate for large magnitudes (above M 6) (Kanamori, 1983) due to the Wood–Anderson instrument response acting as a high-pass filter (Bormann and Saul, 2009), and it breaks in scale around M 2–4 due to the anelastic attenuation in the medium acting as a low-pass filter (Bethmann *et al.*, 2011; Munafò *et al.*, 2016; Deichmann, 2017).

Here, we take a closer look at MFDs of currently available HR catalogs that span more than six orders of magnitude, examining different (1) spatiotemporal scales (individual sequences vs. entire southern California), (2) temporal states (2019 Ridgecrest vs. 1992 Landers sequence), and (3) tectonic environments (southern California vs. Italy). We analyze whether their MFDs agree with an exponential TGR distribution (referring to this agreement as "consistency" hereinafter) and explore whether inconsistencies can be detected through common-practice methods to estimate the parameters of the TGR distribution (M_c and b-value). Understanding in more detail whether, how, and why MFDs of available HR catalogs are inconsistent is fundamental to use them correctly in statistical seismology.

Catalogs and Statistical Methods

For southern California, we consider the following earthquake catalogs (see Data and Resources for accessed repositories):

- Southern California Seismic Network (SCSN) catalog (Hutton *et al.*, 2010; Southern California Earthquake Data Center [SCEDC], 2013);
- U.S. Geological Survey's Advanced National Seismic System (USGS-ANSS) ComCat;
- Hauksson *et al.* (2012), containing relatively relocated hypocenters of SCSN events;
- quake template matching (QTM; Ross, Trugman, *et al.*, 2019), based on template matching (TM) using SCSN events as template set; and
- three dedicated catalogs for the Ridgecrest sequence (Ross, Idini, *et al.*, 2019; Lee *et al.*, 2020; Shelly, 2020b) based on TM using SCSN events as templates.

For the 2009 L'Aquila, Italy, sequence, we use the HR catalog of Valoroso *et al.* (2013) (see Data and Resources). We only focus on the magnitude information contained in these catalogs. All catalogs have a magnitude discretization, or binning, of $\Delta M = 0.01$.

To analyze their MFDs, we calculate the most relevant parameters for an exponential distribution, that is, M_c and the b-value. At first, we apply two common M_c estimation methods (Mignan and Wössner, 2012): (1) the maximum curvature method (Wiemer and Wyss, 2000) that uses the mode of the MFD; we include a correction of $+0.2$ magnitude units (Wössner and Wiemer, 2005), hereinafter referred to as $M_c^{\mathrm{MAXC}}(+0.2)$; and (2) median-based analysis of the segment slope method (Amorèse, 2007) that detects a change point; we use the 2σ confidence interval (~95%) of a 1000-sample bootstrap distribution as the final estimate, hereinafter referred to as $M_c^{\mathrm{MBASS}}(+2\sigma)$. To enhance the stability of both methods, we apply them to magnitudes rounded to one decimal place. The b-value is estimated with a bias-free maximum-likelihood method (Tinti and Mulargia, 1987; Marzocchi and Sandri, 2003; Marzocchi *et al.*, 2020), but only for sample sizes of 100 or larger.

At the same time, we assess whether the magnitude is exponentially distributed using the canonical goodness-of-fit test of Lilliefors (1969). Only a goodness-of-fit test can indicate whether data follow a certain distribution (Clauset *et al.*, 2009). The Lilliefors test is a modification of the Kolmogorov–Smirnov (KS) one-sample test to be used when the parameters of the distribution are unknown and need to be estimated from the sample. (This test is the same as the often termed "modified KS test" referring to either Stephens, 1974 or Pearson and Hartley, 1972, which are based on extensive and revised Monte Carlo [MC] simulations compared with Lilliefors, 1969; our test statistic is similarly based on an extensive MC simulation with 10 million replications; see Data and Resources.) Because the exponential distribution is a continuous probability distribution, the 0.01-binned magnitudes are transformed into a continuous random variable by adding uniformly sampled random noise $\mathcal{U}(-\frac{\Delta M}{2}, \frac{\Delta M}{2})$. For 0.01-binned magnitudes, the added uniform noise does not affect significantly the exponentiality of the distribution up to sample sizes of at least one million, as confirmed by Lilliefors test simulations (not shown). The Lilliefors test is performed as a function of M_c for 50 initializations of the random noise, from which we obtain an average p-value, $\overline{p}_M$, at each magnitude bin. The p-value expresses the probability to observe the data sample assuming the null hypothesis is true (here, the exponential distribution). According to Ronald Fisher's original interpretation, it measures the strength of evidence against the null hypothesis. It is worth remarking that our application cannot be seen as a formal statistical test because of this recursive testing but is used to highlight significant departures from the exponential GR relation. We use $\overline{p}_M$ with a significance level of $\alpha = 0.1$ to obtain the lowest magnitude level above which the MFD can be considered exponential, hereinafter referred to as the Lilliefors-based magnitude of completeness, $M_c^{\text{Lilliefors}}$. Choosing $\alpha = 0.1$ is conservative in a statistical sense (Clauset *et al.*, 2009); less conservative choices $\alpha < 0.1$ increase the probability to not reject models that have only a very small chance to follow an exponential distribution. To improve stability, $\overline{p}_M$ must exceed α for at least five consecutive magnitude bins, in which case the first exceedance, that is, the lowest magnitude bin, yields the eventual $M_c^{\text{Lilliefors}}$.

Departures from the exponential distribution can occur either over a magnitude range or intermittently at various magnitude levels. To facilitate identifying and characterizing MFD inconsistencies, we determine the slope of the MFD (i.e., the b-value) as a function of M_c. Although the b-value of nonexponential MFDs does not have a physical meaning, a systematic dependence on M_c provides clues on the kind of MFD inconsistency.

We do not discuss the existence of various nonexponential MFDs for individual fault segments, such as the characteristic earthquake model (Schwartz and Coppersmith, 1984), which should not present significant differences from an exponential distribution if the characteristic magnitude is much larger than M_c.

Results

Ridgecrest sequence

We first compare the magnitude statistics of the relocated Hauksson *et al.* (2012) catalog to the original SCSN catalog for the Ridgecrest sequence (see Fig. 1a–c) within the time range 1 April–31 December 2019 and a distance of 100 km from the mainshock. The MFD of both catalogs is very similar (gray and yellow in Fig. 1a), because the Hauksson *et al.* (2012) catalog takes over the magnitudes of SCSN and is a subset thereof. Both MFDs feature a discontinuity around M 3.5, which has a strong influence on the b-value (see the abrupt change for $M_c \geq 2.8$ in Fig. 1b), which peaks at M 3.44 with a b-value of ~1.2. Below M 3.5, the Lilliefors p-values (Fig. 1c) indicate a rejection of exponentiality for both catalogs. The composition of SCSN magnitude types in terms of their MFD (red, blue, and green in Fig. 1a) reveals that the discontinuity is caused by an improper merging of the local (M_L) and moment magnitude (M_w) scale (see the Discussion section). $M_c^{\text{Lilliefors}} = 3.54$ accounts for this discontinuity, whereas $M_c^{\text{MAXC}}(+0.2) = 1.10$ and $M_c^{\text{MBASS}}(+2\sigma) = 1.51$ do not and are much lower. The latter two do not comply with the assumed exponential distribution of the GR relation and lead to biased b-value estimates (Fig. 1b). The b-value below M 3.1 (smaller than 1) is considerably different from the one between M 3.1 and 3.6 (well above 1), and above M 3.6 (around 1.05).

The dedicated catalogs for the Ridgecrest sequence with even higher resolution (Ross, Idini, *et al.*, 2019; Lee *et al.*, 2020; Shelly, 2020b) are affected by the same inconsistency (Fig. 1d–f). We restricted all three catalogs—and the SCSN catalog for comparison—to their common spatiotemporal window: 4 July 2019 15:35 (2 hr prior to the M 6.4 foreshock) until 17 July 2019 (~11 days after the M 7.1 mainshock), within a radius of 37 km from the coordinate 35.74° N, 117.54° W (approximately in the middle of both hypocenters). For the Ross, Idini, *et al.* (2019) catalog, the Lilliefors p-values (Fig. 1f) reveal that the exponentiality cannot be rejected around $M_c = 1.8$ and above $M_c = 3.5$, but it is rejected in between ($M_c = 2.0$–3.5). This indicates that this catalog has two different exponential distributions with distinct b-values: a first between ~M 1.8 and M 3.5 containing ~95% of the data above ~M 1.8, dominated by the M_L scaling (b-value about 0.8); and a second above M 3.5 when the discontinuity is overcome, dominated by the scaling of M_w and M_{Lr} (b-value about 1.0). Because of the short-lived exponentiality (i.e., no persistent exponentiality with increasing M_c) far below the discontinuity, $M_c^{\text{Lilliefors}}$ could mislead as it does not relate to the exponential distribution of the largest events but to a secondary one of the smaller events. The MFD of the Shelly (2020b) catalog shows a similar behavior with increased p-values around $M_c = 2.0$, albeit not exceeding the significance level; $M_c^{\text{Lilliefors}}$ is above the discontinuity and similar to the SCSN catalog. The Lee *et al.* (2020) catalog shows a short-lived exponentiality just below the discontinuity as indicated by $M_c^{\text{Lilliefors}} = 3.16$, which means that the discontinuity around M 3.5 is reduced compared with the other catalogs—but it is unclear why. The b-value at $M_c^{\text{Lilliefors}} = 3.16$ (about 1.05) is comparable with the one for $M_c > 3.6$ (Fig. 1e)—a coincidence because it remains anomalous in between. For all catalogs, $M_c^{\text{MAXC}}(+0.2)$ and $M_c^{\text{MBASS}}(+2\sigma)$ again yield underestimated (i.e., overconfident) completeness magnitudes, which do not comply with the exponential assumption.

Landers sequence

Inspecting the Landers sequence as an example for an older catalog period shows that it is composed of even more magnitude types, each having different nonexponential MFDs (Fig. 2a). Most notably, the coda amplitude-based magnitude,

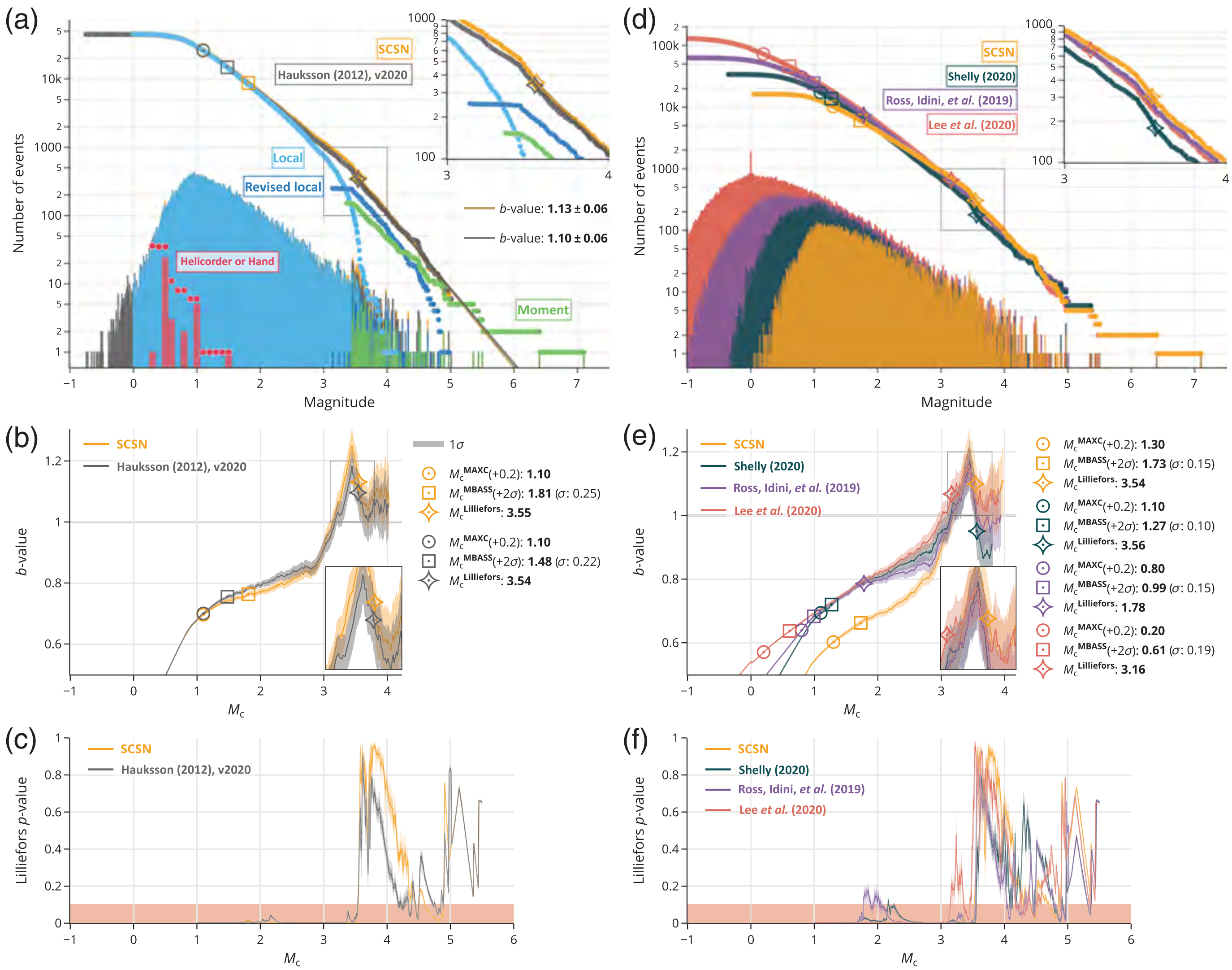


Figure 1. Magnitude statistics of the 2019 Ridgecrest sequence using various catalogs. Panels (a–c) relate to nine-month data extracts (see the Ridgecrest Sequence section) of the Southern California Seismic Network (SCSN) and Hauksson *et al.* (2012) catalog, including the underlying composition of magnitude types; panels (d–f) relate to three template-matching-based (TM-based) catalogs (Ross, Idini, *et al.*, 2019; Lee *et al.*, 2020; Shelly, 2020b) and the SCSN catalog in their common spatiotemporal window (see the Ridgecrest Sequence section). (a and d) The catalogs in terms of their magnitude–frequency distribution (MFD). Panels (b and e) and (c and f) show, as a function of lower magnitude cutoff, or magnitude of completeness, M_c, the *b*-value (the slope of the fitted Gutenberg–Richter relation), and the Lilliefors *p*-value (assuming an exponential distribution as null hypothesis), respectively. Different estimates of M_c are indicated in panels (a,d) and (b,e) (see legend and the Ridgecrest Sequence section). Each inset in those rows magnifies the enframed section of the plot area.

M_{coda}, is overrepresented below M 3.0, whereas the helicorder or hand magnitude, M_h, mixes differently binned magnitudes (0.5, 0.1, and 0.01). These two magnitude types considerably affect the estimated *b*-value of the overall catalog as a function of M_c (Fig. 2b): M_{coda} raises the *b*-value for $M_c = 2.0$–3.0 compared with $M_c > 3.0$, whereas M_h results in distinct *b*-value jumps due to the irregular binning, especially in 0.5 magnitude steps noticeable up to $M_c = 4.0$. The 0.5 binning of M_h causes the overall MFD to be nonexponential for $M_c = 3.2$–3.5 and again briefly for $M_c = 4.0$ (Fig. 2c). Below M 3.0, the MFD is not exponential anymore due to the combined effects of M_h and M_{coda} and the *b*-value is overestimated. The common completeness estimation methods with $M_c^{MAXC}(+0.2) = 1.70$ and $M_c^{MBASS}(+2\sigma) = 2.33$ would result in such a biased *b*-value because they again do not comply with the exponential assumption.

Regional catalog of southern California

We further investigate whether the MFD of a regional seismic catalog is inconsistent (Fig. 3). We compare the 10 yr QTM catalog (Ross, Trugman, *et al.*, 2019) with the SCSN catalog for 2008–2017 to obtain the contributing magnitude types. M_L and M_w merge around M 4.4 (Fig. 3a), but its impact on the *b*-value is too uncertain because of too few data. Yet, the merging might be the reason for the *p*-value decrease around $M_c \approx 4.4$ (Fig. 3c). The M_h magnitude—especially its 0.1 binning—causes some irregularities and fluctuations in the b-value below M 3.5.

The QTM catalog has the same $M_c^{\mathrm{Lilliefors}}$ as SCSN (M 3.24), which implies that QTM does not comply with an exponential distribution to lower magnitudes. This finding is similar to the Ridgecrest sequence (Fig. 1d–f), but without the lower magnitudes having a distinct secondary exponential distribution. Below $M_c^{\mathrm{Lilliefors}}$, the MFD gradually curves toward low magnitudes, accompanied by continuously decreasing b-values. $M_c^{\mathrm{MAXC}}(+0.2)$ and $M_c^{\mathrm{MBASS}}(+2\sigma)$ again do not comply with the exponential assumption and differ much more from $M_c^{\mathrm{Lilliefors}}$ than in the catalogs for the individual sequences.

Particular attention must be paid to the "reloc" subset of the QTM catalog; it excludes many events (such as the 2010 M 7.2 Baja California Sierra El Mayor–Cucapah event). As a consequence, the b-value diverges from the SCSN catalog for $M_c > 3$. This subset should be used with great care for statistical analyses because its MFD is apparently not a good representation of the actual MFD.

We also investigate temporal changes in the proportion of magnitude types for southern California. Figure 4 summarizes these proportions for the periods of the SCSN catalog analyzed previously. The 1999 M 7.1 Hector Mine sequence was added as an intermediate temporal sample. (For the sake of completeness, we have also applied the MFD analysis to the Hector Mine sequence; see

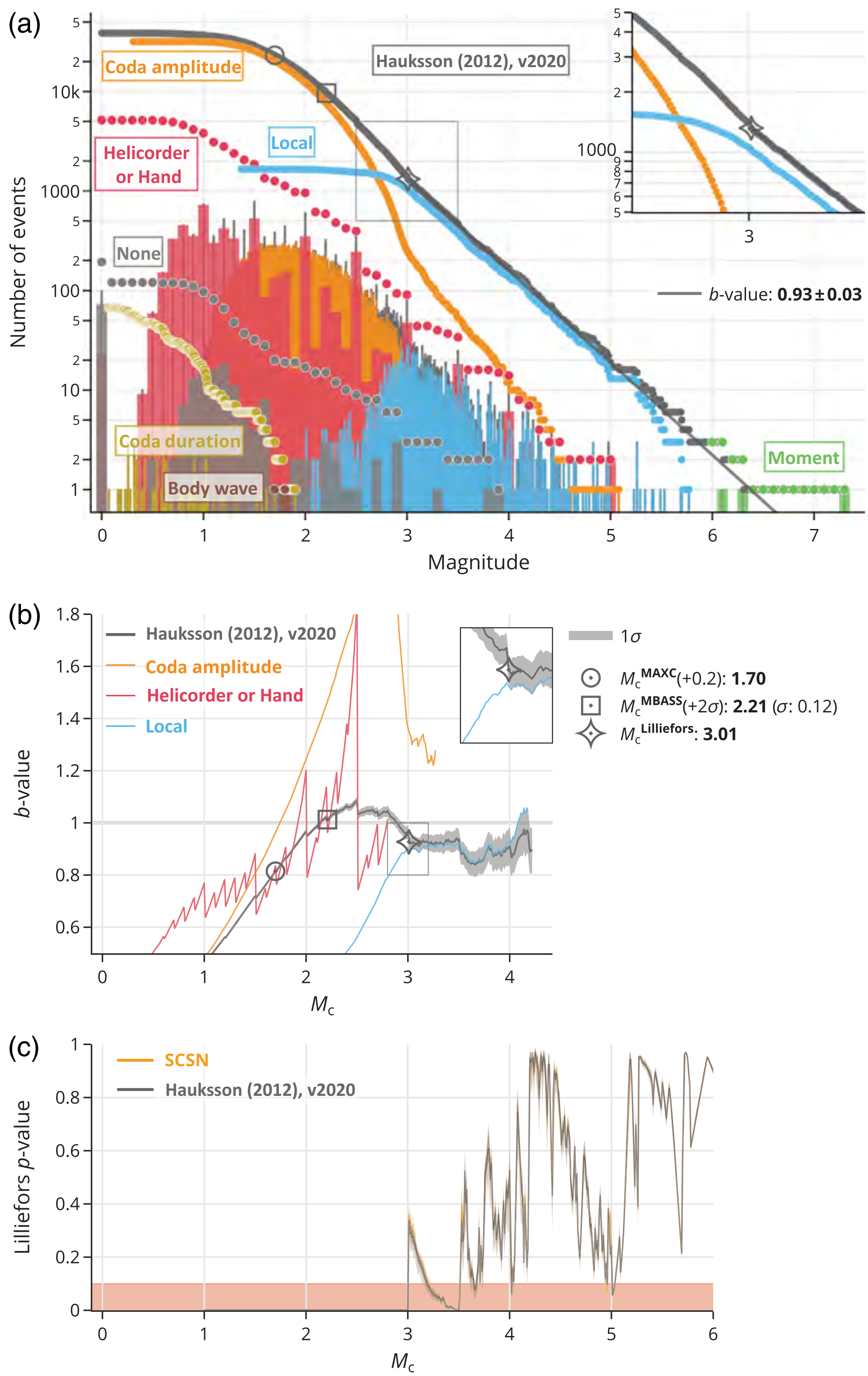


Figure 2. Magnitude statistics of the 1992 Landers sequence using 1 yr of data (1 March 1992 until 28 February 1993, within 100 km from the mainshock). Analogous to Figure 1a–c, panel (a) shows catalog data in terms of their magnitude–frequency distribution; panels (b) and (c) show the b-value and the Lilliefors p-value as function of M_c. In addition, panel (b) shows the b-value as function of M_c for three dominating magnitude types (see legend).

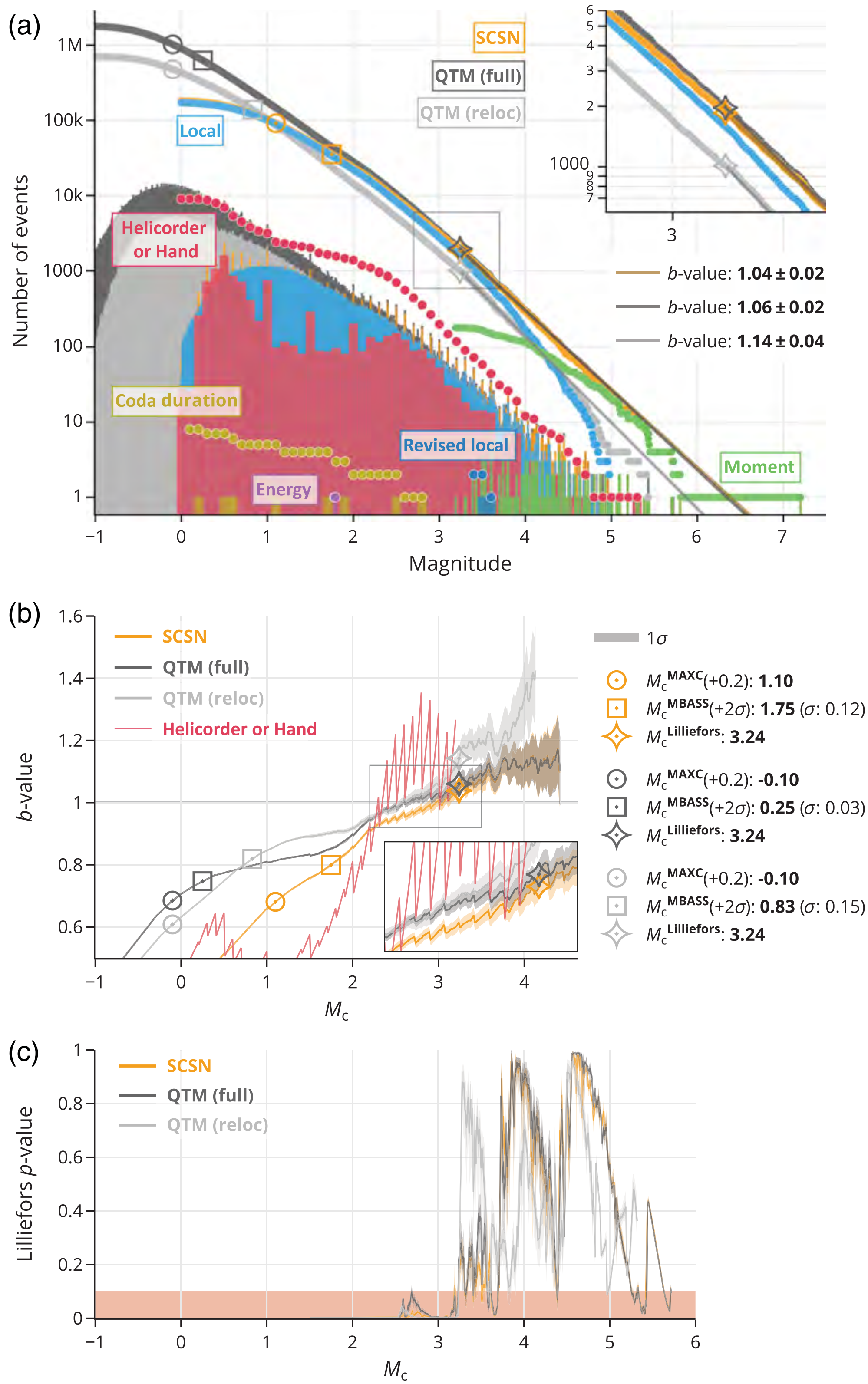


Figure 3. Magnitude statistics for catalog data of entire southern California in 2008–2017, the time period of the quake template matching (QTM) catalog (Ross, Trugman, *et al.*, 2019) ("QTM (full)," dark gray). Its subset of only relocated events ("QTM (reloc)") is shown in light gray. Analogous to Figures 1 and 2, panel (a) shows catalog data in terms of their magnitude–frequency distribution; panels (b) and (c) show the *b*-value and the Lilliefors *p*-value as function of M_c.

Fig. S3 and Text S2 in the supplemental material, available to this article.) The most apparent change over time is the gradual replacement of both M_{coda} and M_h by M_L. Simultaneously, M_L gets replaced by M_w as can be seen from the individual MFDs (Figs. 1d, 2a, and 3a and Fig. S3a). As a consequence of the changing magnitude proportions over time, the magnitude types merge at different magnitude levels, which may cause one or more discontinuities as shown earlier. We summarize those in Text S1 for the four analyzed periods of the SCSN or Hauksson *et al.* (2012) catalog.

Figure 5 shows that the time dependence of the discontinuities propagates to $M_c^{Lilliefors}$, making it time dependent as well. $M_c^{Lilliefors}$ is elevated during 1985–1995 and again from 2010 onward compared to 1980–1985 and 1995–2010. Figure S1 shows the same analysis for 2 yr time intervals in which $M_c^{Lilliefors}$ fluctuates more strongly; its highest estimates typically dominate the respective 5 yr interval in Figure 5. These estimates match $M_c^{Lilliefors}$ found for the individual sequences and entire southern California in 2008–2017 (filled symbols in Fig. 5). For comparison, M_c^{MAXC} and M_c^{MBASS} generally decrease over time, reflecting that smaller events are increasingly being added to the catalog. In the last 5 yr catalog period, $M_c^{Lilliefors}$ differs from both M_c^{MAXC} and M_c^{MBASS} by ~2.5 magnitude units, highlighting again their discrepancy already found for the Ridgecrest sequence.

It should be noted that the USGS-ANSS ComCat, which depends on SCSN as a regional seismic network, inherits the

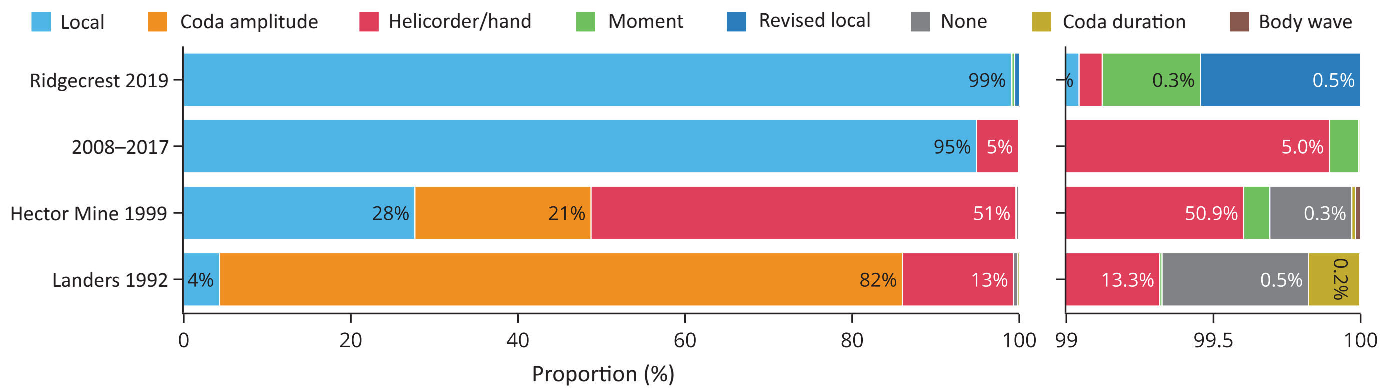


Figure 4. Proportion of magnitude types for different temporal periods of the SCSN catalog. Magnitude types are sorted by the order given in the legend. The subfigure in the right zooms into the 99%–100% range.

same magnitude composition (see Fig. S2) and therefore the same inconsistencies outlined previously.

L'Aquila sequence

The HR catalog of Valoroso *et al.* (2013) for the L'Aquila sequence suffers from a similar apparent overrepresentation of low-magnitude events as observed for the Landers sequence. Here, it is the single cause for the rejection of the exponential distribution below $M_c^{\mathrm{Lilliefors}} = 1.81$ (see Fig. 6, black curve). We investigated the catalog within consecutive nonoverlapping time windows (indicated in Fig. S4) and found a time-dependent degree of overrepresentation (Fig. 6a, colored MFDs): It is not evident in the first four days (red) but starts to appear in the subsequent week (yellow) and is more dominant for the following time periods after about 1.5 weeks after the mainshock (green, blue, and dark blue). Accordingly, the b-value below $M_c^{\mathrm{Lilliefors}} = 1.81$ becomes exceptionally overestimated at these later times (Fig. 6b). The MFD of the whole catalog (black curve) below $M_c^{\mathrm{Lilliefors}}$ mixes the MFD behavior in the individual time windows: the overall overrepresentation is weaker than in the last periods but greater than in the initial periods. This compensation effect also applies to the b-value below $M_c^{\mathrm{Lilliefors}}$ (Fig. 6b).

The overconfident $M_c^{\mathrm{MAXC}}(+0.2)$ and $M_c^{\mathrm{MBASS}}(+2\sigma)$ point to a completeness magnitude in which the b-value is maximally biased, whereas $M_c^{\mathrm{Lilliefors}}$ yields a b-value of about 1.0. Although the b-values differ among the individual time periods at this magnitude level (see inset of Fig. S4b), the uncertainty estimates indicate that they are not significantly different from 1.0. (Only a 1σ, i.e., 68%, confidence interval is shown.)

For the sake of completeness, we performed the time-window analysis for the Ridgecrest sequence using the catalog of Hauksson *et al.* (2012) (Figs. S5 and S6) and Ross, Idini, *et al.* (2019) (Figs. S7 and S8), and for the Landers sequence (Figs. S9 and S10). The analyses of both catalogs for the Ridgecrest sequence show that MFDs of the earliest, mostly incomplete, period are much more curved than MFDs of later periods. The MFDs of later periods appear exponential down to M 1.5 mostly because they are more complete, but partly because the discontinuity around M 3.4 is undersampled and barely noticeable. Yet, the Lilliefors test still detects the discontinuity and results in short-lived rejections below it (Figs. S6c and S8c). The earliest period is the most active, and its MFD shape dominates the overall MFD (e.g., it contains more $M > 2.5$ events than all other periods combined). For the Landers sequence, the individual MFDs behave similar to the L'Aquila sequence: the early period tends to a lower b-value, whereas the later periods increase in b-value with decreasing M_c below $M_c^{\mathrm{Lilliefors}}$.

Discussion

In all inspected HR catalogs, the exponential distribution does not hold toward low-magnitude ranges. This undesired shortcoming applies to both sequence-based and regional catalogs. Noteworthy, common completeness estimation methods (such as MAXC and MBASS) cannot capture these departures,

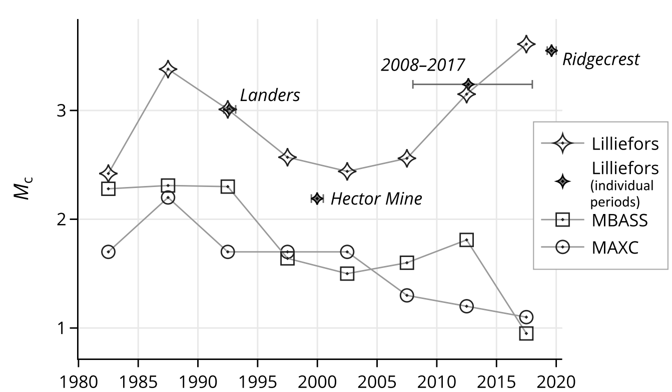


Figure 5. Estimates of M_c using three different methods for 5 yr intervals of the (Hauksson *et al.*, 2012) catalog. The symbols are aligned with the midpoint of each time interval. For comparison, $M_c^{\mathrm{Lilliefors}}$ estimated for four individual periods of this catalog (Figs. 1–3 and Fig. S3) are indicated with filled symbols; the horizontal whiskers relate to their time span.

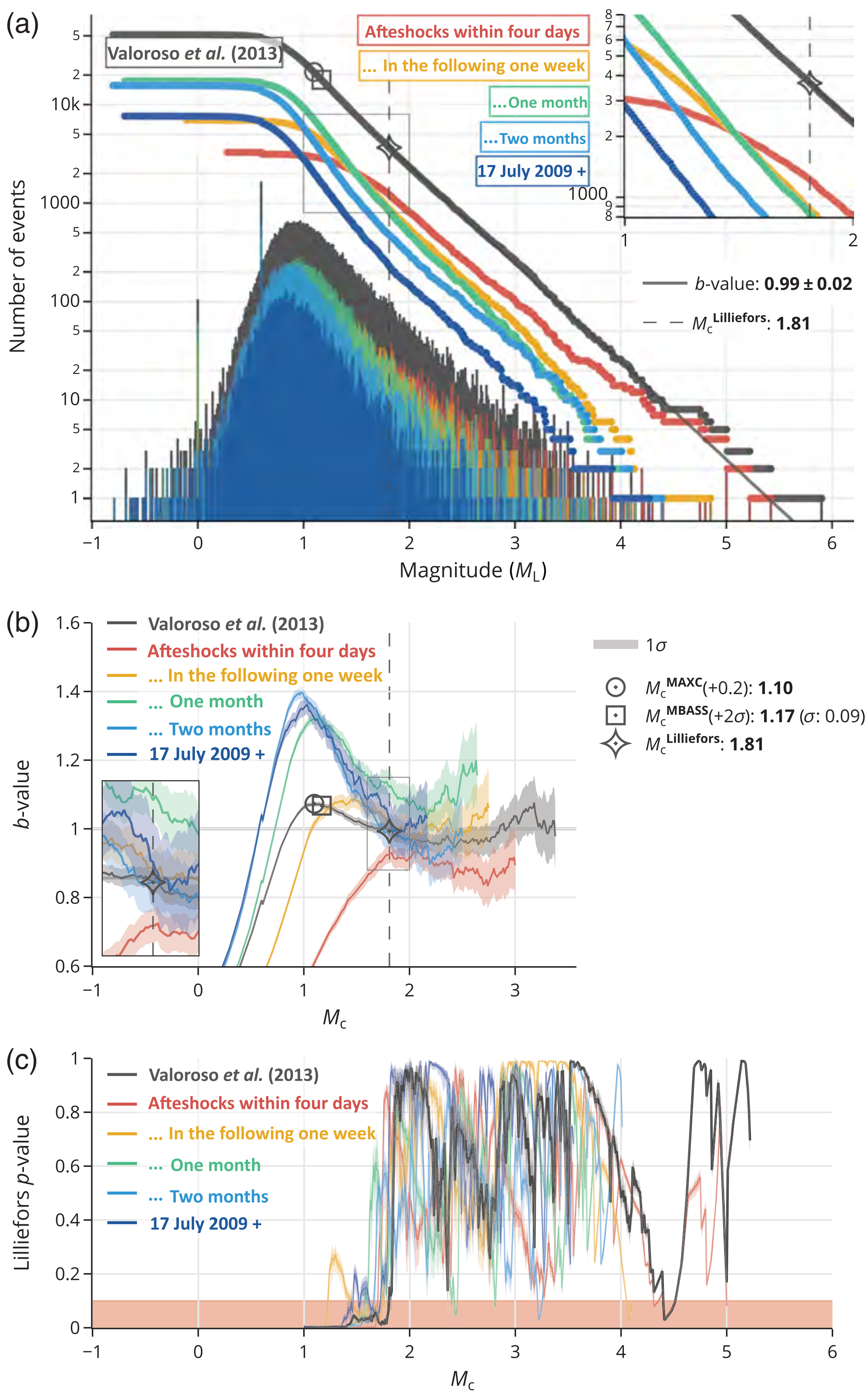


Figure 6. Magnitude statistics of the 2009 L'Aquila, Italy, sequence using the catalog of Valoroso *et al.* (2013). The colored curves relate to periods of the catalog in nonoverlapping time windows after the mainshock (see legend and Fig. S3). Accordingly, the last time window starts about 3.3 months after the mainshock. Analogous to Figures 1–3, panel (a) shows catalog data in terms of their magnitude–frequency distribution; panels (b) and (c) show the *b*-value and the Lilliefors *p*-value as function of M_c.

leading to (1) M_c that do not comply with the exponential distribution and (2) biased *b*-values that do not describe the magnitude distribution of the largest events adequately. It is already known that the MAXC method may underestimate M_c (Wössner and Wiemer, 2005; Mignan *et al.*, 2011) unless spatiotemporally confined samples are used (not the focus of our study). The MBASS method is considered to estimate M_c more conservatively (Mignan and Wössner, 2012). Here, we showed that these two methods provide much lower M_c estimates than the Lilliefors test in every investigated HR catalog, which strongly affects the estimated *b*-value. For TM-based catalogs in particular, $M_c^{\mathrm{Lilliefors}}$ is either equal to the one of the network-based catalog (Fig. 3), or it indicates that the exponential MFD of small events differs from the one of larger events (Fig. 1d–f). Both cases imply that adding more small earthquakes with advanced detection methods does not preserve the exponential shape of the MFD.

According to our observations, HR catalogs should be used with caution for estimating any property of the MFD. In the following, we discuss the different kinds of MFD inconsistencies, when they compensate, and how to possibly overcome them.

Different kinds of MFD inconsistency and their origin

Our findings show that MFD inconsistencies have different origins and may be divided into three categories.

The first category contains MFDs with abrupt discontinuities; we observed those specifically in catalogs for the Ridgecrest sequence and entire southern California. These discontinuities are due to the mixture of different magnitude types. Because the magnitude composition changes over time for southern California, one or more discontinuities can occur at different magnitude levels (especially between M_L vs. M_w for more recent catalogs, and between M_{coda} vs. M_L for older catalog periods). The revised local magnitude (M_{Lr}) introduced by SCSN for 2016 onward to bring both M_L and M_w "into closer agreement" (SCEDC, 2016) apparently does not solve the issue or is not sufficient. It remains to be seen whether TM-based catalogs could become exponential down to M 2.0 without this discontinuity. For older catalog periods, a culprit for discontinuities is the irregular and mostly coarse binning of M_h.

Besides abrupt changes, MFDs can change gradually in slope toward low magnitudes. The second category is composed of MFDs characterized by an overrepresentation of low magnitudes with respect to an exponential distribution, as observed for Landers and L'Aquila. For Landers, it could be attributed to the M_{coda} scale. Although we cannot inspect the origin of this kind of inconsistency, we suspect that it is caused by data recording or processing issues leading to inappropriate magnitude estimates. Overrepresentation results in an increasing b-value with decreasing M_c. Like many catalog inconsistencies, this effect can induce fake time variations of the b-value (Fig. 6b and Fig. S9b), especially when obtained with common completeness estimation methods. The observed time dependence of this inconsistency can be explained with the improved completeness over time. Because low magnitudes are overrepresented compared with higher ones, their increasing contribution to the catalog over time makes the inconsistency more prevalent, changing the b-value.

The third category is composed of MFDs characterized by an underrepresentation of low magnitudes with respect to an exponential distribution. Underrepresentation (i.e., a gradual curvature in the MFD; Mignan, 2012) results in a continuously decreasing b-value with decreasing M_c. This effect is very dominant in more recent catalog periods for southern California including the TM-based catalogs. The explanation of underrepresentation is probably more challenging. We argue that the most likely origin is the mixture of spatiotemporally inhomogeneous (in)completeness. As shown for the sequences (Fig. 6 and Figs. S4–S10), the effect is dominant immediately after a large earthquake and vanishes over time, which is commonly known as short-term aftershock incompleteness (STAI, Kagan, 2004; Helmstetter *et al.*, 2006; Hainzl, 2016; de Arcangelis *et al.*, 2018). For instance, Kagan (2004) estimated that up to 28,000 early aftershocks after the Landers mainshock are missing (or two-thirds of M 2 events). Our observations corroborate the hypothesis of underreporting low-magnitude events. As other scientists have pointed out (Wiemer and Wyss, 2000; Wössner and Wiemer, 2005; Mignan *et al.*, 2011), a gradual curvature in regional catalogs (Fig. 3) can additionally arise from the spatial inhomogeneity of completeness due to the varying seismic network density. The contribution of the temporal evolution of the network is maybe a weaker factor, because $M_c^{MAXC}(+0.2)$, which is related to the strongest curvature in the MFD, decreases only marginally in the period from 2005–2010 to 2015–2020 (see Fig. 5).

In our observations, even TM-based methods can apparently not sufficiently improve the underreporting (Figs. 1 and 3)—possibly due to their selectiveness, that is, strong dependence on events in the network-based catalog. To investigate the influence of temporal incompleteness, we removed the period in which STAI is most evident—(until four days after the M 7.1 Ridgecrest mainshock; see Figs. S11 and S12). Accordingly, a strong gradual curvature remains in the MFDs and $M_c^{Lilliefors}$ of the TM-based catalogs is very close to the one of the network-based catalogs at $M_c \approx 1.65$ (except for the Shelly, 2020b catalog). This proximity indicates that (1) the abundance of small earthquakes from advanced detection methods does not necessarily make the MFD more exponential toward low magnitudes even in a more complete period; and (2) the underrepresentation in HR catalogs may last well beyond the short-term incompleteness. Moreover, these $M_c^{Lilliefors}$ relate only to short-lived exponentiality (Fig. S12c). When removing the first about nine days after the M 7.1 mainshock (see Figs. S13 and S14), $M_c^{Lilliefors}$ improved for the SCSN and Ross, Idini, *et al.* (2019) catalog to 1.34 and 0.90, respectively. The former change indicates an improved completeness and the latter relates again to a short-lived exponentiality (Fig. S14c). Worthy of note, incompleteness is not detected by the common methods to estimate M_c; even after removing STAI, their estimates are lower than $M_c^{Lilliefors}$, especially for the TM-based catalogs, leading to strongly biased b-values (Fig. S14b).

An additional explanation for the apparent underrepresentation of low magnitudes (which does not preclude the previous ones) is the scaling break of the (amplitude-based) local magnitude M_L, which, as shown by several studies (e.g., Bakun, 1984; Hanks and Boore, 1984; Ben-Zion and Zhu, 2002; Edwards *et al.*, 2010; Zollo *et al.*, 2014; Staudenmaier *et al.*, 2018; Lanzoni *et al.*, 2019), scales differently with M_w below M 2–4 (with $M_L \propto 1.5 M_w$) due to the attenuation of the higher frequency content in the medium (i.e., their corner frequencies remain constant) (Bethmann *et al.*, 2011; Munafò *et al.*, 2016; Deichmann, 2017). Antialiasing in the digital sampling process (an additional low-pass filter) can contribute to the scaling break (Uchide and Imanishi, 2018). Finally, we argue that even when accounting for this scaling break, a gradually curved MFD at very low magnitudes may remain, for example, as observed for induced seismicity (Herrmann *et al.*, 2019). Underrepresentation may, therefore, be further related to underlying physical processes such as a minimum rupture size (see also Ellsworth, 2019).

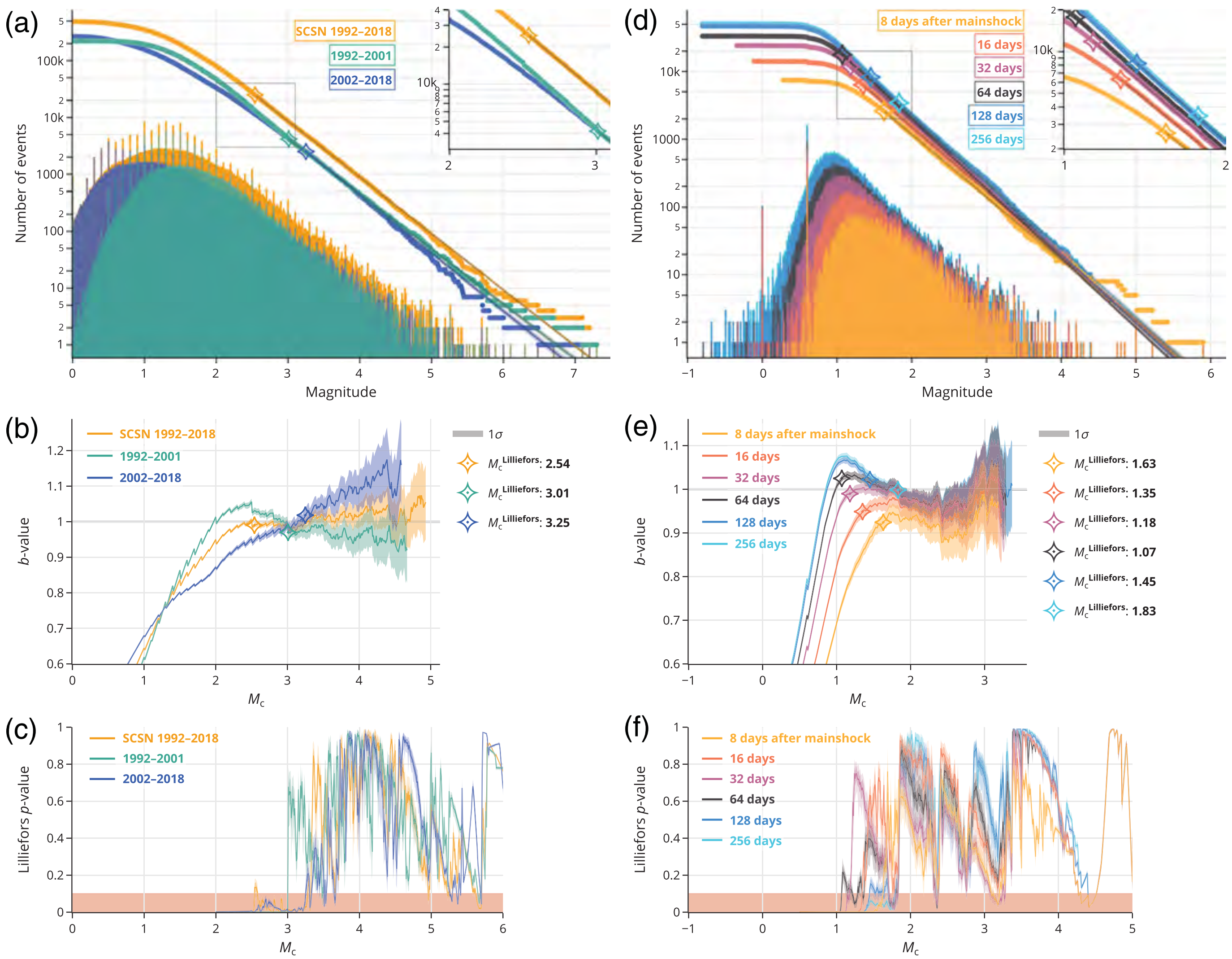


Figure 7. Demonstrating for two catalogs that the cancellation of over- and underrepresentation of low magnitudes can lead to an apparently wider exponential MFD. (a–c) Regional catalog of southern California for 1992–2018, in which $M_c^{\mathrm{Lilliefors}}$ is lower (yellow) than in the two individual time periods (blue and green, see legend). (d–f) Catalog of the L'Aquila sequence, in which $M_c^{\mathrm{Lilliefors}}$ depends on the time window length of the considered aftershock sequence.

Apparent compensation of inconsistencies

Sometimes, the over- and underrepresentation can cancel out and lead to an apparently (and unknowingly) wider exponential MFD when choosing an unfortunate time window, so that both effects are approximately in balance. For the entire southern California region, such a compensation can happen, for instance, in the period 1992–2018 (Fig. 7a–c). (We did not include the period after 2019 due to the inconsistency at M 3.5 outlined for the Ridgecrest sequence.) The compensation gives the impression of an apparent validity of the exponential GR relation at $M_c \approx 2.5$, although $M_c^{\mathrm{Lilliefors}}$ of the two individual time periods is higher (~3.0 and 3.25, respectively). A similar compensation may happen for the sequence-specific catalogs of L'Aquila and Landers, when the underrepresentation at early times after the mainshock due to STAI cancels out with the later overrepresentation. For L'Aquila, $M_c^{\mathrm{Lilliefors}}$ reaches its lowest level for a time window of 64 days after the mainshock (gray in Fig. 7d–f). For shorter or longer time windows of the aftershock sequence, the under- and overrepresentation dominates, respectively.

These compensation effects are an unfortunate consequence of the mixture of different nonexponential MFDs. It remains uncertain whether they can correct for the low-magnitude inconsistencies. Possibly, previous studies inferred lower completeness levels than reasonable.

A possible way to reduce MFD inconsistencies

Estimating magnitudes that produce a consistent MFD over a wide magnitude range appears to be a major challenge, and are currently a limiting factor to exploit HR catalogs in terms of magnitude statistics. Hence, magnitude estimation will require a different treatment than what is currently common practice. A possible remedy is to directly estimate moment magnitudes (M_w) for every earthquake. The estimation of M_w is well developed, robust with small uncertainty, and in principle consistent over the entire magnitude range (Deichmann, 2018). It has, therefore, been established as the standard magnitude scale, such as by the International Seismological Center (Di Giacomo *et al.*, 2015). M_w has the further benefit that it is directly related to earthquake source physics (e.g., seismic moment) and is therefore seismologically and physically well defined. Several studies have demonstrated for natural microearthquakes that a direct estimation of M_w (i.e., without magnitude regressions) is feasible for event sizes approaching M_w 0.0 (Atkinson *et al.*, 2014; Ross *et al.*, 2016; Moratto *et al.*, 2017; Staudenmaier *et al.*, 2018; Uchide and Imanishi, 2018; Butcher *et al.*, 2020). However, also the estimation of M_w and source parameters might practically not be free of biases and limitations, for example, due to near-surface amplifications at low frequencies especially for small events (Abercrombie and Leary, 1993).

Conclusions

Our study highlighted that HR catalogs usually do not preserve the exponential MFD that characterizes ordinary catalogs and that common methods to estimate the completeness magnitude, and consequently the b-value, underestimate severely the magnitude level below which the MFD departs from an exponential distribution. Moreover, MFDs of both HR catalogs based on the seismic network and on advanced detection methods depart from exponentiality at a similar magnitude level. These departures are mostly due to an improper mixing of different magnitude types, spatiotemporal incompleteness, or recording and processing issues. Another possible explanation is the intrinsic scaling break toward low magnitudes, such as for M_L. Observed inconsistencies make it necessary to set a considerably higher completeness level than often anticipated, for instance using Lilliefors' goodness-of-fit test with the exponential distribution as we did here.

Our findings have implications for both HR catalog producers and modelers that use MFDs of such catalogs. Modelers should be cautious when using HR catalogs that are composed of different magnitude types, span several orders of magnitude (especially below ~M 3), and cover wide spatiotemporal scales. The different kinds of inconsistencies outlined in our study for a selection of catalogs are usually not detected by common methods to estimate M_c, leading to strongly biased b-values and, as a consequence, to an inappropriate extrapolation of the rate of large earthquakes from low-magnitude events. This deficiency calls into question b-value-related studies that used those catalogs without a proper check of exponentiality. Moreover, the time dependence of inconsistencies introduces spurious b-value variations in time.

There is no doubt that the advent of HR catalogs brought great benefits in many aspects, but the results reported here may encourage HR catalog producers to evaluate carefully the homogeneity of the magnitude scales in their catalog (so that the MFD becomes consistent). Because it is not trivial to merge different magnitude scales into one consistent MFD, a possible solution may be to establish the estimation of M_w for each earthquake as common practice. Such an effort could reduce the observed and outlined inconsistencies and make the MFD more physically interpretable.

MFDs conceal inconsistencies more than it seems at first glance. Although they can be revealed and accounted for with deliberate methods like the one presented here, it may be more rewarding to make MFDs themselves more consistent, which would provide greater opportunities for the statistical analysis of existing and future catalogs.

Data and Resources

The southern California catalogs were downloaded from these repositories: Southern California Seismic Network (SCSN) (Southern California Earthquake Data Center [SCEDC], 2013, last accessed June 2020), Hauksson *et al.* (2012) (https://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html, version "1981–2019," last accessed June 2020), U.S. Geological Survey's Advanced National Seismic System (USGS-ANSS) ComCat (https://earthquake.usgs.gov/data/comcat, last accessed June 2020), quake template matching (QTM) of Ross, Trugman, *et al.* (2019) (https://scedc.caltech.edu/research-tools/QTMcatalog.html, last accessed June 2020), Ross, Idini, *et al.* (2019) (https://scedc.caltech.edu/research-tools/QTM-ridgecrest.html, last accessed June 2020), Shelly (2020b) (data release: Shelly, 2020a), and Lee *et al.* (2020) (http://bit.ly/2WswZQk, last accessed June 2020). The catalog of Valoroso *et al.* (2013) was provided by L. Chiaraluce (personal comm., November 2019). For the Lilliefors test, we used the implementation of statsmodels version 0.11.1 (https://www.statsmodels.org, last accessed June 2020; Seabold and Perktold, 2010). The supplemental material for this article includes further information and results referred to in the text, such as a summary of the magnitude–frequency distribution (MFD) inconsistencies for the SCSN catalog, MFD analyses in time windows during the aftershock sequence of Ridgecrest and Landers (as done for L'Aquila), and MFD analyses of the template matching (TM)-based Ridgecrest catalogs excluding the evident short-term incompleteness period. Our method to calculate $M_c^{\mathrm{Lilliefors}}$ is available as a Python class and demonstrated for an example catalog at DOI: 10.5281/zenodo.4162497. Data about Python graphing library are available at www.plotly.com/python (last accessed October 2020).

Acknowledgments

The authors sincerely appreciated the discussions with Ian Main, Egill Hauksson, and Greg Beroza. The authors also thank two anonymous reviewers for their comments, which helped improve the article. This study was supported by the "Real-time Earthquake Risk Reduction for a Resilient Europe" (RISE) project, funded by the European Union's Horizon 2020 research and innovation program (Grant Agreement Number 821115). The authors thank the creators and maintainers of plotly's Python graphing library (see Data and Resources).

References

Abercrombie, R., and P. Leary (1993). Source parameters of small earthquakes recorded at 2.5 km depth, Cajon Pass, southern California: Implications for earthquake scaling, *Geophys. Res. Lett.* **20,** no. 14, 1511–1514, doi: 10.1029/93GL00367.

Amorèse, D. (2007). Applying a change-point detection method on frequency–magnitude distributions, *Bull. Seismol. Soc. Am.* **97,** no. 5, 1742–1749, doi: 10.1785/0120060181.

Atkinson, G. M., D. W. Greig, and E. Yenier (2014). Estimation of moment magnitude (M) for small events (M < 4) on local networks, *Seismol. Res. Lett.* **85,** no. 5, 1116–1124, doi: 10.1785/0220130180.

Bakun, W. H. (1984). Seismic moments, local magnitudes, and coda-duration magnitudes for earthquakes in central California, *Bull. Seismol. Soc. Am.* **74,** no. 2, 439–458, available at https://pubs.geoscienceworld.org/ssa/bssa/article/74/2/439/102169 (last accessed June 2020).

Ben-Zion, Y., and L. Zhu (2002). Potency-magnitude scaling relations for southern California earthquakes with $1.0 < M_L < 7.0$, *Geophys. J. Int.* **148,** no. 3, F1–F5, doi: 10.1046/j.1365-246X.2002.01637.x.

Bethmann, F., N. Deichmann, and P. M. Mai (2011). Scaling relations of local magnitude versus moment magnitude for sequences of similar earthquakes in Switzerland, *Bull. Seismol. Soc. Am.* **101,** no. 2, 515–534, doi: 10.1785/0120100179.

Bohnhoff, M., M. Rische, T. Meier, D. Becker, G. Stavrakakis, and H. P. Harjes (2006). Microseismic activity in the Hellenic Volcanic Arc, Greece, with emphasis on the seismotectonic setting of the Santorini–Amorgos zone, *Tectonophysics* **423,** nos. 1/4, 17–33, doi: 10.1016/j.tecto.2006.03.024.

Bormann, P., and J. Saul (2009). Earthquake magnitude, in *Encyclopedia of Complexity and Systems Science*, Springer, New York, New York, 2473–2496, doi: 10.1007/978-0-387-30440-3_151.

Brodsky, E. E. (2019). The importance of studying small earthquakes, *Science* **364,** no. 6442, 736–737, doi: 10.1126/science.aax2490.

Bulut, F., M. Bohnhoff, W. L. Ellsworth, M. Aktar, and G. Dresen (2009). Microseismicity at the North Anatolian fault in the Sea of Marmara offshore Istanbul, NW Turkey, *J. Geophys. Res.* **114,** no. B9, doi: 10.1029/2008JB006244.

Butcher, A., R. Luckett, J. M. Kendall, and B. Baptie (2020). Seismic magnitudes, corner frequencies, and microseismicity: Using ambient noise to correct for high-frequency attenuation, *Bull. Seismol. Soc. Am.* **110,** no. 3, 1260–1275, doi: 10.1785/0120190032.

Chiaraluce, L., C. Chiarabba, C. Collettini, D. Piccinini, and M. Cocco (2007). Architecture and mechanics of an active low-angle normal fault: Alto Tiberina fault, northern Apennines, Italy, *J. Geophys. Res.* **112,** no. B10, doi: 10.1029/2007JB005015.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-law distributions in empirical data, *SIAM Rev.* **51,** no. 4, 661–703, doi: 10.1137/070710111.

deArcangelis, L., C. Godano, and E. Lippiello (2018). The overlap of aftershock coda waves and short-term postseismic forecasting, *J. Geophys. Res.* **123,** no. 7, 5661–5674, doi: 10.1029/2018JB015518.

Deichmann, N. (2017). Theoretical basis for the observed break in M_L/M_w scaling between small and large earthquakes, *Bull. Seismol. Soc. Am.* **107,** no. 2, 505–520, doi: 10.1785/0120160318.

Deichmann, N. (2018). The relation between M_E, M_L and M_w in theory and numerical simulations for small to moderate earthquakes, *J. Seismol.* **22,** no. 6, 1645–1668, doi: 10.1007/s10950-018-9786-1.

DiGiacomo, D., I. Bondár, D. A. Storchak, E. R. Engdahl, P. Bormann, and J. Harris (2015). ISC-GEM: Global instrumental earthquake catalogue (1900–2009), III. Recomputed M_s and m_b, proxy M_w, final magnitude composition and completeness assessment, *Phys. Earth Planet. In.* **239,** 33–47, doi: 10.1016/j.pepi.2014.06.005.

Ebel, J. E. (2008). The importance of small earthquakes, *Seismol. Res. Lett.* **79,** no. 4, 491–493, doi: 10.1785/gssrl.79.4.491.

Edwards, B., B. Allmann, D. Fäh, and J. F. Clinton (2010). Automatic computation of moment magnitudes for small earthquakes and the scaling of local to moment magnitude, *Geophys. J. Int.* **183,** no. 1, 407–420, doi: 10.1111/j.1365-246X.2010.04743.x.

Ellsworth, W. L. (2019). From foreshocks to mainshocks: Mechanisms and implications for earthquake nucleation, in *Mechanics of Earthquake Faulting*, A. Bizzarri, S. Das, and A. Petri (Editors), IOS Press, Amsterdam, The Netherlands, 95–112, doi: 10.3254/978-1-61499-979-9-95.

Fischer, T., and J. Horálek (2003). Space-time distribution of earthquake swarms in the principal focal zone of the NW Bohemia/Vogtland seismoactive region: Period 1985–2001, *J. Geodyn.* **35,** nos. 1/2, 125–144, doi: 10.1016/S0264-3707(02)00058-3.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** no. 7777, 193–199, doi: 10.1038/s41586-019-1606-4.

Habermann, R. E. (1987). Man-made changes of seismicity rates, *Bull. Seismol. Soc. Am.* **77,** no. 1, 141–159, available at https://pubs.geoscienceworld.org/ssa/bssa/article/77/1/141/118906 (last accessed June 2020).

Hainzl, S. (2016). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs, *J. Geophys. Res.* **121,** no. 9, 6499–6509, doi: 10.1002/2016JB013319.

Hainzl, S., T. Fischer, H. Čermáková, M. Bachura, and J. Vlček (2016). Aftershocks triggered by fluid intrusion: Evidence for the aftershock sequence occurred 2014 in West Bohemia/Vogtland, *J. Geophys. Res.* **121,** no. 4, 2575–2590, doi: 10.1002/2015JB012582.

Hanks, T. C., and D. M. Boore (1984). Moment-magnitude relations in theory and practice, *J. Geophys. Res.* **89,** no. B7, 6229–6235, doi: 10.1029/JB089iB07p06229.

Hatzfeld, D., V. Karakostas, M. Ziazia, I. Kassaras, E. Papadimitriou, K. Makropoulos, N. Voulgaris, and C. Papaioannou (2000). Microseismicity and faulting geometry in the Gulf of Corinth (Greece), *Geophys. J. Int.* **141,** no. 2, 438–456, doi: 10.1046/j.1365-246x.2000.00092.x.

Hauksson, E., W. Yang, and P. M. Shearer (2012). Waveform relocated earthquake catalog for southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.* **102,** no. 5, 2239–2244, doi: 10.1785/0120120010.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-dependent earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.* **96,** no. 1, 90–106, doi: 10.1785/0120050067.

Herrmann, M., T. Kraft, T. Tormann, L. Scarabello, and S. Wiemer (2019). A consistent high-resolution catalog of induced seismicity in Basel based on matched filter detection and tailored postprocessing, *J. Geophys. Res.* **124,** no. 8, 8449–8477, doi: 10.1029/2019JB017468.

Hutton, K., J. Woessner, and E. Hauksson (2010). Earthquake monitoring in southern California for seventy-seven years (1932–2008), *Bull. Seismol. Soc. Am.* **100,** no. 2, 423–446, doi: 10.1785/0120090130.

Improta, L., D. Latorre, L. Margheriti, A. Nardi, A. Marchetti, A. M. Lombardi, B. Castello, F. Villani, M. G. Ciaccio, F. M. Mele, *et al.* (2019). Multi-segment rupture of the 2016 Amatrice–Visso–Norcia seismic sequence (central Italy) constrained by the first high-quality catalog of early aftershocks, *Sci. Rep.* **9,** no. 1, 1–13, doi: 10.1038/s41598-019-43393-2.

Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.* **94,** no. 4, 1207–1228, doi: 10.1785/012003098.

Kamer, Y., and S. Hiemer (2015). Data-driven spatial b value estimation with applications to California seismicity: To *b* or not to *b*, *J. Geophys. Res.* **120,** no. 7, 5191–5214, doi: 10.1002/2014JB011510.

Kanamori, H. (1983). Magnitude scale and quantification of earthquakes, *Tectonophysics* **93,** nos. 3/4, 185–199, doi: 10.1016/0040-1951(83)90273-1.

Kijko, A., and A. Smit (2017). Estimation of the frequency–magnitude Gutenberg–Richter *b* value without making assumptions on levels of completeness, *Seismol. Res. Lett.* **88,** no. 2A, 311–318, doi: 10.1785/0220160177.

Lanzoni, A., L. Moratto, E. Priolo, and M. A. Romano (2019). Fast M_w estimation of microearthquakes recorded around the underground gas storage in the Montello–Collalto area (southeastern Alps, Italy), *J. Seismol.* **24,** 1029–1043, doi: 10.1007/s10950-019-09889-0.

Lee, E.-J., W.-Y. Liao, D. Mu, W. Wang, and P. Chen (2020). GPU-accelerated automatic microseismic monitoring algorithm (GAMMA) and its application to the 2019 Ridgecrest earthquake sequence, *Seismol. Res. Lett.* **91,** no. 4, 2062–2074, doi: 10.1785/0220190323.

Lilliefors, H. W. (1969). On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown, *J. Am. Stat. Assoc.* **64,** no. 325, 387–389, doi: 10.2307/2283748.

Martinsson, J., and A. Jonsson (2018). A new model for the distribution of observable earthquake magnitudes and applications to *b*-value estimation, *IEEE Geosci. Remote Sens. Lett.* **15,** no. 6, 833–837, doi: 10.1109/LGRS.2018.2812770.

Marzocchi, W., and L. Sandri (2003). A review and new insights on the estimation of the *b*-value and its uncertainty, *Ann. Geophys.* **46,** 1271–1282, doi: 10.4401/ag-3472.

Marzocchi, W., I. Spassiani, A. Stallone, and M. Taroni (2020). How to be fooled searching for significant variations of the *b*-value, *Geophys. J. Int.* **220,** no. 3, 1845–1856, doi: 10.1093/gji/ggz541.

Marzorati, S., M. Massa, M. Cattaneo, G. Monachesi, and M. Frapiccini (2014). Very detailed seismic pattern and migration inferred from the April 2010 Pietralunga (northern Italian Apennines) micro-earthquake sequence, *Tectonophysics* **610,** 91–109, doi: 10.1016/j.tecto.2013.10.014.

Meng, X., and Z. Peng (2016). Increasing lengths of aftershock zones with depths of moderate-size earthquakes on the San Jacinto fault suggests triggering of deep creep in the middle crust, *Geophys. J. Int.* **204,** no. 1, 250–261, doi: 10.1093/gji/ggv445.

Mignan, A. (2012). Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude, *J. Geophys. Res.* **117,** no. B8, doi: 10.1029/2012JB009347.

Mignan, A. (2014). The debate on the prognostic value of earthquake foreshocks: A meta-analysis, *Sci. Rep.* **4,** 1–5, doi: 10.1038/srep04099.

Mignan, A. (2019). Generalized earthquake frequency–magnitude distribution described by asymmetric Laplace mixture modelling, *Geophys. J. Int.* **219,** no. 2, 1348–1364, doi: 10.1093/gji/ggz373.

Mignan, A., M. J. Werner, S. Wiemer, C.-C. Chen, and Y.-M. Wu (2011). Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs, *Bull. Seismol. Soc. Am.* **101,** no. 3, 1371–1385, doi: 10.1785/0120100223.

Mignan, A., and J. Wössner (2012). Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, 1–45, doi: 10.5078/corssa-00180805.

Moratto, L., A. Saraò, and E. Priolo (2017). Moment magnitude (M_w) estimation of weak seismicity in northeastern Italy, *Seismol. Res. Lett.* **88,** no. 6, 1455–1464, doi: 10.1785/0220170063.

Munafò, I., L. Malagnini, and L. Chiaraluce (2016). On the relationship between M_w and M_L for small earthquakes, *Bull. Seismol. Soc. Am.* **106,** no. 5, 2402–2408, doi: 10.1785/0120160130.

Pearson, E. S., and H. O. Hartley (1972). *Biometrika Tables for Statisticians*, Vols. I and II**,** Cambridge University Press, Cambridge.

Piccinini, D., M. Cattaneo, C. Chiarabba, L. Chiaraluce, M. D. Martin, M. D. Bona, M. Moretti, G. Selvaggi, P. Augliera, D. Spallarossa, *et al.* (2009). A microseismic study in a low seismicity area of Italy: The Città di Castello 2000–2001 experiment, *Ann. Geophys.* **46,** no. 6, doi: 10.4401/ag-3476.

Ross, Z. E., Y. Ben-Zion, M. C. White, and F. L. Vernon (2016). Analysis of earthquake body wave spectra for potency and magnitude values: Implications for magnitude scaling relations, *Geophys. J. Int.* **207,** no. 2, 1158–1164, doi: 10.1093/gji/ggw327.

Ross, Z. E., E. S. Cochran, D. T. Trugman, and J. D. Smith (2020). 3D fault architecture controls the dynamism of earthquake swarms, *Science* **368,** no. 6497, 1357–1361, doi: 10.1126/science.abb0779.

Ross, Z. E., B. Idini, Z. Jia, O. L. Stephenson, M. Zhong, X. Wang, Z. Zhan, M. Simons, E. J. Fielding, S. H. Yun, *et al.* (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science* **366,** no. 6463, 346–351, doi: 10.1126/science.aaz0109.

Ross, Z. E., D. T. Trugman, E. Hauksson, and P. M. Shearer (2019). Searching for hidden earthquakes in southern California, *Science* **364,** no. 6442, 767–771, doi: 10.1126/science.aaw6888.

Schorlemmer, D., and J. Woessner (2008). Probability of detecting an earthquake, *Bull. Seismol. Soc. Am.* **98,** no. 5, 2103–2117, doi: 10.1785/0120070105.

Schwartz, D. P., and K. J. Coppersmith (1984). Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones, *J. Geophys. Res.* **89,** no. B7, 5681–5698, doi: 10.1029/JB089iB07p05681.

Seabold, S., and J. Perktold (2010). Statsmodels: Econometric and statistical modeling with python, *Proc. of the 9th Python in Science Conf.*, available at http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf (last accessed June 2020).

Shelly, D. R. (2020a). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: U.S. Geological Survey data release, doi: 10.5066/P9JN6H0N.

Shelly, D. R. (2020b). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity, *Seismol. Res. Lett.* **91,** no. 4, 1971–1978, doi: 10.1785/0220190309.

Shelly, D. R., W. L. Ellsworth, and D. P. Hill (2016). Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm, *J. Geophys. Res.* **121,** no. 3, 1776–1795, doi: 10.1002/2015JB012719.

Southern California Earthquake Data Center (SCEDC) (2013). Southern California Earthquake Data Center, *Caltech. Dataset*, doi: 10.7909/C3WD3xH1.

Southern California Earthquake Data Center (SCEDC) (2016). SCSN catalog change history, available at https://scedc.caltech.edu/eq-catalogs/change-history.html (last accessed June 2020).

Staudenmaier, N., T. Tormann, B. Edwards, N. Deichmann, and S. Wiemer (2018). Bilinearity in the Gutenberg–Richter relation based on M_L for magnitudes above and below 2, from systematic magnitude assessments in Parkfield (California), *Geophys. Res. Lett.* **45,** no. 14, 6887–6897, doi: 10.1029/2018GL078316.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons, *J. Am. Stat. Assoc.* **69,** no. 347, 730–737, doi: 10.2307/2286009.

Tinti, S., and F. Mulargia (1987). Confidence intervals of *b* values for grouped magnitudes, *Bull. Seismol. Soc. Am.* **77,** no. 6, 2125–2134, available at https://pubs.geoscienceworld.org/ssa/bssa/article/77/6/2125/119039 (last accessed June 2020).

Tormann, T., S. Wiemer, and E. Hauksson (2010). Changes of reporting rates in the Southern California Earthquake Catalog, introduced by a new definition of M_L, *Bull. Seismol. Soc. Am.* **100,** no. 4, 1733–1742, doi: 10.1785/0120090124.

Tormann, T., S. Wiemer, and A. Mignan (2014). Systematic survey of high-resolution *b* value imaging along Californian faults: Inference on asperities, *J. Geophys. Res.* **119,** no. 3, 2029–2054, doi: 10.1002/2013JB010867.

Uchide, T., and K. Imanishi (2018). Underestimation of microearthquake size by the magnitude scale of the Japan Meteorological Agency: Influence on earthquake statistics, *J. Geophys. Res.* **123,** no. 1, 606–620, doi: 10.1002/2017JB014697.

Valoroso, L., L. Chiaraluce, D. Piccinini, R. Di Stefano, D. Schaff, and F. Waldhauser (2013). Radiography of a normal fault system by 64,000 high-precision earthquake locations: The 2009 L'Aquila (central Italy) case study, *J. Geophys. Res.* **118,** no. 3, 1156–1176, doi: 10.1002/jgrb.50130.

Waldhauser, F., W. L. Ellsworth, D. P. Schaff, and A. Cole (2004). Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett.* **31,** no. 18, doi: 10.1029/2004GL020649.

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011). High-resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.* **101,** no. 4, 1630–1648, doi: 10.1785/0120090340.

Wiemer, S., and D. Schorlemmer (2007). ALM: An asperity-based likelihood model for California, *Seismol. Res. Lett.* **78,** no. 1, 134–140, doi: 10.1785/gssrl.78.1.134.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan, *Bull. Seismol. Soc. Am.* **90,** no. 4, 859–869, doi: 10.1785/0119990114.

Wössner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** no. 2, 684–698, doi: 10.1785/0120040007.

Zollo, A., A. Orefice, and V. Convertito (2014). Source parameter scaling and radiation efficiency of microearthquakes along the Irpinia fault zone in southern Apennines, Italy, *J. Geophys. Res.* **119,** no. 4, 3256–3275, doi: 10.1002/2013JB010116.

Zúñiga, F. R., and M. Wyss (1995). Inadvertent changes in magnitude reported in earthquake catalogs: Their evaluation through *b*-value estimates, *Bull. Seismol. Soc. Am.* **85,** no. 6, 1858–1866, available at https://pubs.geoscienceworld.org/ssa/bssa/article/85/6/1858/102703 (last accessed June 2020).

Manuscript received 22 September 2020
Published online 16 December 2020

Geophysical Journal International

Geophys. J. Int. (2022) **230,** 2073–2088
Advance Access publication 2022 May 6
GJI Seismology

https://doi.org/10.1093/gji/ggac160

3-D spatial cluster analysis of seismic sequences through density-based algorithms

Ester Piegari, Marcus Herrmann and Warner Marzocchi

Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli Federico II, 80126 Naples, Italy.
E-mail: Piegari ester.piegari@unina.it

Accepted 2022 April 22. Received 2022 April 3; in original form 2022 January 10

SUMMARY

With seismic catalogues becoming progressively larger, extracting information becomes challenging and calls upon using sophisticated statistical analysis. Data are typically clustered by machine learning algorithms to find patterns or identify regions of interest that require further exploration. Here, we investigate two density-based clustering algorithms, DBSCAN and OPTICS, for their capability to analyse the spatial distribution of seismicity and their effectiveness in discovering highly active seismic volumes of arbitrary shapes in large data sets. In particular, we study the influence of varying input parameters on the cluster solutions. By exploring the parameter space, we identify a crossover region with optimal solutions in between two phases with opposite behaviours (i.e. only clustered and only unclustered data points). Using a synthetic case with various geometric structures, we find that solutions in the crossover region consistently have the largest clusters and best represent the individual structures. For identifying strong anisotropic structures, we illustrate the usefulness of data rescaling. Applying the clustering algorithms to seismic catalogues of recent earthquake sequences (2016 Central Italy and 2016 Kumamoto) confirms that cluster solutions in the crossover region are the best candidates to identify 3-D features of tectonic structures that were activated in a seismic sequence. Finally, we propose a list of recipes that generalizes our analyses to obtain such solutions for other seismic sequences.

Key words: Machine learning; Statistical methods; Seismicity and tectonics; Statistical seismology.

1 INTRODUCTION

In recent years, machine learning algorithms have been increasingly used in many different research fields due to the availability of large data sets and new software tools. Clustering is a type of unsupervised machine learning (Mehta *et al*. 2019; Bhattacharya 2021; Zhang *et al.* 2022) that groups data by means of a similarity measure. In the last decades, many clustering algorithms based on different similarity measures have been proposed (Kaufman & Rousseeuw 1990; Jain *et al.* 1999) and applied to a variety of scientific problems (Aggarwal & Reddy 2013; Lyra *et al.* 2014; Lindsey *et al.* 2018; Karpatne *et al.* 2019; Abdideh & Ameri 2020) with the aim of identifying hidden patterns in data. Regarding applications to seismicity, a fuzzy clustering algorithm was used to partition earthquake epicentres of Iranian seismic catalogues (Ansari *et al.* 2009), while approaches based on k-means (Ouillon *et al.* 2008), Gaussian Mixture models (Ouillon & Sornette 2011) and more recently agglomerative hierarchical clustering (Kamer *et al.* 2020) have been proposed for fault network reconstruction. Furthermore, Konstantaras *et al.* (2012), Schoenball & Ellsworth (2017) and Fan & Xu (2019) have applied the density-based (DB) algorithm DBSCAN for cluster analyses of earthquake epicentres, while Cesca *et al.* (2014), Cesca (2020) and Petersen *et al.* (2021) have developed a software tool based on DBSCAN for implementing multidimensional clustering that accounts for other properties (such as origin, times, focal mechanisms, moment tensors and waveform similarity).

The choice of the most appropriate clustering algorithm depends on the application at hand and is related to the definition of a cluster. Clusters are commonly identified either as groups of data that minimize the intracluster distance (and maximize intercluster distance) or as dense data regions separated by sparse regions. Here, we are interested in discovering spatial features of seismicity by density rather than distances between data points. This decision is crucial to identify clusters of arbitrary shapes and anisotropic structures in a 3-D space. Partitioning algorithms like k-means or Gaussian Mixture models instead minimize the distances between data points, which generally leads to identify convex (i.e. spherical) regions around denser groups of data points. Instead, DB connections among data points allow recognizing preferential alignments of anisotropic structures and provide information about their size (Ester *et al.* 1996). Another advantage of DB algorithms is their efficiency on large data sets compared to hierarchical clustering algorithms. Furthermore, DB clustering does not require every data point to be part of a cluster, which makes it possible to account for noise in data.

In the following, we will explore the two most popular DB clustering algorithms, DBSCAN (Ester *et al.* 1996) and its extension OPTICS (Ankerst *et al.* 1999). They are based on a simple set of instructions and require only two input parameters. The problem is that depending on the spatial distribution of earthquakes, even small changes of these parameters can lead to very different cluster solutions, ranging from many small to very few large clusters. For this reason, we explore the challenges in the calibration of these procedures to obtain stable cluster solutions. We deal with this sensitivity aspect by first exploring the whole parameter space and then discussing DB cluster solutions for different catalogues. Specifically, we perform cluster analyses of earthquake catalogues of the 2016 Kumamoto and 2016 Central Italy sequence and identify their main spatial features. Finally, on the basis of the findings from clustering, a tentative recipe with instructions to explore a seismic sequence and identify its main spatial features through DB algorithms is proposed. Then, an application to better characterize the region of the 2016 Kumamoto sequence where the main shocks occurred is illustrated. All the numerical analyses have been performed by using software packages available in the Statistics and Machine Learning Toolbox of MATLAB R2021a.

2 DB ALGORITHMS

2.1 DBSCAN

DBSCAN stands for *Density Based Spatial Clustering of Application with Noise* and was introduced by Ester *et al.* (1996) with the aim to discover clusters of arbitrary shapes in large spatial databases with noise. The algorithm is based on only two input parameters (see Fig. 1a): ε, the neighbourhood distance around a given point; and Z, the minimum number of points in a neighbourhood. Once the values of ε and Z are assigned, DBSCAN classifies data points, p, into three categories as follows:

(1)A a *core point*, if the number of points in its ε-neighbourhood, N_ε(p), is greater than or equal to Z, that is $N_\varepsilon(\mathrm{p}) \geq Z$.
(2)As a *boundary point*, if two conditions are satisfied: (i) the number of points in its neighbourhood is less than Z, that is $N_\varepsilon(\mathrm{p}) < Z$, (ii) p is in the ε-neighbourhood of a core point.
(3)As a *noise point*, if it is neither a core point nor a boundary point, that is N_ε (p) $< Z$.

Initially, DBSCAN searches for core points, assigns them a cluster index (hereafter called 'colour'), and gives the same colour to all core points that are in the ε-neighbourhood of each other. These points are called density connected core points (see Fig. 1a) and their spatial distribution determines the shape and the number of clusters. Boundary points take the colour of the nearest core point, while noise points are discarded. We notice that setting the values of ε and Z is equivalent to introducing a density threshold to influence which points become clustered. Thus, varying ε and Z corresponds to increase or decrease this threshold, that means clustering smaller or larger groups of points. Looking at the distribution of points in Fig. 1(a), for example, if $Z = 3$, all points belong to the same cluster except for one noise point; instead if $Z = 5$, the algorithm does not find any cluster because all points are noise points. One of the most striking features of this algorithm is that the cluster geometry is not predefined and clusters of any shape can be identified just grouping paths of density connected points. This is particular useful for cluster analyses of 3-D spatial distribution of earthquakes as it might be of help in discovering complex networks of fault systems.

Finally, we note that the number of clusters retrieved by DBSCAN does not depend on the order in which the data points are processed. Instead, boundary points might belong to adjacent clusters and the algorithm assigns them to the first discovered cluster.

2.2 OPTICS

OPTICS stands for *Ordering Points To Identify Clustering Structure* and is an extension of DBSCAN proposed by Ankerst *et al.* (1999). Actually, it is not a clustering algorithm but an ordering algorithm introduced to overcome the main drawback of DBSCAN, that is, not being able to distinguish regions with different densities. The basic idea is that for a given Z, denser clusters may be completely contained in clusters of lower density. Therefore, if higher density points are processed first, a clustering order can be obtained, which contains information about hierarchically nested clustering structures.

To identify the clustering structure, the algorithm computes for each point, p, two additional quantities called core distance, d_{C}, and reachability distance, d_{R}, as follows (see also Fig. 1b):

$$d_{\mathrm{C}}(\mathrm{p};\ \varepsilon, Z) = \begin{cases} \text{undefined if } N_\varepsilon(\mathrm{p}) < Z \\ \varepsilon' = \min(\varepsilon) \mid \ N_{\varepsilon'}(\mathrm{p}) \geq Z. \end{cases}$$

$$d_{\mathrm{R}}(\mathrm{p}, \mathrm{q};\ \varepsilon,\ Z) = \begin{cases} \text{undefined if } N_\varepsilon(\mathrm{p}) < Z \\ \max(\varepsilon',\ \mathrm{dist}(\mathrm{p}, \mathrm{q})) \text{ otherwise.} \end{cases}$$

In other words, for a given Z, d_{C} is the minimum neighbourhood distance (i.e. minimum ε) to make the point p a core point, whereas d_R between q and p is defined only if p is a core point, in which case d_R equals the maximum of d_{C} and the Euclidean distance between p and q. It is worth noting that the algorithm does not necessarily need the parameter ε because the search radius can span all possible values between zero and infinity, that is exploring all possible values for d_{C}. Practically, to save computation time, ε is set to a reasonably large value that serves as the maximum distance to consider.

The algorithm starts similar to DBSCAN with finding core points, but then explores new points in the order of lowest to highest d_{C}. The result is a reachability plot that represents d_{R} of each point as a function of the cluster-ordered list of points and provides information about the clustering structure. An example reachability plot is shown in Fig. 2 for a data set with 300 data points and five clusters. Such a graph can be considered as a special type of dendrogram (Sander *et al*. 2003), since the obtained clustering structure is hierarchical and indicates the existence of nested clusters. In Fig. 2(b), the points belonging to clusters have very low d_{R} (<1.5), and correspond to apparent 'valleys'; the smaller d_{R}, the denser are the corresponding clusters. The peaks represent points with larger d_{R} and separate individual clusters. The higher are the peaks, the more separated are the clusters. Clusters can be extracted from the reachability plot by selecting a threshold value of ε, that is drawing a horizontal line in Fig. 2(b). The number of valleys below such a threshold results in the exact same cluster solution as DBSCAN for the same ε and Z.

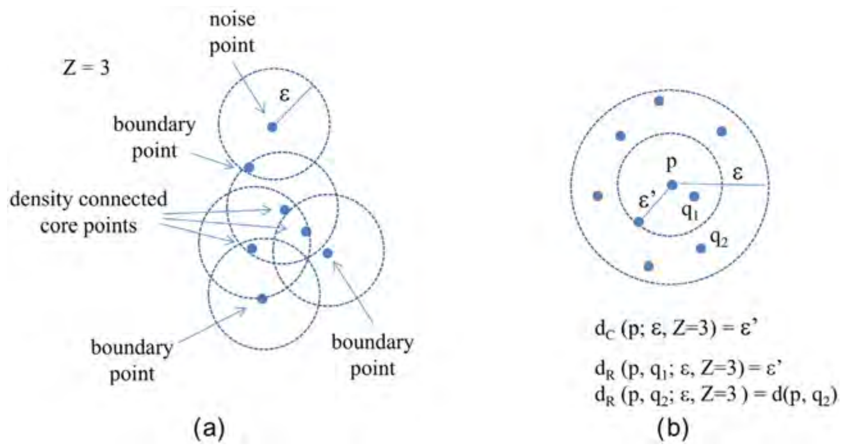


Figure 1. Graphical representation of (a) DBSCAN classification of data points basic concept and (b) OPTICS definitions of core and reachability distances.

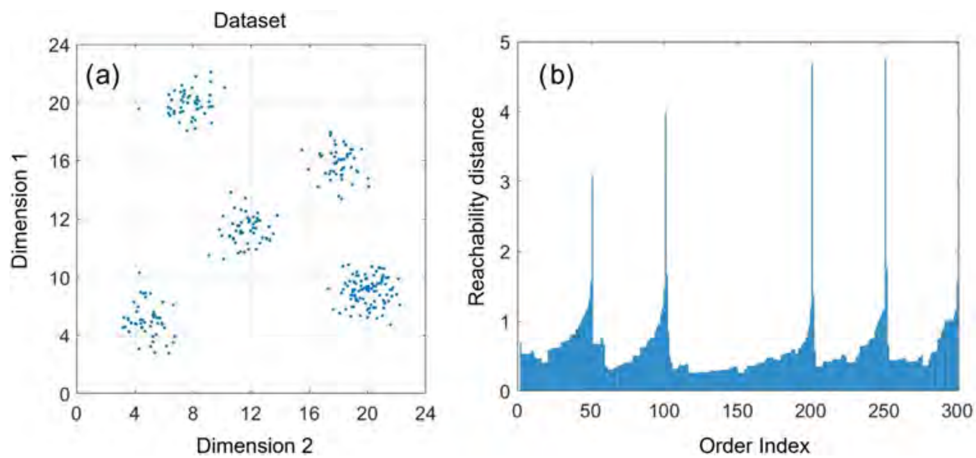


Figure 2. Example data set (a) and the corresponding reachability plot for $Z = 6$ (b).

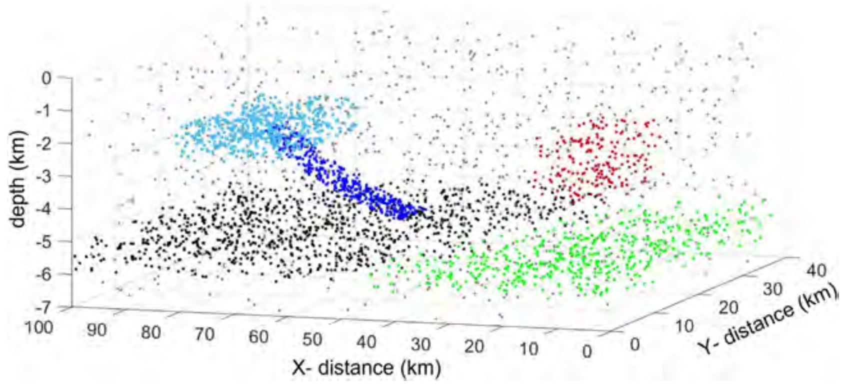


Figure 3. Spatial distribution of the synthetic data set consisting of five manually defined structures. The structures (coloured dots) and background activity in the whole volume (grey dots) are represented by uniformly distributed random points of varying density, totalling 3280 points.

3 APPLICATION TO A SYNTHETIC DATA SET

To illustrate how DB algorithms operate, we apply them to a synthetic data set of hypocentres. This analysis has multiple purposes summarized as follows:

(i) Illustrating how DB clustering works in principle by using an example with simple structures of known geometry.

(ii) Visualizing cluster solutions as function of the parameters.

(iii) Demonstrating that rescaling the data can help to recognize the largest structural features in presence of highly anisotropic structures.

3.1 Data set presentation

Fig. 3 shows the synthetic data set consisting of five manually defined large structures represented by uniformly distributed random

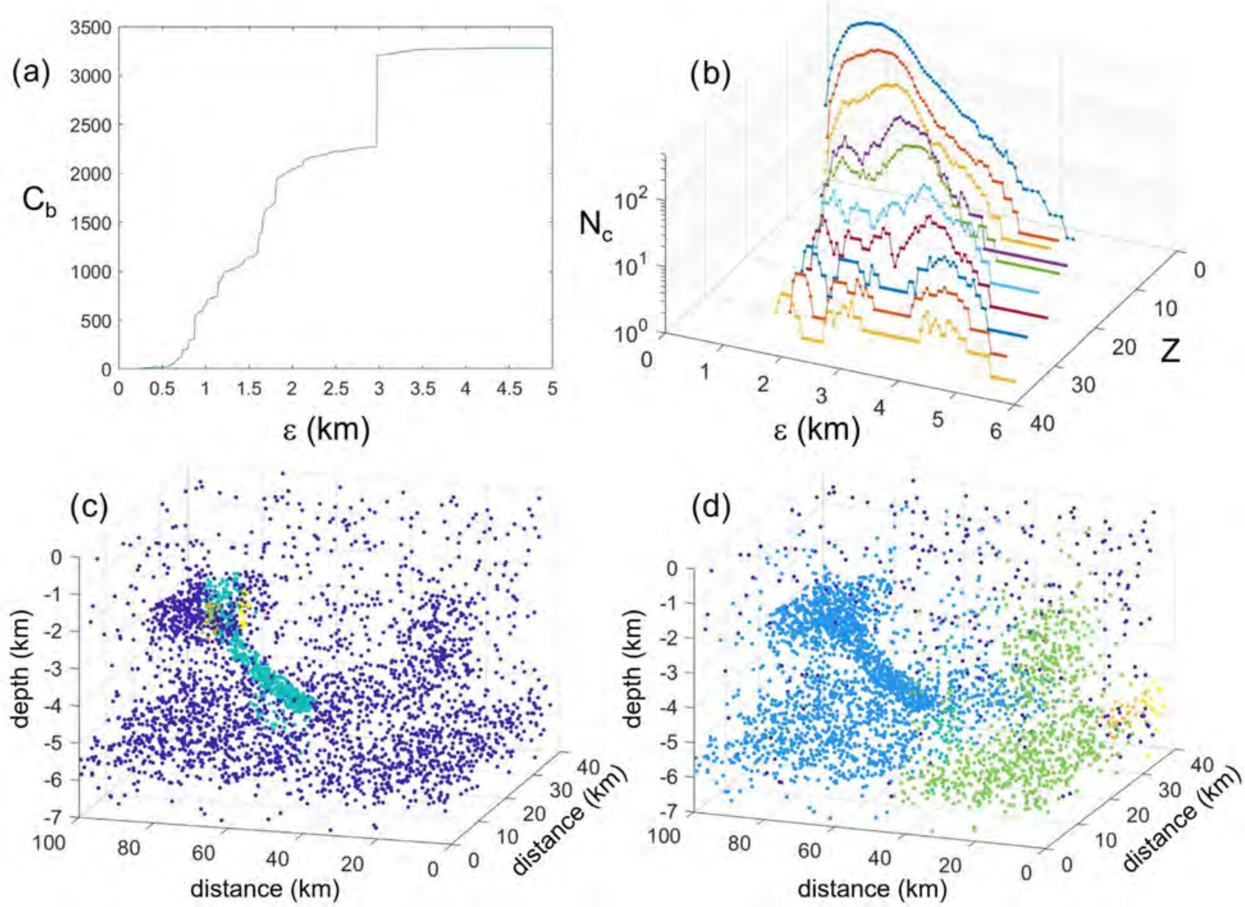


Figure 4. Exploring the influence of input parameters on DBSCAN solutions: (a) Number of points belonging to the biggest cluster, C_b, as a function of ε for $Z = 1$; (b) Number of clusters, N_c, as function of input parameters ε and Z; (c) DBSCAN solution for $\varepsilon = 1.4$ km and $Z = 15$; (d) DBSCAN solution for $\varepsilon = 3.4$ km and $Z = 15$. Dark blue points in (c) and (d) represent noise points and do not belong to any cluster.

points of varying density (2480 points in total). In addition, a uniform noise consisting of 800 uniformly distributed random points (about 25 per cent of the total points) was added to the whole volume (40 km × 100 km × 7 km) to represent background activity.

In particular, the synthetic geometric structures are polygonal regions representing (i) two planar structures at ∼6 km depth (black and green in Fig. 3, slightly shifted in depth), which extend up to 40 km and 60 km horizontally, respectively and about 1 km vertically; (ii) a shallow planar structure occupying a volume of about $10 \times 30 \times 1$ km^3 (cyan in Fig. 3); (iii) an inclined surface extending for about 2 km in depth and connecting two other structures (dark blue in Fig. 3) and (iv) a square prism-shaped volume of about 2 km height (red in Fig. 3). Choosing these structures has been motivated by the following reasons: (i) shallow and deep planar structures with different orientations mimic horizontal planes associated with thrust shear zones; (ii) intersections between structures mimic intersecting faults; (iii) strong anisotropy mimics larger sequences that propagate along a fault system and (iv) various orientations and overall 3-D interconnectedness mimics a fractured volume without preferential fault planes.

3.2 Cluster solutions in the parameter space

DBSCAN provides a wide range of solutions with clusters differing in number, shape and size depending on the value of ε and Z. For $Z = 1$, all points become clustered (i.e. belong to one or more clusters). Hints about the number of the largest structures can be derived from the number of stepwise increases of the biggest cluster size, C_b, as a function of ε. The behaviour of $C_b(\varepsilon, Z = 1)$ for the synthetic data set is shown in Fig. 4(a). $C_b(\varepsilon, Z = 1)$ grows step-wise every time a clustered region joins the biggest cluster. By increasing ε, the density threshold, Z/ε, for identifying core and boundary points decreases, leading to the clustering of larger regions with lower density. Small jumps in C_b indicate that small and dense regions are incorporated into the biggest cluster. Bigger jumps in C_b instead indicate the presence of large and dense regions that are spatially distant, as the clusters they belong to must increase in size before joining the biggest cluster. This is more easily understood for a data set with two dense regions that are separated by a large gap. By increasing ε, two big clusters in each of the two regions will form and continue to increase in size (simultaneously and independently of each other) until they merge. At this point, the larger the spatial distance between the two dense regions, the larger the corresponding stepwise increase in C_b will be (because the ε range in which both clusters grow separately increases with the separation gap). Therefore, jumps in C_b are controlled by the size and spatial distance of dense regions.

For small Z (< 5), the number of clusters, N_c, typically becomes very large and then gradually decreases to 1 for increasing ε (see Fig. 4b). For larger Z, $N_c(\varepsilon)$ becomes more complex including minor fluctuations before reaching 1 for large ε.

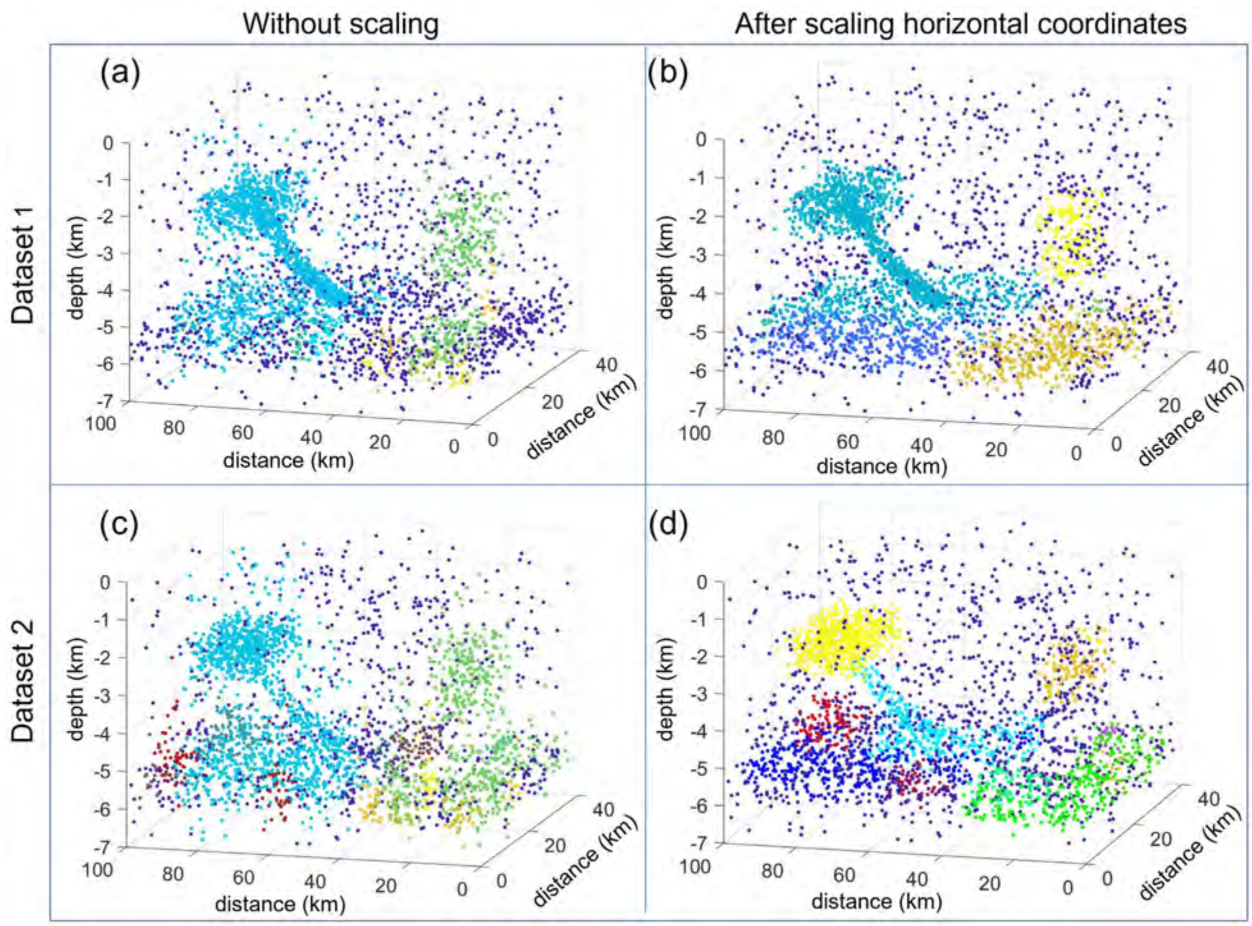


Figure 5. Influence of data scaling on DBSCAN solutions of two synthetic data sets for $Z = 15$. The two data sets differ only in the number of random points representing the inclined surface intersecting horizontal structures. (a) $\varepsilon = 2.6$ km; (b) $\varepsilon = 0.45$ km; (c) $\varepsilon = 2.8$ km and (d) $\varepsilon = 0.43$ km.

If ε is small, Z/ε becomes large, causing only regions with locally high densities to become clustered; most points are classified as noise. If ε is large, Z/ε becomes small, causing an inclusion of less dense regions into the clustering and most points to end up in the biggest cluster. Examples for these two extreme cluster solutions are shown in Figs 4(c) and (d), respectively: Fig. 4(c) illustrates that the region with the highest density becomes clustered, whereas Fig. 4(d) illustrates a separation of the large horizontal structures at depth, which resembles a characteristic feature of the synthetic data set.

Examples of cluster solutions for intermediate values of the threshold density Z/ε are reported in Fig. 5. In particular, Fig. 5(a) shows that DBSCAN produces two large clusters that do not separate shallow and deep structures. This limitation is related to the isotropic neighbour searching, that is processing points by using spheres of radius ε, for which even a small increase in ε leads to incorporate structures into the clusters that are outside the planar structures or linked to them. This can be more easily understood by focusing on the structures that form the big cyan cluster of Fig. 5(a). In presence of intersecting structures, like a planar structure and an inclined surface, DBSCAN is not able to distinguish them as individual structures even though the point density in the planar structure is high enough and the value of the neighbourhood search radius ε does not exceed its thickness. This indiscernibility happens for two main reasons: (i) decreasing ε while Z is kept fixed leads to a considerable increase of noise points (see Fig. 4c) and (ii) DBSCAN gives the same colour to paths of density connected points of any shape, therefore making intersecting structures inseparable unless they are characterized by different densities.

In an attempt to overcome this limitation, we scaled the data by homogenizing horizontal and depth ranges before clustering, using the latter as a reference (here: 0–7 km). To translate each coordinate individually to a common range, we used the min–max scaling for each horizontal coordinate x:

$$x_{new} = \frac{(max_{new} - min_{new})}{(max_{old} - min_{old})}\,(x_{old} - min_{old}) + min_{new}, \tag{1}$$

with $\min_{\mathrm{new}} = 0$ km and $\max_{\mathrm{new}} = 7$ km. For intermediate values of the threshold density Z/ε, Fig. 5(b) shows a cluster solution after applying this scaling, performing the clustering with DBSCAN and mapping the results back to the original space. In this case, clustering is more effective in resolving the shallow planar structure (yellow points) and one of the two horizontal structures (brown points). However, the shallow planar structure together with the inclined structure and a large part of a deep horizontal structure still belong to the same cluster (cyan cluster). The scaling-based cluster analysis fails in this part because the point density within the inclined structure is very high. To demonstrate the influence of this high density, we repeat the analysis for a subset of the synthetic data set in which the inclined structure has only 25 per cent of the points, that is a four times lower density (see Figs 5c and d).

Accordingly, clustering without data scaling is again not able to discriminate shallow and deep structures, whereas they can be

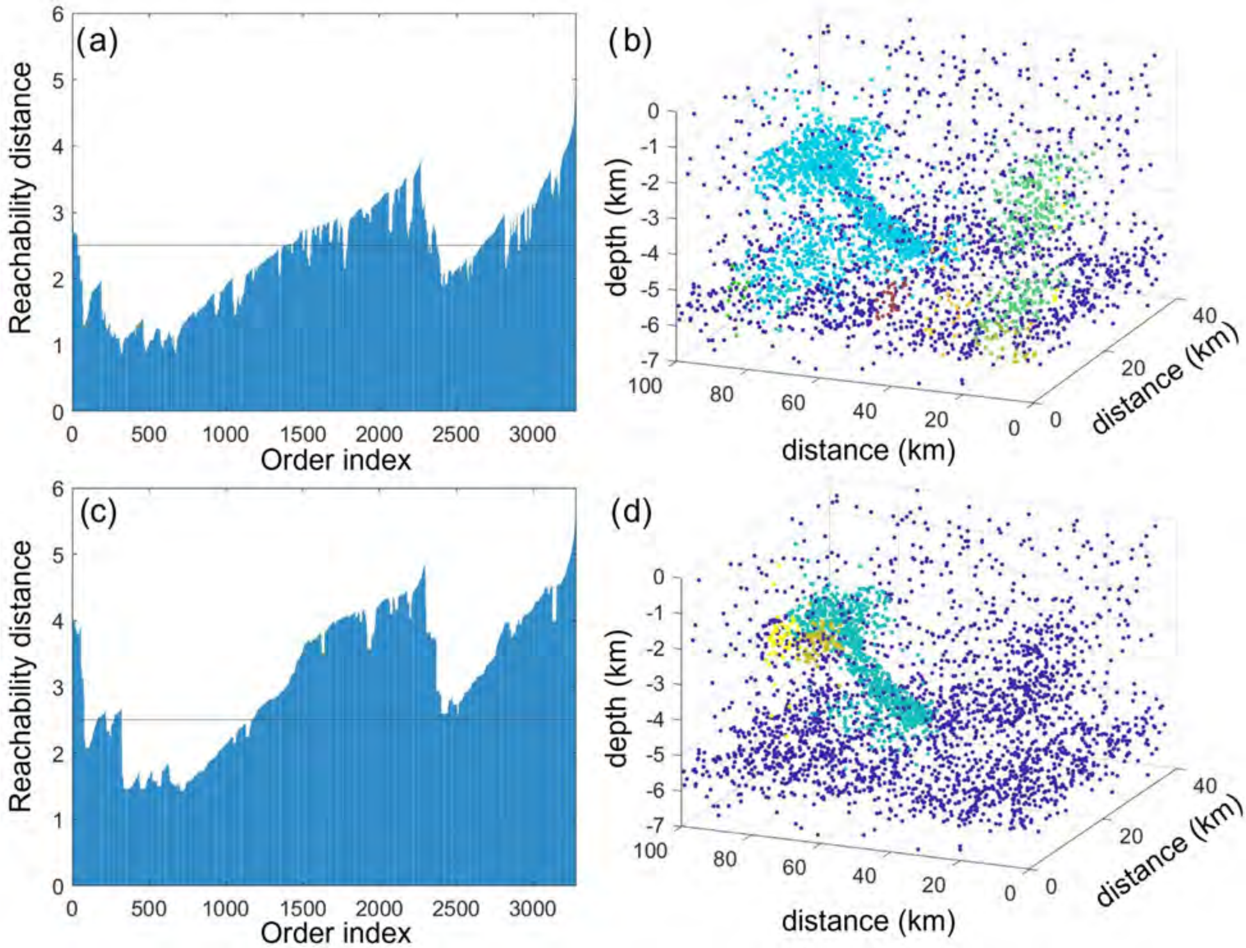


Figure 6. Reachability plots of the OPTICS algorithm for the synthetic data set in (a) and (c), and corresponding DBSCAN solutions for $\varepsilon = 2.5$ km in (b) and (d), respectively. Figures in the top row relate to $Z = 15$ and those in the bottom row to $Z = 30$. Dark blue points are noise points.

identified when data are scaled beforehand, even if the inclined structure is still not well defined. It is worth noting that the min–max scaling has some caveats. For instance, it may amplify the effect of local depth uncertainties and does not preserve the absolute distances among event pairs when the spatial boundary of the catalogue changes (but here we do not consider temporal changes of the catalogue). We suggest the min–max scaling only in the presence of strong anisotropies, that is when the horizontal extension of large dense regions is much larger than the vertical extension, $Lx,y/Lz >> 1$. Figs 5(b) and (d) show that the scaling-based cluster analysis fails in identifying intersecting structures as distinct objects if their contrast in point density values is not high enough. However, we think that such a scaling is useful to identifying planar structures that could be at least partially hidden by the uniform point distributions in depth caused for instance by uncertainties.

For two different choices of parameter Z (15 and 30), Fig. 6 shows reachability plots (left-hand column) and the corresponding DBSCAN solutions for $\varepsilon = 2.5$ km (right-hand column). The comparison reveals that a small Z produces more small-scale valleys in the reachability plot than a larger Z, which reduced their widths. Accordingly, a smaller Z results in a larger number of clusters because the ε threshold crosses more valleys horizontally than for larger Z. Although it is theoretically possible to get information about the number of characteristic structures by simply counting the number of the crossed valleys, practically this is not an easy task if Z is too small, because the meaning of a valley may be ambiguous. Both reachability plots reveal two main valleys, which can be considered as the main features of the data set. Such valleys do not correspond to any of the five manually defined structures but contain them. In particular, the relative locations of the main valleys suggest a spatial separation that divides the investigated volume into two parts, which are illuminated in Fig. 5(d) by cyan and green dots, respectively. The fact that the ε threshold cannot cross all the nested clusters ('sub-valleys') inside the biggest clusters (main valleys) indicates that isotropic neighbour searching is not effective for our synthetic data set and that scaling improves its characterization in DB clustering.

4. REPRESENTING DB CLUSTER SOLUTIONS IN A PHASE DIAGRAM

DB algorithms provide cluster solutions that can vary greatly in size and shape depending on the values of ε and Z. So, the question is how to choose these parameter values. A common strategy for estimating an appropriate ε is to detect the 'knee' in a k-distance graph, which plots the distances of each point to its kth nearest point in sorted order (see Ester *et al.* 1996). However, this approach does not always return an optimal ε, especially when a certain number of large clusters is desired instead of a single big one. As noted by Cesca (2020), a general rule to determine the best value of ε and

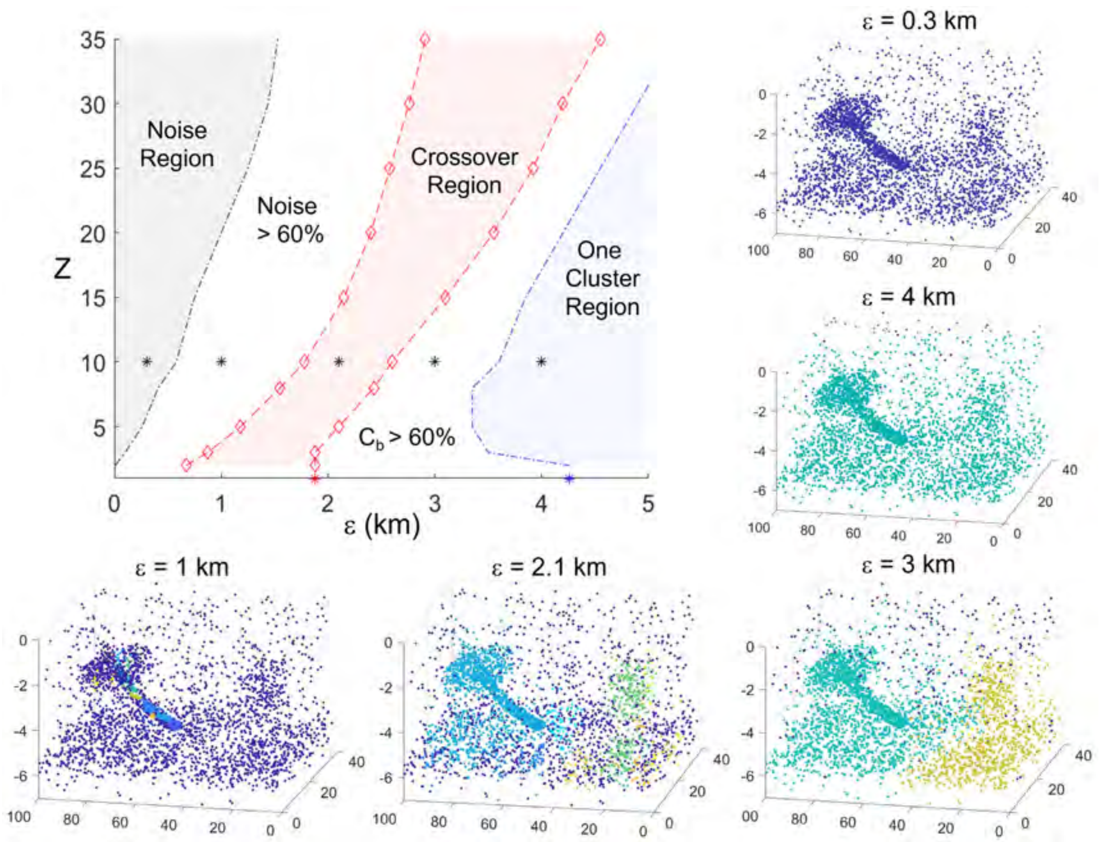


Figure 7. Phase diagram of DBSCAN solutions for the synthetic data set. The dotted–dashed and the red dashed lines divide the parameter space into five regions with different types of cluster solutions (see annotations and main text). Cluster solutions corresponding to the five points in the diagram for $Z = 10$ (marked by asterisks) are reported in separate subplots. In each subplot, dark blue points are noise points.

Z cannot be provided because DB algorithms are used for different purposes.

To get a better understanding of what types of information can be retrieved from DB clustering algorithms by varying the input parameters, we explore the whole space of solutions for synthetic and real seismic catalogues. The numerical analysis has shown that the phase diagram in the parameter space can be divided into five areas, which represent different classes of cluster solutions. As an example, Fig. 7 shows the phase diagram of the analysed synthetic data set. At the opposite ends of ε axis in the phase diagram, we find two extreme conditions: for very low ε all data points become noise points, whereas for very high ε all data points become connected to a single cluster. Moving horizontally from right to left in the phase diagram (i.e. decreasing ε and increasing the density threshold Z/ε), the size of the biggest cluster, C_b, decreases and other clusters appear. The locations of the jumps in C_b occur every time it splits into two or more clusters. Based on this behaviour, we obtain a first critical ε value when C_b contains 60 per cent of all points, that is for larger ε, cluster solutions are characterized by a big cluster that contains more than 60 per cent of the data. Similarly, by moving from left to right in the phase diagram (i.e. increasing ε and decreasing the density threshold Z/ε), we obtain another critical ε value when 60 per cent of the data are noise points, that is for lower ε, cluster solutions are characterized by more than 60 per cent of noise points. For the synthetic data set presented in the previous section, examples of cluster solutions for which Noise > 60 per cent and $C_b > 60$ per cent are shown in Figs 4(c) and (d), respectively, for $Z = 15$. The two critical ε are determined for various Z to construct the phase diagram (red markers in Fig. 7). The area between them is a transition zone named 'crossover region', which represents cluster solutions with many large clusters. Cluster solutions in Fig. 5 all belong to the crossover region. We are most interested in cluster solutions belonging to this region since they maximize the number of large clusters and help us to identify volumes with the highest density (*natural clustering*). Note that by using other jumps in C_b (i.e. other splits of the biggest cluster), it is possible to reconstruct a cluster hierarchy and divide the crossover region into subregions that differ in the number of large, stable clusters—depending on the event density distribution.

For the special case $Z = 1$, there are no noise points, but it is still possible to define two critical ε values, above which the biggest cluster contains more than 60 per cent of the data (red star) and all data (blue star), respectively.

Generally, decreasing ε for a fixed Z leads to more clusters, while increasing Z for a fixed ε leads to a fewer clusters. However, the number of clusters as a function of ε and Z does not behave monotonic (see Fig. 4b).

It is worth noting that increasing the height of the horizontal ε threshold in the OPTICS' reachability plot is equivalent to moving from left to right in the phase diagram. Obtaining cluster solutions in combination with the reachability plot has the advantage of accounting for the nested clustering structure—because the reachability plot visualizes, for a fixed Z, all cluster solutions of DBSCAN for a broad range of ε values.

The phase diagram also shows that cluster solutions depend slightly on Z; an increase of Z generally leads to an increase of noise points and to clusters that are more convex

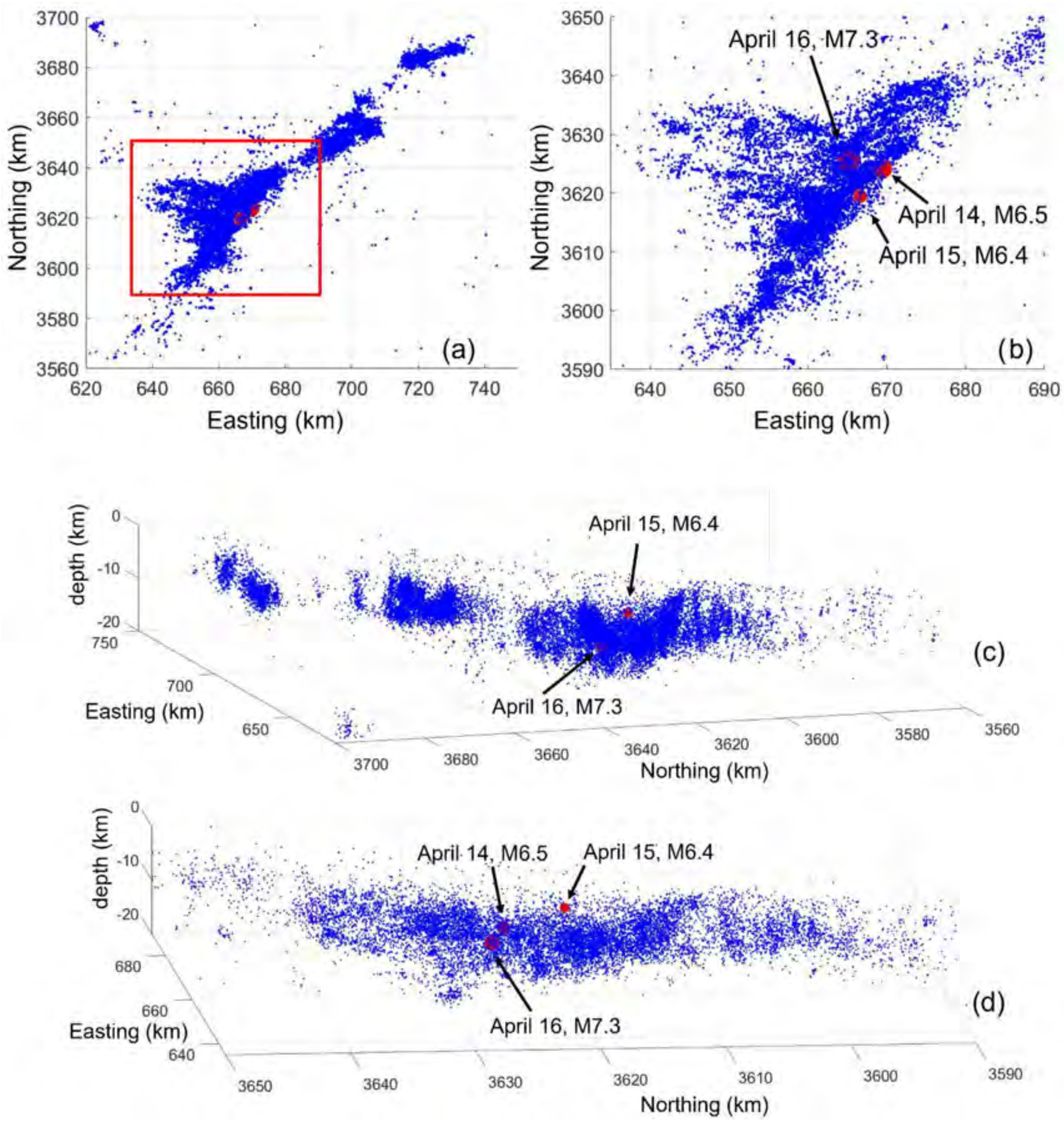


Figure 8. Overview of the 2016 Kumamoto sequence using catalogue extracts between 1 April 2016 and 31 August 2016. (a) Map view; (b) zoom into the area where the three largest earthquakes occurred [see red frame in '(a)']; (c) 3-D representation of (a); (d) 3-D representation of (b). The three largest earthquakes are represented with red markers and annotated with their magnitude and day of occurrence.

in shape. Recall that a larger Z produces less small-scale valleys in the reachability plot. Thus, even if the reachability plot helps in identifying the number of large clusters in the data set, the choice of Z still affects the characterization of the cluster hierarchy.

Finally, we want to point out that the areas covered by the five regions in the phase diagram of Fig. 7 depend on the spatial distribution of the data, which is an intrinsic property of a data set. Thus, changing the data distribution will change the width of the crossover region. Furthermore, exploring the whole parameter space is computationally expensive and practically unnecessary. Our analysis shows that only solutions in the crossover region are representative to extract meaningful information about the characteristic largest structures of a data set. In Section 6.1, we suggest a procedure for finding them, and selecting those with the desired level of nesting structure by using OPTICS.

5 APPLICATION TO REAL EARTHQUAKE CATALOGUES

5.1 The 2016 Kumamoto earthquake sequence

We performed DB cluster analysis of events that occurred between 1 April 2016 and 31 August 2016 (4 months) in the Kumamoto area, southwest of Japan. The earthquake catalogue was obtained from the Seismological Bulletin of Japan as provided by the Japan Meteorological Agency (JMA) and contains 163 988 events. We only use events with $M > 1$ and hypocentral depths shallower than 20 km within the spatial range of UTM coordinates from 621 to 745 km Easting and from 3564 to 3699 km Northing (WGS coordinates: 130.3–131.6°E, 32.2–33.4°N), totaling 20 887 events. Fig. 8 shows 2-D and 3-D representations of the hypocentral locations with the three largest earthquakes highlighted with a red marker (M6.5, M6.4 and M7.3).

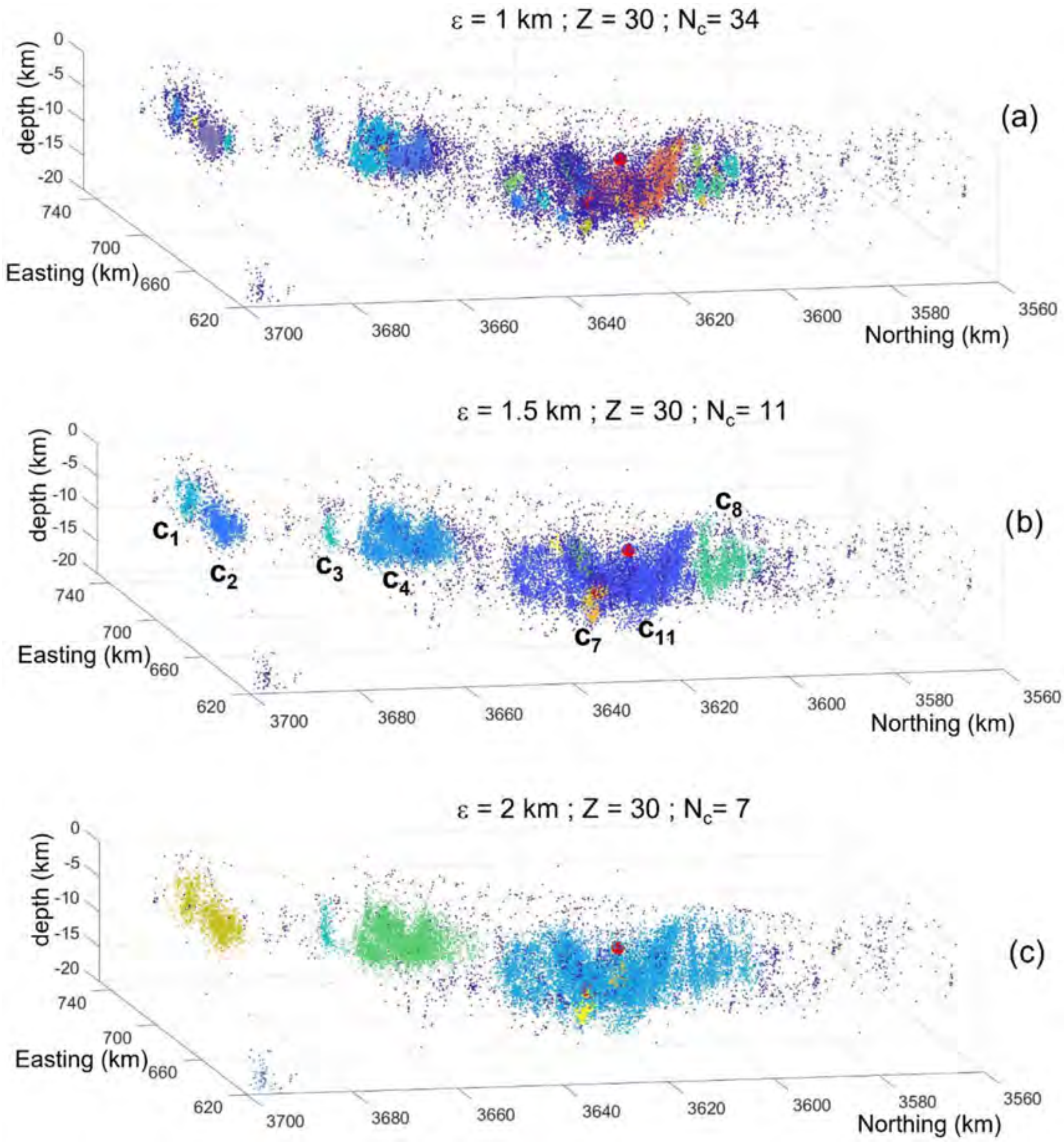


Figure 9. DBSCAN solutions of the 2016 Kumamoto sequence for $Z = 30$ and varying ε: (a) $\varepsilon = 1$ km, (b) $\varepsilon = 1.5$ km and (c) $\varepsilon = 2$ km. Red markers represent the location of the three largest earthquakes. Dark blue points represent noise points.

From Figs 8(a) and (c), we visually recognize a few big clusters as denser areas that are separated by areas with sparse seismicity. However, a zoom into the region where the largest earthquakes occurred (Figs 8b and d) blurs the sharp borders of the denser areas and reveals finer details, making a visual recognition of clusters ambiguous. With DB clustering, instead, we can divide the catalogue into natural groups in an exploratory way and identify patterns within it.

Fig. 9 shows three DBSCAN solutions for different choices of the input parameters inside the crossover region, that is for which both the number of noise points and C_b are less than 60 per cent of the data. These choices for $\varepsilon = 1$, 1.5 and 2 km divide the seismic sequence into 34, 11 and 7 clusters, respectively. With an increasing ε, the number of clusters and the number of noise points decrease, whereas the largest clusters increase in size by incorporating more adjacent hypocentres. Even though the shapes of the clusters change by varying ε, the centres of the largest clusters remain the same; the clusters always represent the most active zones and the largest earthquakes always belong to the same cluster (coluored orange, purple and light blue in Figs 9a–c, respectively).

Fig. 10 shows three reachability plots for $Z = \{15, 30, 70\}$ and the DBSCAN solution for $Z = 30$ and $\varepsilon = 3$ km, which is characterized by the presence of three largest clusters. These three clusters are evident in each of the three reachability plots as the deepest and best-defined valleys, named C1, C2 and C3, which can be considered as the main features of the sequence. As indicated in Figs 10(b) and (c), this cluster solution can be obtained for a wide range of ε, that is many horizontal lines lead to a division into three big clusters. However, for a small Z (Fig. 10a), the number of clusters increases significantly as revealed by the many narrow valleys inside the largest valleys. Consequently, a small variation of ε can lead to new clusters that include very little data due to the narrowness of the valleys.

The height of the peaks in the reachability plots represents another feature of the seismic sequence, namely the spatial separation of the clusters. In particular, the highest peaks correspond to the

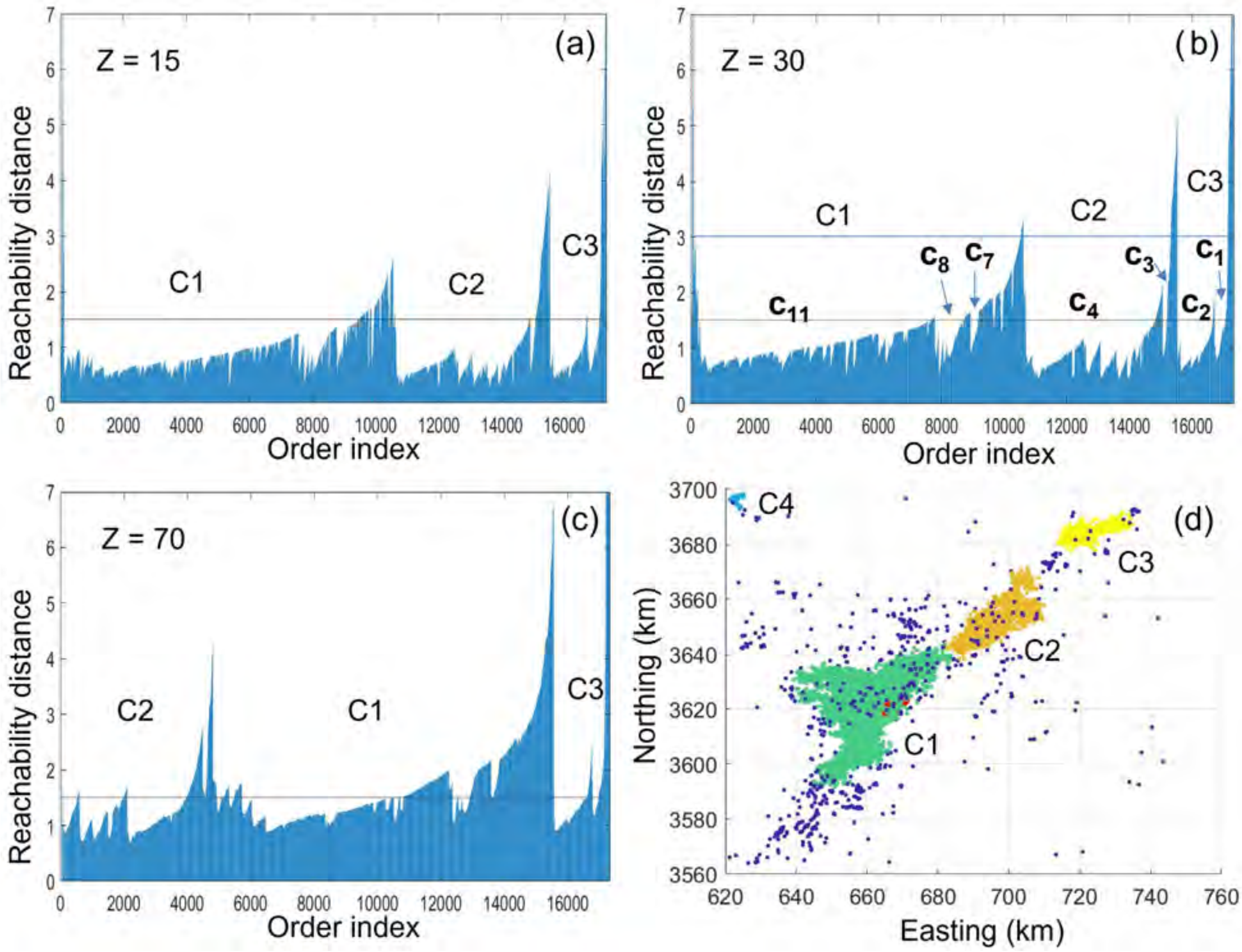


Figure 10. Cluster analysis of the 2016 Kumamoto sequence. (a–c) Reachability plots for different Z: (a) $Z = 15$, (b) $Z = 30$ and (c) $Z = 70$. Panel (d) shows a map view of the DBSCAN solution for $Z = 30$ and $\varepsilon = 3$ km. Red markers represent the location of the three largest earthquakes and dark blue points the noise points.

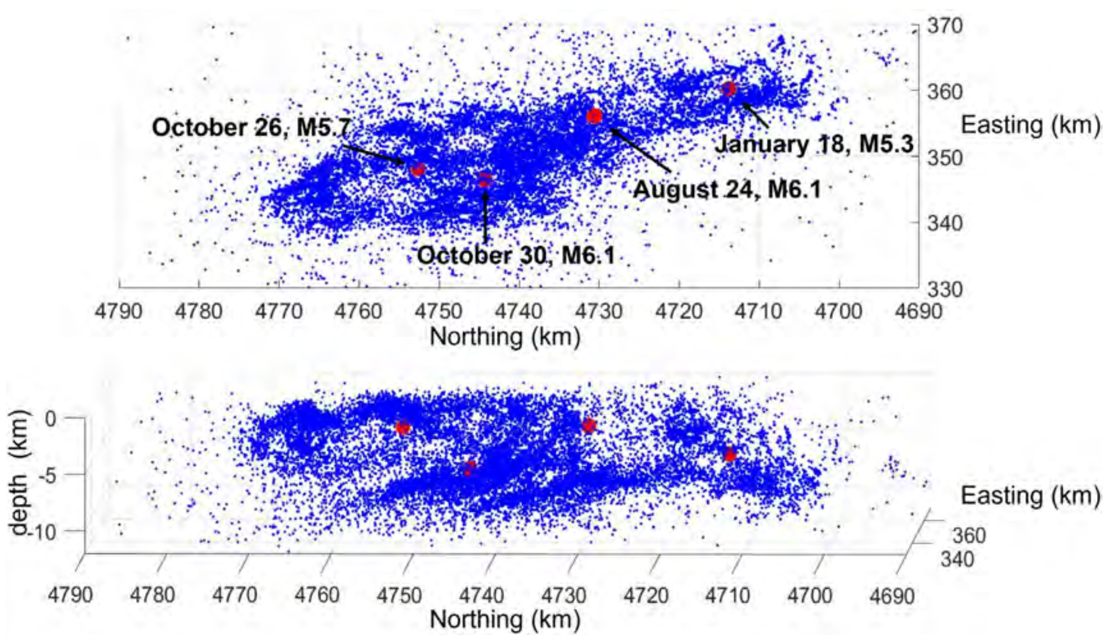


Figure 11. Overview of the 2016 Central Italy sequence between 15 August 2016 and 15 August 2017. Map view (top panel) and 3-D view (bottom panel). The four largest earthquakes are highlighted with a red marker and annotated with their magnitude and day of occurrence in the top panel.

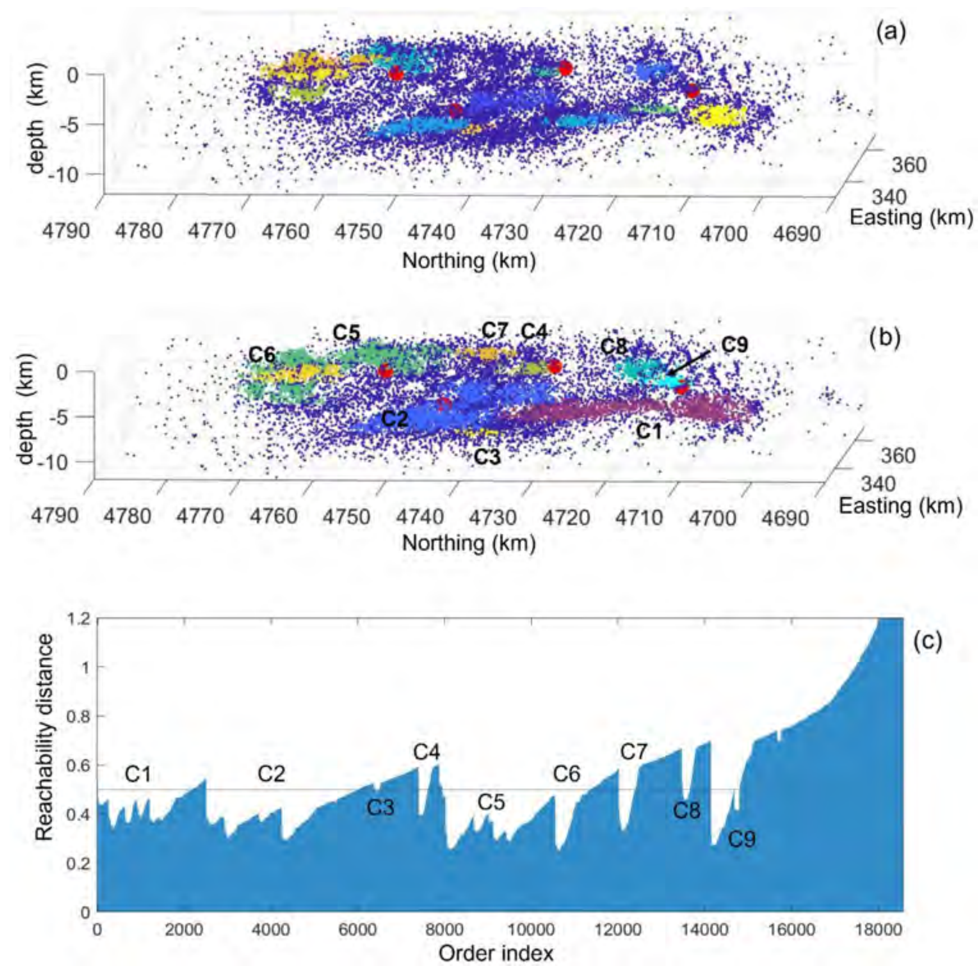


Figure 12. Cluster analysis of the 2016 Central Italy sequence by applying the DBSCAN algorithm for a fixed $Z = 100$. (a) and (b) Cluster solutions for $\varepsilon =$ 0.4 km (a) and $\varepsilon = 0.5$ km (b). Red markers represent the location of the four largest earthquakes. Dark blue points represent noise points. (c) Reachability plot with the black horizontal line corresponding to the ε threshold shown in (b). The annotated cluster names in (c) correspond to the ones in (b).

points with the largest d_R, which indicate the most separated clusters (see Section 2.2). The difference in height of the three main peaks in Figs 10(a)–(c) indicates that C1 and C2 are less spatially separated with respect to C3. In addition, Fig. 10(b) ($\varepsilon = 1.5$ km and $Z = 30$) lets us identify two well-defined nested structures related to the smaller valleys inside both C2 and C3 (named c_1, c_2, c_3 and c_4), which are visible as light blue and green coloured clusters in Fig. 9(b). Note that C4, which corresponds to a small event group in the northwestern sector of the area (see Fig. 10d), is not easily recognizable in the reachability plots due to its very deep and narrow valley on the left-hand side.

Finally, the horizontal $\varepsilon = 1.5$ km thresholds in the reachability plots produce a different number of clusters for different Z: 14 for $Z = 15$ (Fig. 10a), 11 for $Z = 30$ (Fig. 10b) and 12 for $Z = 70$ (Fig. 10c).

5.2 The 2016 Central Italy seismic sequence

For the 2016 Central Italy sequence, we used the high-resolution earthquake catalogue of Tan *et al*. 2021 spanning from 2016-08-15 to 2017-08-15. We only considered events with $M_w > 2$ and hypocentral depths shallower than 12 km within the spatial range of UTM coordinates from 330 to 370 km Easting and from 4690 to 4790 km Northing (12.9°–13.4°E, 42.3°–43.2°N), totalling 18 595 events (see Fig. 11). The locations of the four largest earthquakes are indicated with a red marker (M_w6.1 Amatrice event on 24 August 2016, M_w5.7 Visso event on 26 October 2016, M_w6.1 Norcia event on 30 October 2016 and M_w5.3 Campotosto event on 18 January 2017).

Since the data set is characterized by horizontally extended structures within a limited vertical range, the cluster analysis was applied after scaling the horizontal coordinates to the depth range (0–12 km) as discussed in Section 3. After remapping into the original coordinate system, Fig. 12 visualizes the obtained clusters for two different ε, but a fixed $Z = 100$. The first parameter set ($\varepsilon = 0.4$ km, $Z = 100$, see Fig. 12a) represents a cluster solution whose proportion of noise points is larger than 60 per cent, that is left of the crossover region. As expected, many small clusters are returned (13 clusters, maximally about 1000 events each). Instead, the second parameter set ($\varepsilon = 0.5$ km, $Z = 100$, see Fig. 12b) is located inside the crossover region and produces a balance between the amount of noise points and density-connected points, maximizing the number of large clusters.

Fig. 12(c) shows the reachability plot for the same $Z = 100$ and reveals several well-defined valleys corresponding to many high-density zones. The threshold $\varepsilon = 0.5$ km (black horizontal line in Fig. 12c) crosses nine valleys, which correspond to the DBSCAN

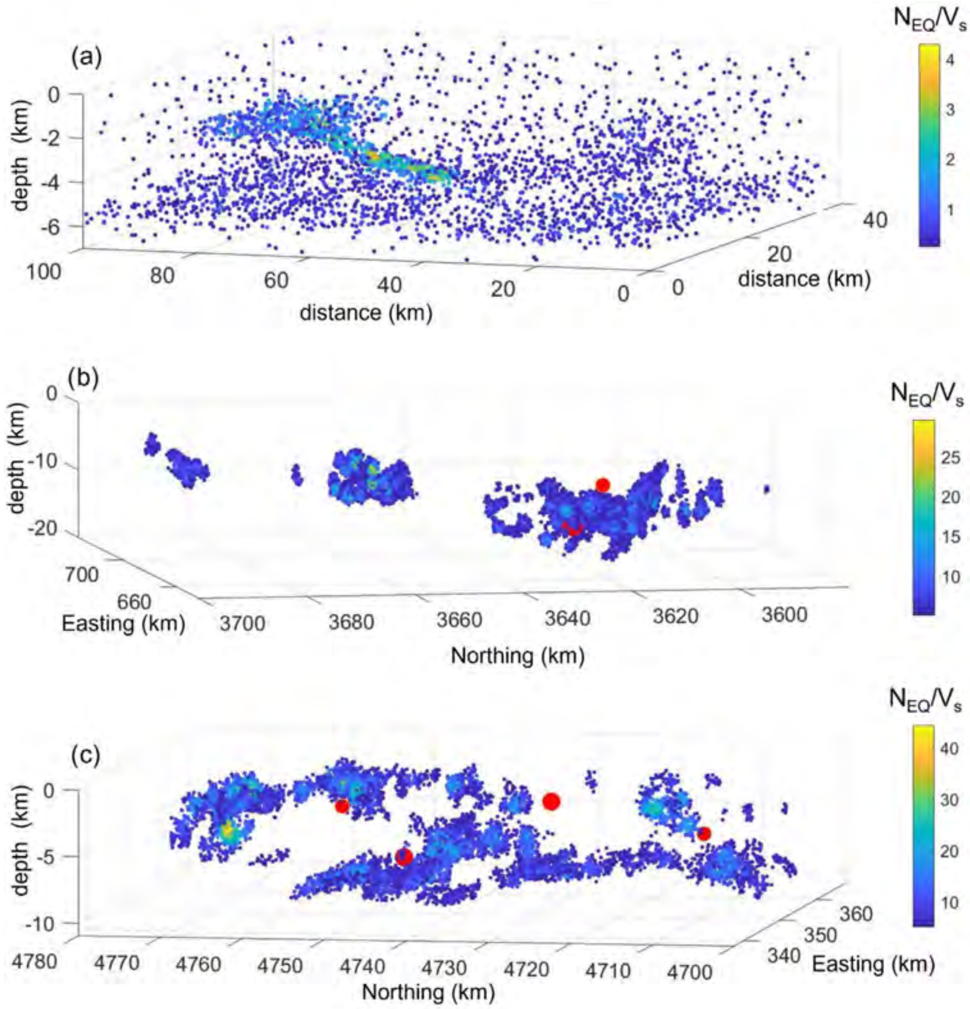


Figure 13. Hypocentre density of the (a) synthetic data set, (b) 2016 Kumamoto sequence and (c) 2016 Central Italy sequence. The hypocentre density (see colour bar) is represented at each event, but for the real cases only if it is larger than 5. Red markers represent the location of the largest earthquakes.

solution shown in Fig. 12(b). The reachability plot provides information not only on the presence of nested structures but also on the size and the number of the largest clusters. The main features of the catalogue are the three largest earthquake clusters, named C1, C2 and C5 in Figs 12(b) and (c). C1, which represents the extended structure at depth in the south, contains four smaller valleys. These four valleys were identified as individual clusters for a smaller $\varepsilon = 0.4$ km and are visible in Fig. 12(a) as clusters of different colour in this region. C2, which represents the extended structure at depth in the centre of the sequence and includes the Norcia main shock, is characterized by two larger and one smaller valleys—three substructures also visible in Fig. 12(a). C5, which represents a shallower structure in the north and contains the Visso event, contains five valleys of which three correspond to structures identified with $\varepsilon = 0.4$ km in this region (Fig. 12a). The spatial volumes illuminated by C1, C2 and C5 are also the main features of this catalogue with a lower magnitude cut-off ($M_w > 1.5$, totalling 76 055 events), which have been statistically analysed to characterize the behaviour of the magnitude distribution during and within this complex sequence (Herrmann *et al*. 2021). The remaining clusters in Fig. 12(b) either did not change significantly between the two parameter sets (e.g. C4, C6 and C8), or were added for the higher ε (e.g. C7).

6 A GENERALIZED APPROACH TO DB CLUSTER ANALYSIS AND A FURTHER APPLICATION

DB clustering algorithms undoubtedly facilitate the analysis of large catalogues by only using two input parameters. Yet, these two parameters can lead to a variety of cluster solutions, making their choice difficult. Ultimately, the preferred clustering solution depends on the purpose (i.e. the desired grouping of the data set), because a single best clustering solution does not exist. The cluster hierarchy of the catalogue can serve as key information for choosing the preferred solution and can be retrieved from the reachability plot of the OPTICS algorithm. Our analyses showed that parameter Z is crucial when the interest is in finding not only regions with the highest hypocentre density but also large clusters that represent the main structures. We have shown that parameter sets lying in the crossover region of the phase diagram are good candidates for exploring the catalogue in a meaningful way. However, finding all cluster solutions in the crossover region by exploring the entire parameter space is impractical (and needless) especially for large catalogues. Based on our findings and some general considerations, we describe below a recipe for finding a representative cluster solution in the crossover region.

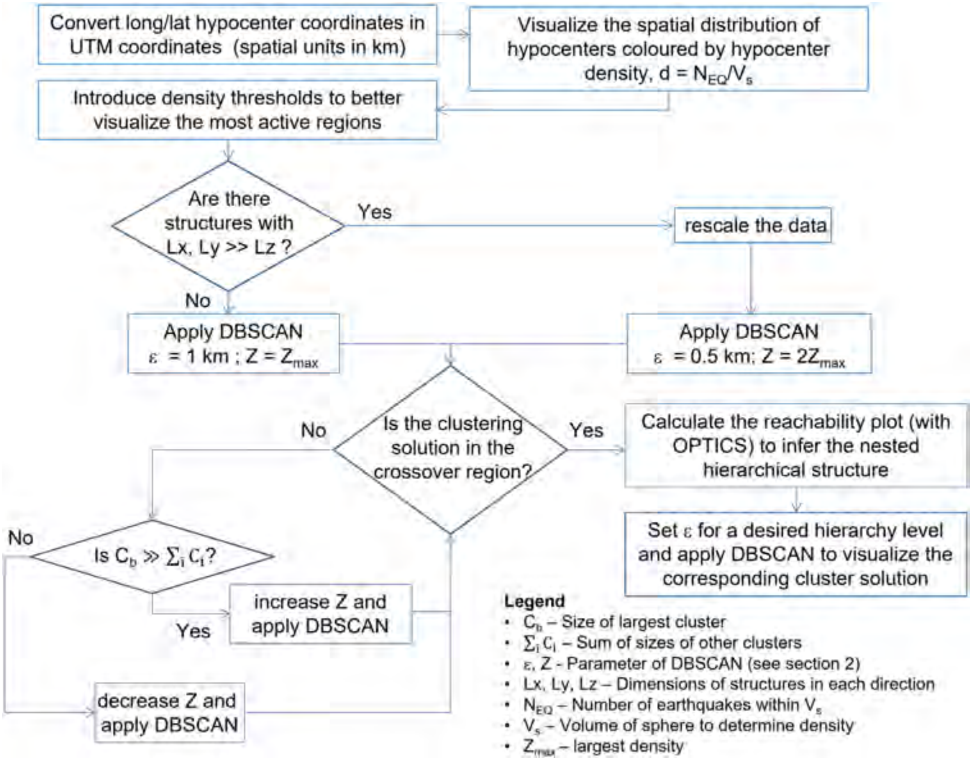


Figure 14. Flow diagram of our proposed DB cluster analysis of a seismic sequence.

6.1 A tentative recipe for finding cluster solutions in the crossover region

Because DB algorithms identify clusters as dense regions separated by sparse regions, their main drawback relates to the identification of cluster boundaries, that is if density drops are absent, cluster boundaries are not well defined. Therefore, we suggest their use with a foregoing inspection of the spatial distribution of earthquake hypocentres, which add information about their density and can help in the cluster analysis. Fig. 13 illustrates this for the investigated catalogues. The colour scale represents the hypocentre density defined as the number of earthquakes, N_{EQ}, in a sphere of radius 1 km, V_s. The number of events displayed in Figs 13(b) and (c) differs from Figs 8 and 11 because we have only visualized hypocentres for which the corresponding density is above a threshold of five events per V_s. This threshold simply avoids that many irrelevant events (in low-density areas) prevent the view of areas of interest (those that have high density). Fig. 13 highlights that density information is fundamental not only to locate the most active regions, but also to quantify the intensity of seismicity. In particular, the maximum value of the density, $Z_{\max}$, is approximately equal to 5, 30 and 45 for the synthetic data set and our extracted catalogues of the 2016 Kumamoto and Central Italy sequences, respectively. Interestingly, the most active regions in both investigated real cases do not include the largest earthquakes. We can use $Z_{\max}$ to find solutions in the crossover region. In Fig. 14, we propose a diagram that shows the main steps to obtain such solutions without performing an extensive exploration of the phase diagram.

Given an earthquake catalogue for a region of interest, the first step consists in converting the map units into km units (e.g. UTM coordinates), because an orthogonal coordinate system is required to correctly measure Euclidean distances between hypocentres. Then, the density of hypocentres needs to be computed and visualized for every event to infer the spatial distribution of hypocentres and obtain the $Z_{\max}$ and the geometry of dense regions. The cluster analysis starts with $\varepsilon = 1$ km and $Z = Z_{\max}$. If strong anisotropic structures characterize the data set, data scaling is suggested and to start cluster analysis with $\varepsilon = 0.5$ km and $Z = 2Z_{\max}$. We note that the evaluation of anisotropic structures is done retrospectively considering the spatial distribution of the whole seismic sequence and when the depth range of hypocentres significantly differs from their horizontal range. Such an evaluation becomes more feasible when the catalogue increases in size, and cannot be done at the beginning of a sequence.

These initial choices for ε and Z were motivated by investigating several catalogues (also catalogues not discussed here), because they proved effective in providing solutions in the crossover region or its proximity. Regarding the earthquake density definition, changing the sphere size V_s does not change the spatial distribution of points in Fig. 13, but only their colour. Generally, for an increasing sphere radius, the earthquake density decreases because the volume increases faster than N_{EQ}. Thus, if density values decrease, also $Z_{\max}$ decreases. Consequently, small values of Z in the initial configuration require small values of ε to find solutions in the crossover region, which however are associated with the undesired feature of a big amount of noise points. In addition, uncertainties of hypocentral depths are typically of the order of 1 km, so that smaller values of V_s might not be useful. Regarding the initial choice for ε, we set the radius of the spherical neighbourhood to 1 km because earthquakes usually occur at a depth of about 0–15 km. If ε is larger than 1 km, DBSCAN likely returns solutions with vertical extensions of the clusters spanning the entire depth range (see Fig. 9), not allowing to distinguish shallow and deep structures. Note that larger ε require larger Z to avoid solutions with only one or two huge clusters but instead remain in the crossover region. Besides, larger ε and Z result in more convex cluster shapes.

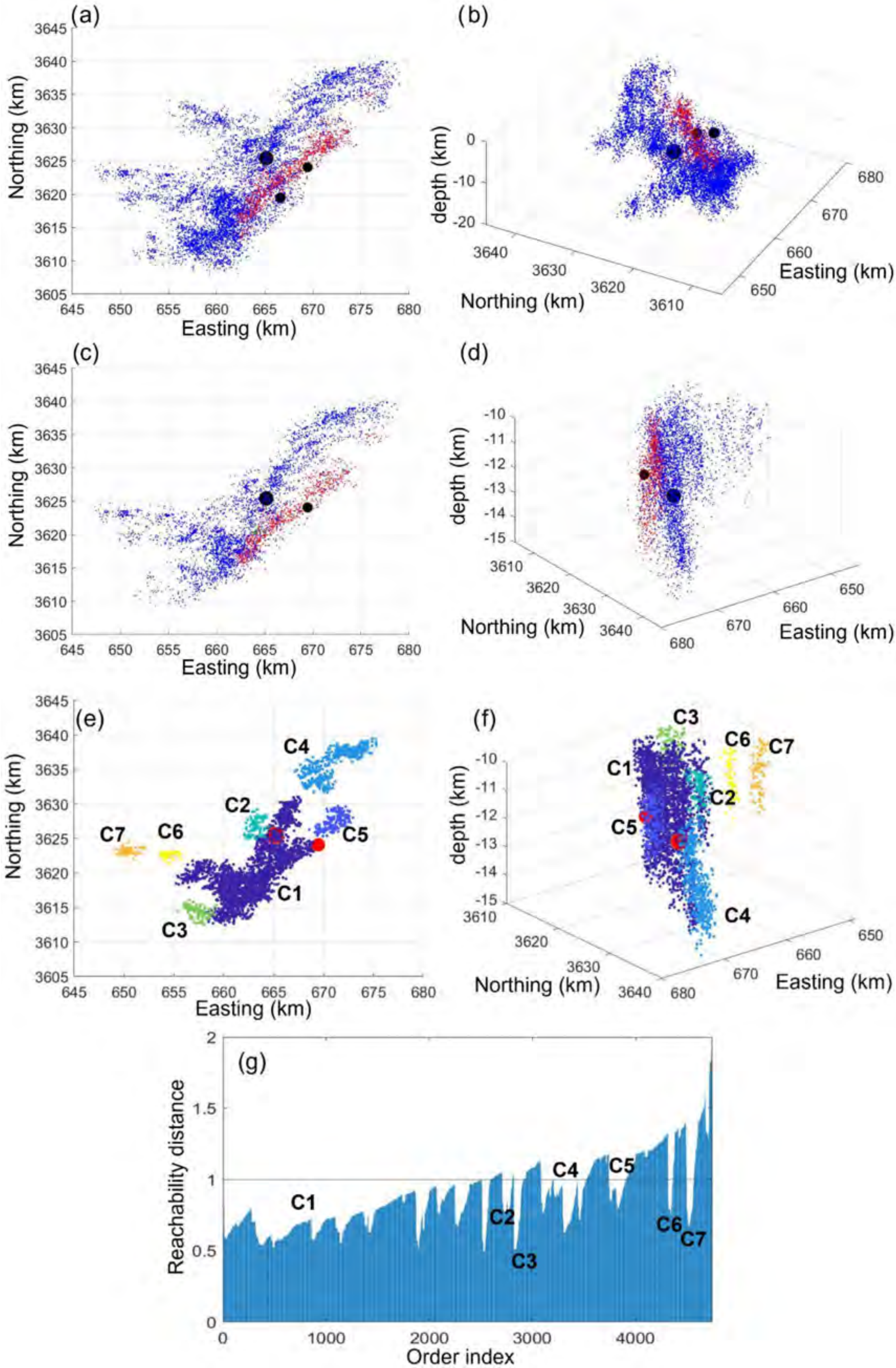


Figure 15. Performing cluster analysis for the largest cluster of the 2016 Kumamoto sequence shown in Fig. 9(b), totalling 9414 events. (a) Map view and (b) 3-D view. (c) Map view and (d) 3-D view using data in the depth range 10–15 km, totalling 4742 events. Events occurring between the two largest events (*M*6.5 and *M*7.3) are shown in red, the rest in blue (16 April 2016 to 31 August 2016). Black markers represent the three largest events of the sequence. (e–g) Cluster analysis for $Z = 25$ applied to the depth-constrained subset shown in panels (c) and (d). The map view and 3-D view in panels (e) and (f) relate to a DBSCAN solution using $\varepsilon = 1$ km, which is indicated by a horizontal line in the reachability plot in panel (g). Noise points are not shown.

Iteratively, the test condition of the cluster solution belonging to the crossover region is checked by computing the number of noise points and the size of the biggest cluster C_b. If this condition is not satisfied, comparing C_b to the size of the other clusters defines how to change the Z value: if C_b is much larger than the sum of sizes of other clusters, it must be decreased, otherwise decreased. Once a solution in the crossover region has been obtained, the reachability plot is computed for this Z with the OPTICS algorithm. Given this visualization of the nested hierarchical structure, the ε value is determined by the desired hierarchy level. This ε completes the parameter set for DBSCAN to obtain the final cluster solution.

6.2 Application to a real case

The proposed recipe (Fig. 14) is applied to the largest cluster of the Kumamoto sequence obtained from the cluster solution shown in Fig. 9(b). Since this group of earthquakes contains the three largest earthquakes of the sequence, we want to investigate if they may belong to different partitions. Fig. 15 emphasizes two periods of the data set: between the two largest events, a M6.5 and a M7.3 (14 and 15 April 2016, show in red) and everything after (16 April to 31 August 2016, show in blue). The earlier events represent a well-known, preferred alignment (Yano & Matsubara 2017), which also persists in the depth range of 10–15 km (Figs 15c and d), and are characterized by a spatial distribution that resembles a branched structure. The two largest events initiated at similar depths and belong to two different branches. From our cluster analysis, we find that both events belong to the same cluster. By applying our proposed procedure only using hypocentres in the depth range of 10–15 km, we again find that the two largest events belong to the same cluster (see Figs 15e and f), supporting the findings of previous studies (Sugito *et al.* 2016; Yue *et al.* 2017). The reachability plot nicely reflects the hierarchy of the data set and its characteristic structures (see Fig. 15g). In particular, a horizontal cut at $\varepsilon = 1$ km crosses seven valleys corresponding to the seven clusters shown in Figs 15(e) and (f) as retrieved by DBSCAN. From the reachability plot, we can infer the density and size of each cluster and already presume what happens when we change the ε threshold: a small increase in ε will cause C2 and C3 to be included in C1, whereas a decrease in ε leads to a splitting of C1 into smaller clusters due to several smaller valleys contained in it.

7 CONCLUSIONS

We performed 3-D spatial cluster analyses of seismic sequences by applying the popular density-based clustering algorithms DBSCAN in combination with the reachability plot of the OPTICS algorithm to synthetic and real hypocentre catalogues. Our analyses address the influence of the input parameters on cluster solutions and provide suggestions for exploring earthquake catalogues more appropriately.

Several studies that applied DBSCAN to earthquake catalogues using hypocentre locations, occurrence times, and/or focal mechanisms all remain vague about the choice of input parameters. Here we showed that such choices are crucial to discover regions of interest for a subsequent analysis and to identify meaningful tectonic structures that were activated in a seismic sequence.

We showed that varying the DBSCAN parameters leads to a variety of cluster solutions that can be classified into five different regions of the phase diagram. Cluster solutions inside the so-called crossover region are the most representative candidates for characterizing 3-D spatial features of seismic sequences, because they represent the individual structures as large clusters. To identify these solutions, we proposed a tentative recipe that includes a density representation of earthquakes and investigating the nested clustering structure.

We draw the following conclusions from our analyses: (i) using DB algorithms for cluster analysis requires utmost care in the selection of input parameters and the type to which the considered solution belongs to; (ii) graphically representing the spatial distribution of hypocentres and their density helps to select the input parameters and (iii) only cluster solutions in the crossover region represent information about the largest characteristic structures of a data set. Investigating such solutions can provide insight into the main features of a seismic sequence (e.g. its 3-D fault geometry) and open new perspectives for studying the spatiotemporal evolution of fault systems.

ACKNOWLEDGMENTS

We thank editor Andrew Valentine, reviewer Simone Cesca and an anonymous reviewer for their careful reading and valuable feedback, which helped us to considerably improve the quality of the manuscript. This study was partially supported by the ‘Real-time Earthquake Risk Reduction for a Resilient Europe’ (RISE) project, funded by the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. 821115).

DATA AVAILABILITY

The synthetic data for this paper are available by contacting the corresponding author at ’ester.piegari@unina.it’.

REFERENCES

Abdideh, M. & Ameri, A., 2020. Cluster analysis of petrophysical and geological parameters for separating the electrofacies of a gas carbonate reservoir sequence, *Nat. Resour. Res.,* **29,** 1843–1856.

Aggarwal, C.C. & Reddy, C.K., 2013. *Data Clustering: Algorithms and Applications,* 1st edn, Chapman & Hall/CRC.

Ankerst, M., Breunig, M.M., Kriegel, H.P. & Sander, J., 1999. OPTICS: ordering points to identify the clustering structure, *ACM SIGMOD Record,* **28**(2), 49–60.

Ansari, A., Noorzad, A. & Zafarani, H., 2009. Clustering analysis of the seismic catalog of Iran, *Comput. Geosci.,* **35,** 475–486.

Bhattacharya, S., 2021. *A Primer on Machine Learning in Subsurface Geosciences,* 1st edn., Vol., **1,** pp. 1–172, Springer.

Cesca, S., Sen, A.T. & Dahm, T., 2014. Seismicity monitoring by cluster analysis of moment tensors, *Geophys J. Int.,* **196,** 1813–1826.

Cesca, S., 2020. Seiscloud, a tool for density-based seismicity clustering and visualization, *J. Seismol.,* **24,** 443–457.

Ester, M., Kriegel, H. P., Sander, J. & Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise, *KDD-96 Proc.,* **34,** 226–231.

Fan, Z. & Xu, X., 2019. Application and visualization of typical clustering algorithms in seismic data analysis, *Proc. Comp. Sci.,* **151,** 171–178.

Herrmann, M., Piegari, E. & Marzocchi, W., 2021. b-Value of what? Complex behavior of the magnitude distribution during and within the 2016-2017 central Italy sequence, *Nature Communications,* (accepted, preprint). DOI:10.21203/rs.3.rs-1210699/v1.

Jain, A.K., Murty, M.N. & Flynn, P.J., 1999. Data clustering: a review, *ACM Comput. Surv.,* **31,** 264–323.

Kamer, Y., Ouillon, G. & Sornette, D., 2020. Fault network reconstruction using agglomerative clustering: applications to southern Californian seismicity, *Nat. Hazards Earth Syst. Sci.,* **20,** 3611–3625.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Ali Babaie, H. & Kumar, V., 2019. Machine learning for the geosciences: challenges and opportunities, *IEEE Trans. Knowled. Data Eng.,* **31,** 1544.

Kaufman, L. & Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis,* Wiley Series in Probability and Statistics, Wiley.

Konstantaras, A.J., Katsifarakis, E., Maravelakis, E., Skounakis, E., Kokkinos, E. & Karapidakis, E., 2012. Intelligent spatial-clustering of seismicity in the vicinity of the hellenic seismic arc, *Earth Sci. Res.,* **1,** 2, E-ISSN 1927-0550.

Lindsey, C.R., Neupaneb, G., Spycher, N., Fairley, J.P., Dobson, P., Wood, T., McLing, T. & Conrad, M., 2018. Cluster analysis as a tool for evaluating the exploration potential of known geothermal resource areas, *Geothermics,* **72,** 358–370.

Lyra, G.B., Oliveira-Júnior, J.F. & Zeri, M., 2014. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil, *Int. J. Climatol.,* **34,** 3546–3558.

Mehta, P., Bukov, M., Wang, C.H., Day, A.G.R., Richardson, C., Fisher, C.K. & Schwab, D.J., 2019. A high-bias, low-variance introduction to Machine Learning for physicists, *Phys. Rep.,* **810,** 1–124.

Ouillon, G., Ducorbier, C. & Sornette, D., 2008. Automatic reconstruction of fault networks from seismicity catalogs: three-dimensional optimal anisotropic dynamic clustering, *J. geophys. Res.,* **113**(B1), doi:10.1029/2007JB005032.

Ouillon, G. & Sournette, D., 2011. Segmentation of fault networks determined from spatial clustering of earthquakes, *J. geophys. Res.,* **116**(B2), doi:10.1029/2010JB007752.

Petersen, G.M., Niemz, P., Cesca, S., Mouslopoulou, V. & Bocchini, G.M., 2021. Clusty, the waveform-based network similarity clustering toolbox: concept and application to image complex faulting offshore Zakynthos (Greece), *Geophys. J. Int.,* **224,** 2044–2059.

Schoenball, M. & Ellsworth, W.L., 2017. A systematic assessment of the spatiotemporal evolution of fault activation through induced seismicity in Oklahoma and Southern Kansas, *J. geophys. Res.,* **122,** 10 189–10 206.

Sugito, N., Goto, H., Kumahara, Y., Tsutsumi, H., Nakata, T., Kagohara, K., Matsuta, N. & Yoshida, H., 2016. Surface fault ruptures associated with the 14 April foreshock (Mj 6.5) of the 2016 Kumamoto earthquake sequence, southwest Japan, *Earth, Planet Space,* **68,** 170.

Yano, T.E. & Matsubara, M., 2017. Effect of newly refined hypocenter locations on the seismic activity recorded during the 2016 Kumamoto Earthquake sequence, *Earth, Planets Space,* **69,** 74.

Yue, H., Ross, Z.E., Liang, C., Michel, S., Fattahi, H., Fielding, E., Moore, A., Liu, Z. & Jia, B., 2017. The 2016 Kumamoto Mw = 7.0 earthquake: a significant event in a fault–volcano system, *J. geophys. Res.,* **122,** 9166–9183.

Zhang, W., Zhang, Y., Gu, X., Wu, C. & Han, L., 2022. *Application of Soft Computing, Machine Learning, Deep Learning and Optimizations in Geoengineering and Geoscience,* 1st edn, Vol., **1,** pp. 1–138, Springer.

Sander, J., Qin, X., Lu, Z., Niu, N. & Kovarsky, A., 2003. *Automatic extraction of clusters from hierarchical clustering representations, in Pacific-Asia Conference on Knowledge Discovery and Data Mining,* Springer, pp.75–87.

An Energy-Dependent Earthquake Moment–Frequency Distribution

Ilaria Spassiani*[1] and Warner Marzocchi[2]

ABSTRACT

The magnitude–frequency distribution (MFD) of many earthquake catalogs is well described by the Gutenberg–Richter (GR) law or its tapered version (TGR). This distribution is usually extrapolated to any subsets of the space–time window covered by the catalog. However, some empirical observations and logical thoughts may raise doubts about the validity of this extrapolation. For example, according to the elastic rebound theory, we may assert that the probability of a strong shock nucleating within a short-time interval in a small area $\mathcal{A}$ just ruptured by another strong event should be lower than that expected by GR (or TGR): a lot of energy has already been released, and it takes time to recover to the previous state. Here, we put forward a space–time modification of the TGR, named energy-dependent TGR (TGRE) in which the corner seismic moment becomes a time-varying energy function depending on (1) the conceivable strongest shock that may nucleate in $\mathcal{A}$; (2) the time elapsed since the last strong earthquake that reset the elastic energy in $\mathcal{A}$ to a residual value; and (3) the rate of the energy recovery, linked to the recurrence time of the fault(s) involved. The model also verifies an invariance condition: for large space–time windows, the occurrence of a strong shock does not affect significantly the whole elastic energy available, that is, the TGRE becomes the TGR. The model is simple and rooted in clearly stated assumptions. To evaluate its reliability and applicability, we apply it to the 1992 Landers sequence. As expected by TGRE, we find that the MFD close to the fault system interested by the mainshock (M_w 7.3) differs from that of earthquakes off-fault, showing a lower corner magnitude. We speculate that TGRE may be profitably used in operational earthquake forecasting and that it explains the empirical observation that the strongest aftershocks nucleate always outside the mainshock fault.

KEY POINTS

- The moment–frequency distribution at small scales should account for residual elastic energy available.
- We propose the energy-dependent tapered Gutenberg–Richter (TGRE) model for seismicity in the short term.
- Epidemic-type aftershock sequence models with TGRE can improve short-term operational earthquake forecasting.

INTRODUCTION

The Gutenberg–Richter (GR) law (Gutenberg and Richter, 1944) and its tapered version (TGR; Kagan, 2002a,b) are the most used magnitude–frequency distributions (MFDs) at quite different space–time windows, such as in operational earthquake forecasting (OEF) models (Jordan *et al.*, 2011; Marzocchi *et al.*, 2017; Omi *et al.*, 2018; Michael *et al.*, 2019). The validity of the (T)GR rests on the assumption that the magnitude of an earthquake is independent of the past seismicity for any dimension of the space–time window. Although this assumption seems appropriate when looking at large spatio-temporal domains, its validity at small space–time scales conflicts with some empirical findings in which the largest triggered events occur outside the fault of the strong triggering earthquake (van der Elst and Shaw, 2015; Stallone and Marzocchi, 2019).

Conceptually, this empirical observation could be explained in the framework of the elastic rebound theory (Reid, 1911), in which one strong earthquake decreases significantly the elastic energy available in the fault that generates the shock, and it takes time to recover it. This means that the probability of a strong shock to nucleate in the same area where another strong earthquake just occurred within a short-time window has to be lower than that predicted by the (T)GR law. Conversely, if we consider a larger spatial scale, the occurrence of a single shock does not affect significantly the elastic energy available in the

1. Istituto Nazionale di Geofisica e Vulcanologia (INGV), Rome, Italy; 2. Department of Earth, Environmental, and Resources Sciences, The University of Naples Federico II, Naples, Italy

*Corresponding author: ilaria.spassiani@ingv.it

Cite this article as Spassiani, I., and W. Marzocchi (2021). An Energy-Dependent Earthquake Moment–Frequency Distribution, *Bull. Seismol. Soc. Am.* **111**, 762–774, doi: 10.1785/012020190

area, so it is expected that the (T)GR holds. In addition to the empirical evidence, we notice that the existence of a possible variability of MFD stems from recent OEF models (Field, Jordan, *et al.*, 2017; Field, Milner, *et al.*, 2017) based on faults system that can produce reliable forecasts only when the MFD is changed in space.

In this article, we put forward a space–time-dependent model that describes the MFD of earthquakes that nucleate in small space–time areas, taking into consideration the elastic energy released by the past seismicity in that area. The use of a small space–time dimension marks a difference from very recent studies on a similar argument (Marsan and Tan, 2020) and from past analyses on the definition of the maximum magnitude expected in fixed (long) time windows (Zöller *et al.*, 2013). The model introduces a time-varying corner seismic moment in the TGR law that results from the level of elastic energy that is currently available to be released in the space–time area of interest. We name the model energy-dependent TGR (TGRE) to explicitly reflect the dependence of the MFD on the elastic energy available. In a nutshell, TGRE inhibits the nucleation of large earthquakes in the area that just experienced a significant release of elastic energy.

An alternative approach to modeling the space–time variability of the elastic energy available is based on quantifying space–time variations of the b-value parameter in the GR law (Gulia and Wiemer, 2019). For instance, a larger b-value diminishes the probability of large earthquakes, but they still remain possible (e.g., if we keep fixed the rate of M 4+ earthquakes, increasing the b-value from 1.0 to 1.2 diminishes the M 7+ rate by a factor of about 4). Empirical evidence seems to show that this chance may be lower because large aftershocks nucleate almost exclusively in the outer regions of the mainshock zone (van der Elst and Shaw, 2015). The model that we put forward in this study is likely more suitable to explain such empirical evidence.

In the first part of this article, we describe the theoretical aspects of the model: we explicitly derive its formulation and that of the time-varying corner seismic moment with respect to which it is conditioned; we also discuss the stability conditions in comparison with that of the classical GR model. In the second part, we analyze the Landers earthquake sequence, which started on 28 June 1992 with an M_w 7.3 event, with a dual purpose: (1) to find empirical evidence corroborating the existence of space–time variability of the MFD and (2) to test if the proposed TGRE model better describes the data than the space–time-independent TGR model.

THE ENERGY-DEPENDENT MFD MODEL—TGRE

For the sake of mathematical simplicity, the TGRE is built in terms of seismic moment instead of magnitude; the transition from one to the other can be easily made by applying the relationship of Kanamori (1977) $m = \frac{2}{3}\log M - 10.73$, in which M stands for seismic moment (in dyn·cm) and m for the corresponding moment magnitude. Such a notation will be adopted in this article hereafter; furthermore, owing to this unambiguous relationship, we will use the acronym MFD for the seismic moment–frequency distribution. The MFD TGR Pareto law introduced by Kagan (2002a,b) reads as follows:

$$\Phi_{\mathrm{TGR}}(M) = \Phi_{\mathrm{GR}}(M) \exp\left\{\frac{M_{\min} - M}{M_{\mathrm{c}}}\right\}, \qquad (1)$$

in which $\Phi_{\mathrm{GR}}(M) = (\frac{M}{M_{\min}})^{-\beta_k}$ is the GR distribution, $\beta_k = \frac{2}{3}b$-value, $M_{\min}$ is the completeness threshold, and M_{c} is the corner seismic moment, which is the value such that events with a higher seismic moment are less likely than what is expected by the decreasing exponential distribution. The tail of the GR law is therefore forced to decay stronger in the TGR model, the decay itself being controlled by the M_{c} value, which is assumed as a fixed parameter and is typically estimated through the maximum-likelihood technique (Kagan and Schoenberg, 2001).

In this article, we introduce the TGRE model for earthquakes that nucleate inside an arbitrary portion $\mathcal{A}$ of the fault (the generalization to a volume is straightforward). The TGRE model relaxes the hypothesis that M_{c} is a fixed parameter, allowing it to vary as a function of the amount of energy E currently available in $\mathcal{A}$, that is, $M_{\mathrm{c}} \equiv M_{\mathrm{c}}(E, t)$, in which t is the time since the last earthquake that reset the energy in $\mathcal{A}$ to a residual value. This function $M_{\mathrm{c}}(E, t)$ has to consider the past earthquakes that nucleated in $\mathcal{A}$, as well as the earthquakes that involved $\mathcal{A}$ in their rupture nucleated somewhere else (we use the term "participation" hereafter as in Parsons *et al.*, 2018). In this way, the TGRE model inhibits a second strong shock to nucleate in a small area that has been involved in a strong earthquake recently, but it does not prevent this area from participating in the rupture of another big event that may nucleate nearby, along the same fault(s) involved. It follows that the nucleation MFDs in two nearby small areas may be different from, but still influenced by, the reciprocal seismicity. For the sake of simplicity, hereafter, we will omit specifying the dependence on t in the notation of $M_{\mathrm{c}}(E)$.

We also constrain the model to respect a sort of "invariance condition," that is, the TGRE turns back to the classical TGR at large spatiotemporal scales. This is to respect the evidence that, at large scales, the TGR law is well validated. Of course, the specific choice of considering a time-varying corner seismic moment is not the only one possible to introduce an energy dependence in the MFD, but it is justified in terms of easy practical use and testing; any other way of including a direct dependence on the energy can be proposed, provided that the higher complexity is balanced by better reliability of the model and that it is consistent with previous pieces of evidence. In the following sections, we define both the time-varying corner seismic moment $M_{\mathrm{c}}(E)$ and the explicit distribution of the TGRE model.

Time-varying corner seismic moment $M_c(E)$

Here, we propose a formulation of $M_c(E)$ based on two main concepts. First, the relevant quantities controlling the earthquake nucleation in $\mathcal{A}$ are the strongest earthquake that can eventually nucleate in $\mathcal{A}$ and the most recent past earthquake, which resets the available energy to the residual minimum value. Specifically, the elastic energy in $\mathcal{A}$ is reset to a residual minimum value by any earthquake that nucleates inside and generates a fractured area larger than $\mathcal{A}$. At the same time, the area is reset when it participates in a large earthquake that nucleates outside but still involves $\mathcal{A}$. In other words, the resetting event must have a seismic moment $M \geq M_A$, in which M_A is the seismic moment of an earthquake with an area equal to $\mathcal{A}$. The latter's dimension, therefore, plays a direct role in the characterization of the relative resetting events. To determine if an event was involved this area, we check if at least part of $\mathcal{A}$ falls in the circle area with the earthquake epicenter. The relative diameter (as well as M_A) may be computed through any proper rupture length–moment magnitude relationship, such as in Wells and Coppersmith (1994), Papazachos *et al.* (2004), or Allen and Hayes (2017).

Second, the elastic energy available in $\mathcal{A}$ scales with time, and it is related to M_c. In elasticity theory, $E \propto \sigma^2$, in which E is the elastic energy accumulated as a consequence of the applied stress σ; because the stress rate due to plate tectonics can be considered a constant value (i.e., $\sigma \propto t$), it follows that $E \propto t^2$. The link between the elastic energy available and seismic moment is more controversial. Here, we assume that the radiated energy is a reliable proxy of the elastic energy drop and that the radiated energy is proportional to the seismic moment, $E \propto M_c$; the latter holds only if the static stress drop of earthquakes is independent of the magnitude. These hypotheses are still matters of intense debate (Ide and Beroza, 2001; Kanamori and Brodksy, 2004; Oth *et al.*, 2010), and their validity is model dependent (Kanamori and Brodksy, 2004). However, we stress that a TGRE may be built adopting a different form of $M_c(E)$ that takes into account different hypotheses.

Going into the detail, we define the following parameters:

1. M_c^* is the maximum corner seismic moment for earthquakes nucleating in $\mathcal{A}$. It is actually the corner seismic moment M_c adopted in the classical TGR distribution (1). We propose that M_c^* could be related, although not necessarily, to the length of the longest fault included in the area: for instance, it can be obtained from any proper rupture length–moment magnitude relationship, such as those proposed in Wells and Coppersmith (1994), Papazachos *et al.* (2004), or Allen and Hayes (2017).
2. t_0 is the occurrence time of the earthquake that reset the elastic energy in $\mathcal{A}$, that is, the past earthquake in which $\mathcal{A}$ participated.
3. $M_{c,0}^*$ sets the minimum value for the corner seismic moment that is achieved after the occurrence of a resetting earthquake in $\mathcal{A}$. In general, $M_{c,0}^* = \rho M_c^*$, in which $\rho < 1$ indicates the fraction of elastic energy that is available after the resetting event. The value of ρ, or equivalently of $M_{c,0}^*$, may be set either theoretically, for instance, by analyzing the stress rotation (Hardebeck and Okada, 2018), or empirically by analyzing one or more stacked similar earthquake sequences.
4. ν is a parameter connected to the recurrence time of the longest fault involved in $\mathcal{A}$, and it controls the velocity of convergence to the maximum value M_c^* after a resetting event.

In the Setting Parameters and Assumptions section, we describe some practical choices for these parameters. Still, we stress again that the choices are not prescriptive for the TGRE's application; different $M_c(E)$ parameterizations, assumptions, and parameters can be used.

According to the previous concepts and definitions, we define the time-varying energy function as

$$M_c(E) = M_{c,0}^* + (M_c^* - M_{c,0}^*)[\nu(t - t_0)]^\alpha, \tag{2}$$

bounded to the values $(t - t_0) \leq \frac{1}{\nu}$, which translates to $(t - t_0) \leq \tau$ when the coefficient of variation (COV) of the interevent times between consecutive earthquakes is zero, that is, τ is the recurrence time between earthquakes. This restriction guarantees that $M_c(E) \in [M_{c,0}^*, M_c^*)$ when $(t - t_0) \in [0, \tau]$, a requirement that is deducible from the earlier argument. The dependence of the corner seismic moment on time is therefore expressed with respect to the time elapsed since the resetting event, and the seismic moments multiplication term allows us to account for the energy reloading process; whereas, $M_{c,0}^*$ is added to ensure that the available energy will not fall below its minimum value, even immediately after the resetting event, that is, when $t - t_0 \sim 0$. In this article, according to the proportionality between elastic energy and seismic moment, we set $\alpha = 2$ ($\alpha = 1$ if the seismic moment is assumed to increase linearly with time).

The temporal trend of $M_c(E)$, as well as its sensitivity to the parameters, can be observed in Figure 1, the plots of which are obtained by considering two parameters among $(M_c^*, M_{c,0}^*, \nu)$ fixed and the third varying; for an easier interpretation, we also display magnitude values instead of seismic moments. An overall increasing trend is shown in all of the plots. As intuition suggests, the time-varying corner seismic moment approaches its maximum more rapidly when ν becomes larger: the lower the recurrence time of the fault is, the faster theM_c^* is reached. The increasing velocity of $M_c(E)$ is also faster as M_c^* is higher, whereas it does not change with $M_{c,0}^*$. This is because the influence of the latter on the taper's trend can be appreciated only within a short-time interval since the resetting event (less than 1 yr in our example), with $M_c(E)$ controlled mainly by M_c^* and ν at just larger scales: this is why the x-axes in plot (c) are cut at 1 yr after the reset; otherwise, the difference would not have been visible. When focusing on the entire time window, we

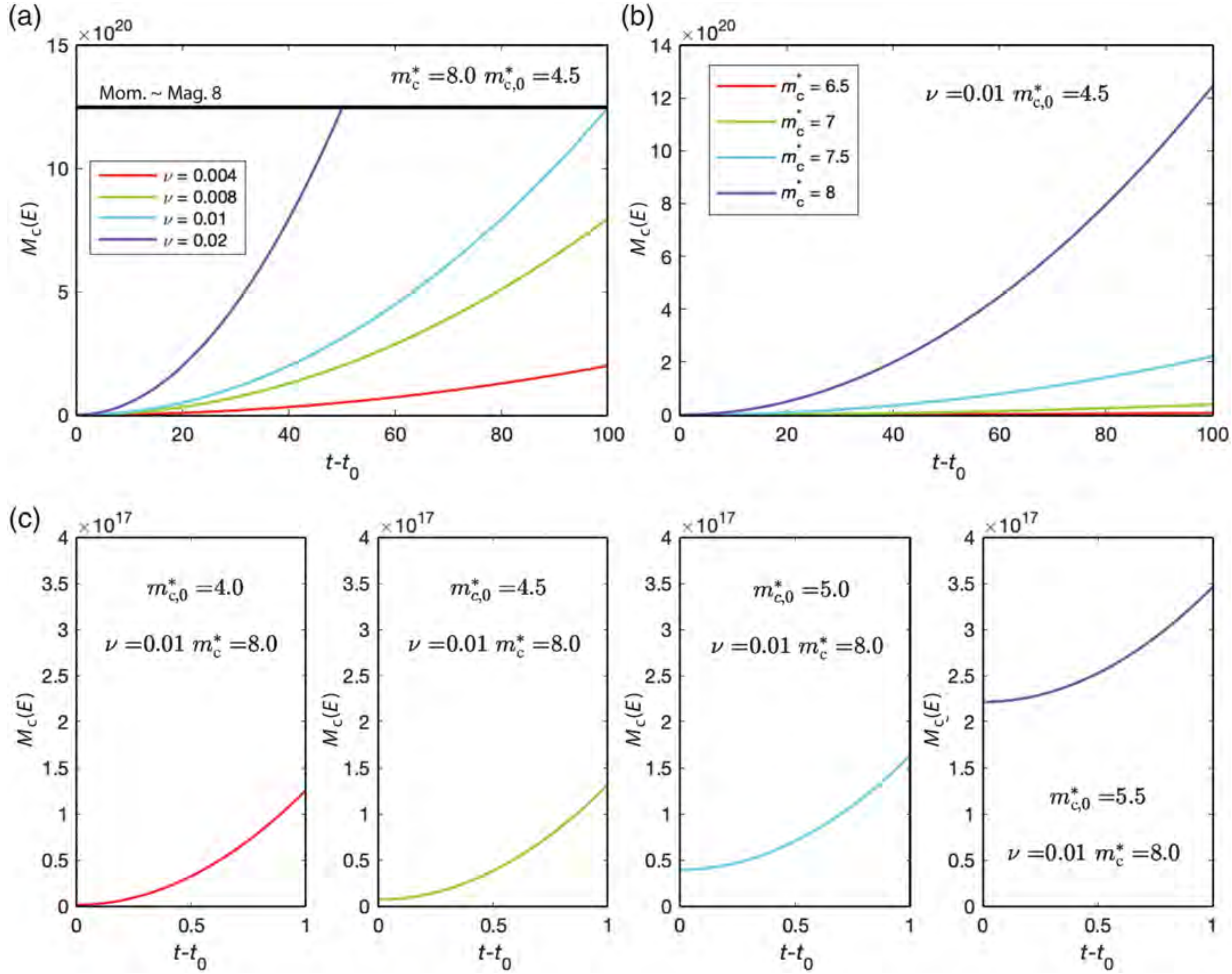


Figure 1. Time-varying corner seismic moment $M_c(E) = M_{c,0}^* + (M_c^* - M_{c,0}^*)[\nu(t - t_0)]^2$ as a function of the elapsed time $t - t_0$ between the event (t, M) and the resetting one (t_0, M_0). Panels (a–c) are obtained, respectively, for fixed $(M_c^*, M_{c,0}^*)$—varying ν, fixed $(\nu, M_{c,0}^*)$—varying M_c^*, and fixed (ν, M_c^*)—varying $M_{c,0}^*$. The latter is obtained for a shorter $t - t_0$ interval because here the differences of the corner seismic moment function can be appreciated: $M_c(E)$ would not change substantially over a longer temporal interval, being M_c^* is predominant over $M_{c,0}^*$. Magnitude values are shown in place of seismic moments for an easier interpretation of the figure.

observe that the influence of ν and M_c^* on $M_c(E)$ is a bit stronger. However, Figure 2 highlights that, in the short term, the time-varying corner seismic moment does not substantially depend on these two values.

To be thorough, we add that $M_c(E)$ could be also interpreted as a random variable with a distribution that takes its cue from the stress level adopted in the stress release model (Vere-Jones, 1978, 1988; Wang *et al.*, 1991; Zheng and Vere-Jones, 1991; Xiaogu and Vere-Jones, 1994). In fact, $M_c(E)$ could consist of a deterministic term of accumulated energy, linked to the elapsed time since the resetting event, and a stochastic term of energy released by each single past earthquake, which is distributed according to TGR. Nevertheless, to gain easy applicability and reliable testing, we assume here that $M_c(E)$ is a deterministic function of time, as defined in (2).

The mathematical description of the TGRE model

The TGRE model that we propose for earthquake seismic moments is simply obtained by including the time-varying corner seismic moment $M_c(E)$ previously derived in the TGR cumulative distribution (1), that is,

$$\Phi_{\mathrm{TGRE}}(M) = \left(\frac{M}{M_{\min}}\right)^{-\beta_k} \exp\left\{\frac{M_{\min} - M}{M_c(E)}\right\}, \qquad (3)$$

in which $M_c(E) \in [M_{c,0}^*, M_c^*]$ is defined in equation (2) with $\alpha = 2$. Figure 3 shows $\Phi_{\mathrm{TGRE}}(M)$ as a function of $M_c(E)$.

If we consider a large spatial domain composed of many faults, the occurrence of one or a few large earthquakes may reset the elastic energy only on a limited portion of the area. This means that, for the whole large spatial domain, $M_c(E) \equiv M_c^*$, which acknowledges the spatial invariance condition. The temporal invariance condition is instead satisfied by construction; in fact, equation (2) gives M_c^* for $t - t_0 \to \tau$.

One obvious application of the TGRE model is in OEF (Jordan *et al.*, 2011). It is expected to solve the main conundrum of existing OEF models (Marzocchi *et al.*, 2017; Omi *et al.*, 2018; Michael *et al.*, 2019), for which the probability of a large aftershock is exactly where the mainshock occurred. For example, the epidemic-type aftershock sequence–TGRE (ETAS-TGRE) (i.e., ETAS with TGRE) rate would be:

$$\lambda(t, x, y, M|\mathcal{H}_t) = [\lambda_0(x, y) + \sum_{\{i|t_i<t\}} \lambda_{\mathrm{tr}}(t - t_i, x - x_i, y - y_i; M_i)] p_{\mathrm{TGRE}}(M|M_c(E)), \quad (4)$$

in which $\mathcal{H}_t$ is the past history up to time t, that is, the past earthquakes $\{(t_i, x_i, y_i, M_i); t_i < t\}$; $\lambda_0(x, y)$ is the rate of the background events; $\lambda_{\mathrm{tr}}(t - t_i, x - x_i, y - y_i; M_i)$ is the rate of the triggered events; $p_{\mathrm{TGRE}}(M|M_c(E))$ is the TGRE probability density function for the seismic moment that is calculated in x, y at time t; and finally, $M_c(E) \equiv M_c(E, x, y, s)$ is linked to the elastic energy available in x, y after time s since the last resetting earthquake. In this framework, the TGRE may

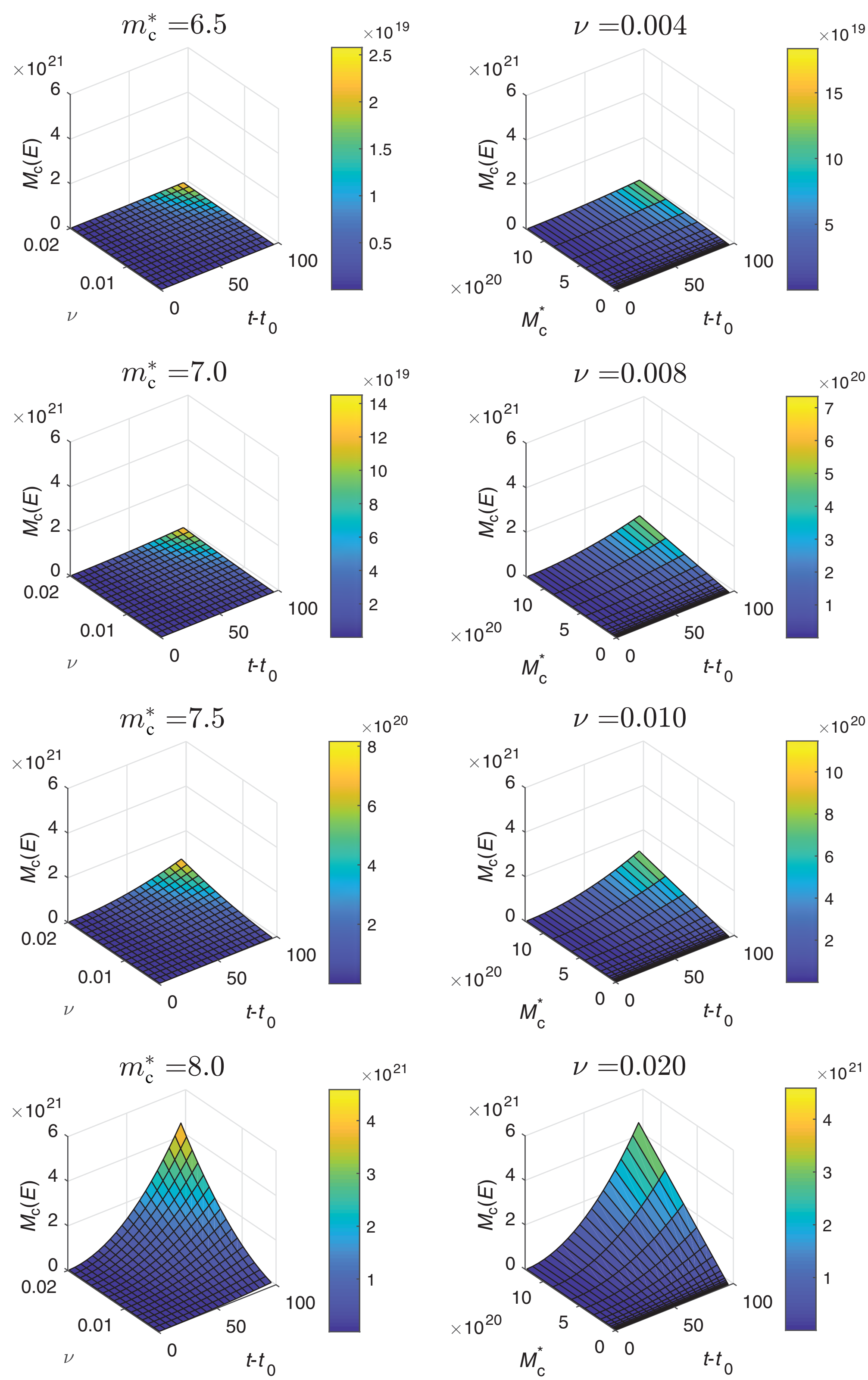


Figure 2. Surface plots of the time-varying corner seismic moment $M_c(E) = M_{c,0}^* + (M_c^* - M_{c,0}^*)[\nu(t - t_0)]^2$ as a function of the time elapsed since the reset $t - t_0$ and the parameter ν with fixed M_c^* in the first column panels, and vice versa in the second column. The minimum corner seismic moment is set at $m_{c,0}^* = 4.5$ in each panel.

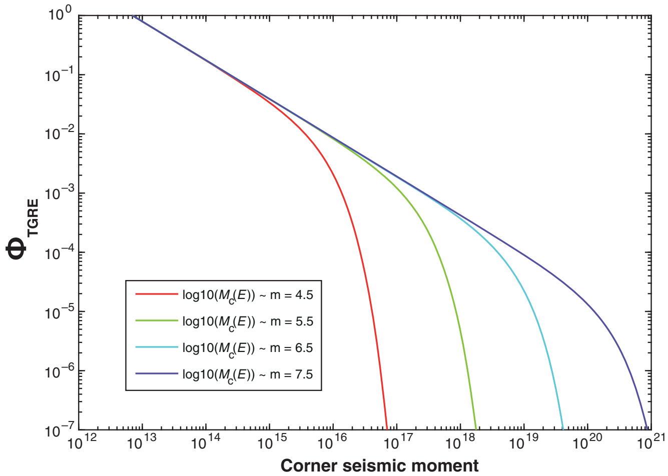


Figure 3. Survival function of the energy-dependent Tapered Gutenberg–Richter (TGRE) model for several values of the available energy (corner seismic moment), corresponding to the $M_c(E)$ indicated in the legend, in a log–log scale.

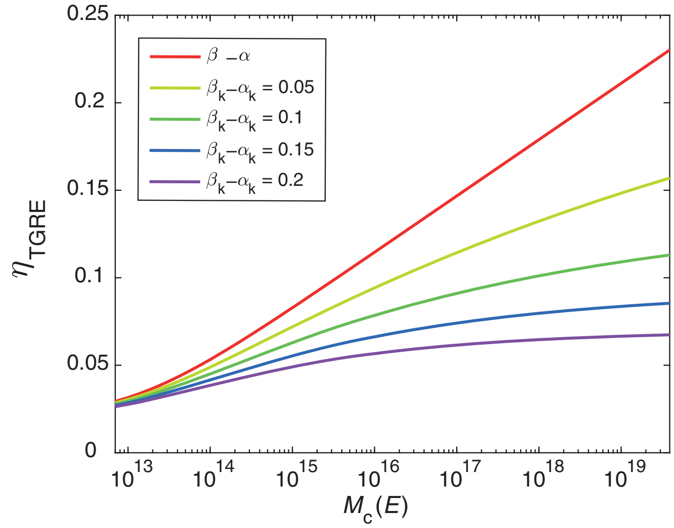


Figure 4. TGRE branching ratio (5) versus its time-varying corner seismic moment, for several values of the difference $\beta_k - \alpha_k$.

be applied to both the background and triggered earthquakes as in the classical ETAS model.

In the ETAS-TGRE setting, it is also interesting to investigate how the shift of the TGRE taper influences the computation of the branching ratio, which we recall is the average number of aftershocks triggered by an arbitrary event (Zhuang *et al.*, 2012). As for the TGR law, the branching ratio of the TGRE model is derived as

$$\eta_{\mathrm{TGRE}} = \kappa + \kappa\alpha_k e^{\frac{M_{\min}}{M_c(E)}} \left(\frac{M_{\min}}{M_c(E)}\right)^{\beta_k-\alpha_k} \Gamma\left(-\beta_k + \alpha_k, \frac{M_{\min}}{M_c(E)}\right), \quad (5)$$

in which $\Gamma(s,t) = \int_t^\infty x^{s-1} e^{-x} dx$ is the upper incomplete gamma function (Bateman, 1953; Temme, 1996; Spassiani, 2020) and κ, α_k are the parameters of the productivity law $\rho(\cdot)$ expressed in terms of the seismic moment through the relationship of Kanamori (1977), that is, $\rho(M) = \kappa(\frac{M}{M_{\min}})^{\alpha_k}$. In Figure 4, we show that η_{TGRE} increases with the time-varying corner seismic moment $M_c(E)$, indicating that, if the taper moves to the left as a consequence of a great amount of energy just released, the average number of aftershocks triggered by a generic event is reduced: in fact, an event with a lower seismic moment will generate a lower number of aftershocks. The plot shows that the increasing behavior is faster as the difference $\beta_k - \alpha_k$ is lower: in the case of the classical ETAS-GR, it has to be $\beta_k > \alpha_k$ for the process not to explode, but this condition becomes unnecessary for the ETAS-TGRE model, and so for ETAS-TGR (Spassiani, 2020). As usual, the stability of the ETAS-TGRE process is guaranteed by imposing $\eta_{\mathrm{TGRE}} < 1$. Because this branching ratio increases with the corner seismic moment, we have $\eta_{\mathrm{TGR}} < 1$ as a sufficient condition for the stability of the ETAS-TGRE process; in fact, $M_c(E)$ assumes its maximum in M_c^*. We also stress that, when $\beta_k > \alpha_k$, $\eta_{\mathrm{TGRE}} < \eta_{\mathrm{GR}}$ holds; therefore, in this case, our model's stability conditions are less restrictive than those of ETAS-GR.

APPLICATION TO REAL EARTHQUAKES: THE LANDERS SEQUENCE

In this section, we test the hypothesis of the space and time independence of MFD, and then we show how the TGRE model works in practice. To do that, we consider the Landers earthquake sequence, which started with an M_w 7.3 event that occurred on 28 June 1992, in southern California. The seismic catalog for such a sequence is rich enough to allow for a statistically significant data-model comparison. Furthermore, the fault segment that generated the initial earthquake is well defined in this case because a detailed mapping of the slip distribution is available; in our analysis, we focus on the fault segments that certainly slipped during the M_w 7.3 event, as shown in Madden and Pollard (2012) and hereafter called the "Landers fault."

Seismic data for the analysis have been taken from the online available Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) earthquake catalog, which covers the entire California region from July 1769 to April 2010 and includes events with M ≥ 4 before 1894 and M ≥ 2.5 after 1894 (Field *et al.*, 2013). The data relative to the Landers fault have been taken from the California Reference Fault Parameter Database–UCERF2 system, which is easily accessible online (Field *et al.*, 2009) and does not present substantial differences with respect to the UCERF3 regarding the faults involved in the Landers rupture. For the websites, see Data and Resources.

In particular, in this application, we test whether the MFDs inside and outside the rupture that generated the Landers earthquake come from the same distribution, which would be expected in the case of space–time independence. Then, we apply the TGRE model to the on-rupture earthquakes, and we quantify the difference in the reliability of the TGR and TGRE models through the log-likelihood ratio test. The red stripe in Figures 5a–8a shows the rupture on the Landers fault. The stripe has a thickness of about 10 km (considering ±5 km around the latitude of each segment fault point). The analysis is conducted in the following four time intervals: 29 June–6 July 1992, 29 June–29 July 1992, 29 June–29 September 1992, and 29 June 1992–29 June 1993, that is, respectively, one week, one month, three months, and 1 yr since the day after the M_w 7.3 resetting event.

Setting parameters and assumptions

The first step is to define $\mathcal{A}$, which sets the spatial resolution of the analysis. We consider the case in which $\mathcal{A}$ covers the whole fault rupture of the M_w 7.3 earthquake (red stripe in Figs. 5a–8a). For this tutorial application, we set the parameters of the TGRE model as follows:

1. M_c^* corresponds to $m_c^* = 7.59$, as proposed in Kagan *et al.* (2010) for active continents.
2. After the resetting Landers earthquake, no other resetting earthquake occurred in $\mathcal{A}$ in the time interval considered; therefore, t_0 corresponds to 28 June 1992.
3. $M_{c,0}^*$ is estimated through a grid search; specifically, we searched the $m_{c,0}^*$ in the set [4,4.1,…,6], which maximizes the likelihood ratio in favor of TGRE in the first week of data. As shown in Figure 9, we find $m_{c,0}^* = 4.3$. Of course, more sophisticated procedures to estimate $M_{c,0}^*$ are possible, but we argue that the results are stable for reasonable variations of this parameter. In particular, the log-likelihood ratio remains well above zero (TGRE explains the data better than TGR) for $4.1 \le m_{c,0}^* \le 4.8$. Then, in Table 1, we show also that the $M_{c,0}^*$ estimated in the first week of data shows the superiority of TGRE with respect to TGR also for other time windows (one month, three months, and 1 yr) (see the Results section for more details).
4. $\nu = \frac{1}{\tau(1-2\mathrm{COV})}$, in which the recurrence time $\tau = 250$ yr is rescaled to account for the covariance coefficient COV = 0.3.

The results are illustrated in the next section. To check their stability and the sensitivity of the model, in addition to using different $M_{c,0}^*$, we perform the analysis for other possible values of the parameters M_c^* and τ. The details are reported in Table 1. We anticipate that the results are not significantly modified, in agreement with what is shown in Figure 2 as previously discussed.

Results

The results are illustrated in Figures 5–8 for one week, one month, three months, and one year since the day after the resetting Landers event, respectively. The space–time windows

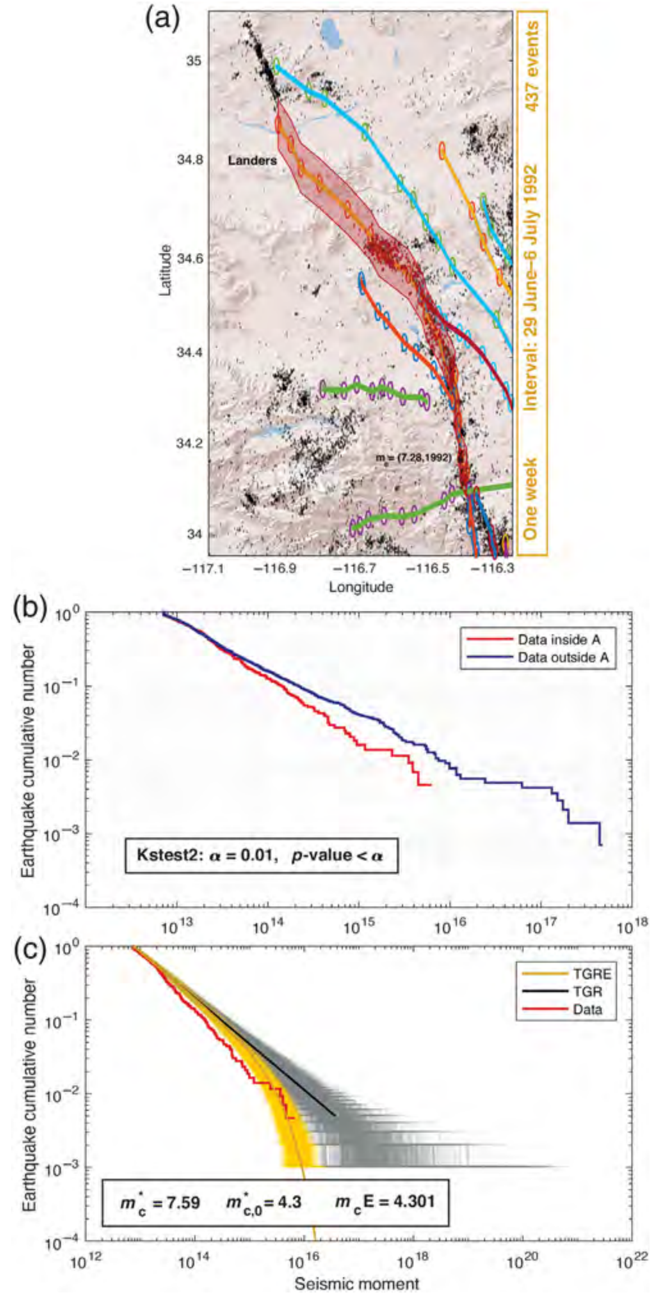


Figure 5. Tapered Gutenberg–Richter (TGR) versus TGRE analysis relative to the considered area $\mathcal{A}$, covering the Landers segment fault, as shown in panel (a) (the colored lines with circles represent the nearby segment faults). The temporal interval here is 29 June–6 July 1992, that is, within one week of the day after the Landers resetting earthquake. The number of events contained in this spatiotemporal window is 437 (red dots in panel (a)). Panel (b) contains the earthquake cumulative number of events inside $\mathcal{A}$ (in red), and outside it (in dark blue). Finally, in panel (c), we compare the fit with the data of the TGR model in black and the TGRE model in yellow, obtained, respectively, with $m_c^* = 7.59$ and $m_c(E) = 4.301$ (the latter derived from equation 2 with $\alpha = 2$). These corner magnitudes are also used to obtain 1000 simulations of 1000 TGR- and TGRE-distributed seismic moments, respectively, which are plotted as light-gray and light-yellow cones, respectively. The data (red step functions) almost completely fall into the TGRE cone. Magnitude values are shown in place of seismic moments for an easier interpretation. Kstest2, two-samples Kolmogorov–Smirnov test.

TABLE 1
Difference between Tapered Gutenberg–Richter (TGR) and Energy-Dependent TGR Log Likelihoods in Bold and $m_c(E)$ Values in Brackets

Model Parameters			One Week	One Month	Three Months	1 Yr
$m_c^* = 7.59$*	$\tau = 250$ yr ‖	$m_{c,0}^* = 4.3$	**3.16** (4.301)	**3.51** (4.32)	**2.76** (4.43)	**1.04** (4.96)
$m_c^* = 7.53$†	$\tau = 100$ yr#	$m_{c,0}^* = 4.3$	**3.15** (4.305)	**3.36** (4.38)	**1.85** (4.69)	**0.26** (5.4)
$m_c^* = 7.5$‡	$\tau = 1000$ yr**	$m_{c,0}^* = 4.3$	**3.16** (4.3)	**3.53** (4.301)	**2.65** (4.31)	**2.62** (4.4)
$m_c^* = 8.0$§	$\tau = 500$ yr††	$m_{c,0}^* = 4.3$	**3.16** (4.301)	**3.51** (4.32)	**2.75** (4.43)	**1.02** (4.96)

*$m_c^* = 7.59$ as in Kagan *et al.* (2010).
†$m_c^* = 7.53$ by relations of Wells and Coppersmith (1994).
‡$m_c^* = 7.5$ such as for the San Jacinto fault (Salisbury *et al.*, 2012).
§$m_c^* = 8$ close to that for the northern San Andreas fault.
‖$\tau = 250$ such as for the northern San Andreas fault.
#$\tau = 100$ such as for the San Jacinto fault.
**$\tau = 1000$ as in Sieh *et al.* (1993).
††$\tau = 500$ close to a mean value.

in which the analysis is performed are shown in the map of panels (a), in which the on-rupture seismicity of $\mathcal{A}$ (red dots) is reported inside the red stripe.

In Figures 5b–8b, we show the results of the null hypothesis of having the same MFD inside and outside the ruptured area. In particular, we plot the earthquake cumulative number of events inside (in red) and outside (in dark blue) $\mathcal{A}$ in different time windows. In each of the four temporal intervals, red and dark-blue step functions are clearly different, and the two-samples Kolmogorov–Smirnov test (Massey, 1951) confirms the rejection of the null hypothesis that the data are drawn from the same continuous distribution, at a significance level of 0.01 that was chosen before carrying out the analysis (the p-values on the figures are always smaller than the significance level). We stress that these results are completely independent of the modeling because they are obtained by considering only earthquake data. At the same time, these results support the main motivation of this work, that is, empirical data support the hypothesis of different MFDs on- and off-rupture just after a large shock.

Figures 5c–8c show the goodness of fit of the TGRE and TGR models with respect to the earthquake data inside $\mathcal{A}$. Specifically, we plot the TGR model in black versus the TGRE one in yellow, orange, green, and blue for one week (Fig. 5), one month (Fig. 6), three months (Fig. 7), and one year (Fig. 8) since the reset, respectively. We also show 1000 simulations of 1000 magnitudes each, obtained both with $m_c^* = 7.59$, which is drawn from a TGR (light-gray cones), and with the new corner magnitudes $m_c(E)$ obtained for the TGRE model (light-yellow, orange, green, and blue cones for the four temporal intervals considered). The results show that within one week of the Landers earthquake, the TGRE corner seismic moment is reduced to a value $\sim M_{c,0}^*$ corresponding to the minimum energy and then it increases with the energy-reloading process.

In all four cases, the TGRE model gives visually a better fit to earthquake data than the TGR model: the red step functions representing the recorded magnitudes are almost completely contained in the nongray cones, indicating our model's capability to better reproduce the time evolution of the real seismicity occurring in $\mathcal{A}$ that just experienced the strong resetting Landers event. We argue that this general observation is independent of the choice of $M_{c,0}^*$ because of the clear bending in the MFD of the earthquakes inside $\mathcal{A}$.

We explore further the suitability of TGRE for calculating the likelihood ratio for the nested TGR and TGRE models (King, 1998). The likelihood ratio is a measure of how much the TGRE is supported by the data with respect to TGR. In particular, the log-likelihood function,

$$\log L(\theta) = N\beta_k \log M_{\min} + \frac{NM_{\min} - \sum_{i=1}^{N} M_i}{\theta} + \sum_{i=1}^{N} \log\left(\frac{\beta_k}{M_i} + \frac{1}{\theta}\right) - \beta_k \sum_{i=1}^{N} \log M_{\min}, \quad (6)$$

is the same for both the models, and it represents the TGRE when $\theta = M_c(E)$ and the TGR when $\theta = M_c^*$. In Table 1, we show the difference $\log L(M_c(E)) - \log L(M_c^*)$ between the two log likelihoods, computed for the four space–time windows considered. The results in the first row of Table 1 are relative to the earthquake data used in Figures 5c–8c, that is, with $m_c^* = 7.59$, $m_{c,0}^* = 4.3$, and $\tau = 250$ yr. As anticipated in the previous section, to verify the stability of the results as a function of these parameters, we calculate the likelihood ratio also for different m_c^* and τ (see the first three columns in Table 1); for all of these cases, we found that $m_{c,0}^* = 4.3$ maximizes the likelihood ratio in favor of TGRE in the first week of data. The results of this stability test are shown in the second row and on. Borrowing the terminology adopted by Kass and Raftery (1995) for the Bayes factor, we may say that the evidence in favor of TGRE with respect to TGR is, most of the time, "substantial" and "strong." As expected, this evidence diminishes only in some cases for long temporal windows, but it still remains >0, showing a superiority of TGRE with respect to TGR independent of the parameters. In general,

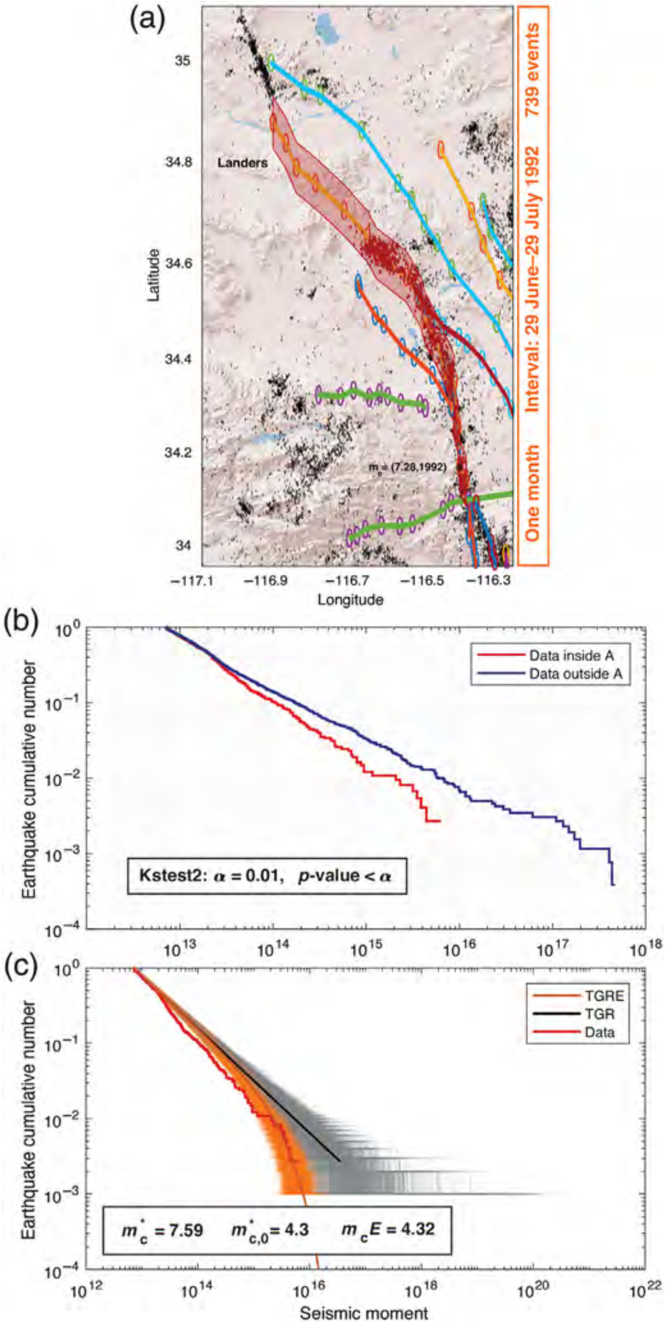


Figure 6. Same as Figure 5, but relative to the temporal interval 29 June–29 July 1992, that is, within one month of the day after the Landers resetting earthquake. The number of events contained in this spatiotemporal window is 739. The color used for the TGRE model is orange, and $m_c(E) = 4.32$.

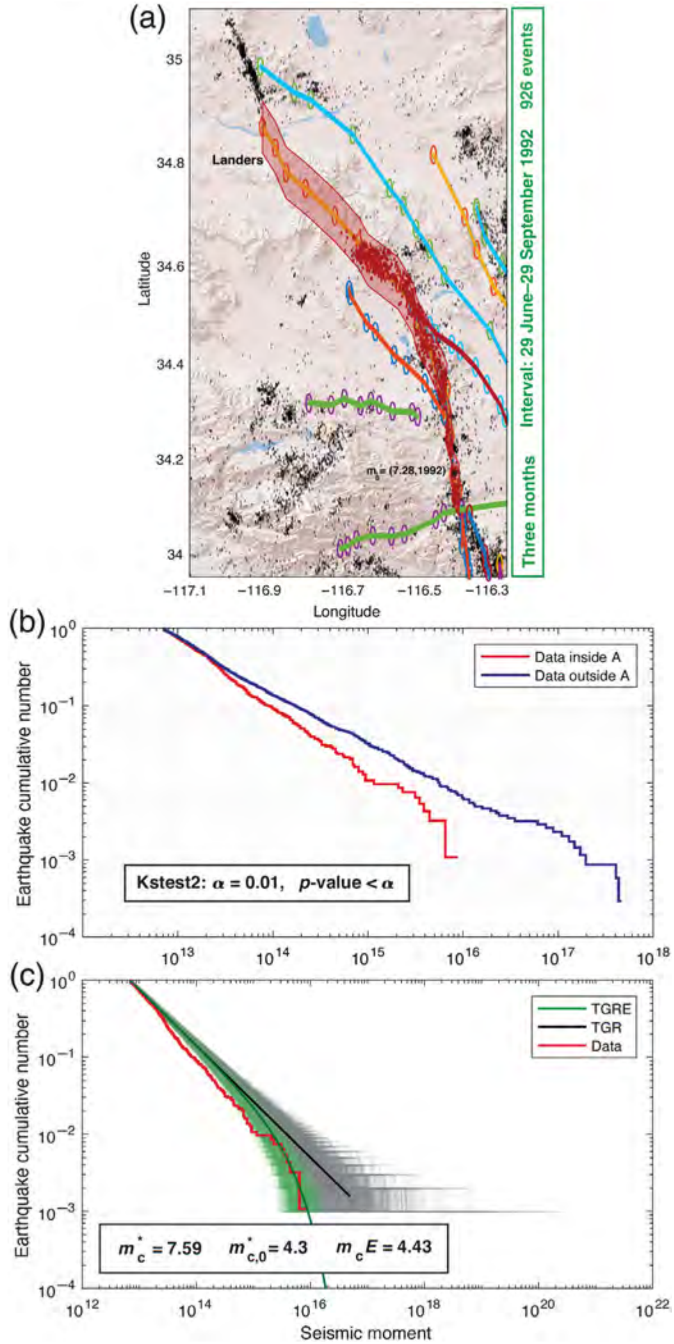


Figure 7. Same as Figure 5, but relative to the temporal interval 29 June–29 September 1992, that is, within three months of the day after the Landers resetting earthquake. The number of events contained in this spatiotemporal window is 926. The color used for the TGRE model is green, and $m_c(E) = 4.43$.

the overall first-increasing-then-decreasing trend of the log-likelihood differences when moving to longer time periods is expected, as a trade-off between the number of events and the recharging of the elastic energy of the system.

Finally, we find that the results remain stable also when considering completeness thresholds $m_{\min}$ higher than 2.5 or when removing the first few days just after the resetting M_w 7.3 event, in which $m_{\min}$ may be higher than in the following days. As a matter of fact, any problem in the completeness magnitude should have equally affected the MFD of both events inside and outside $\mathcal{A}$, leaving the difference between the two distributions unchanged.

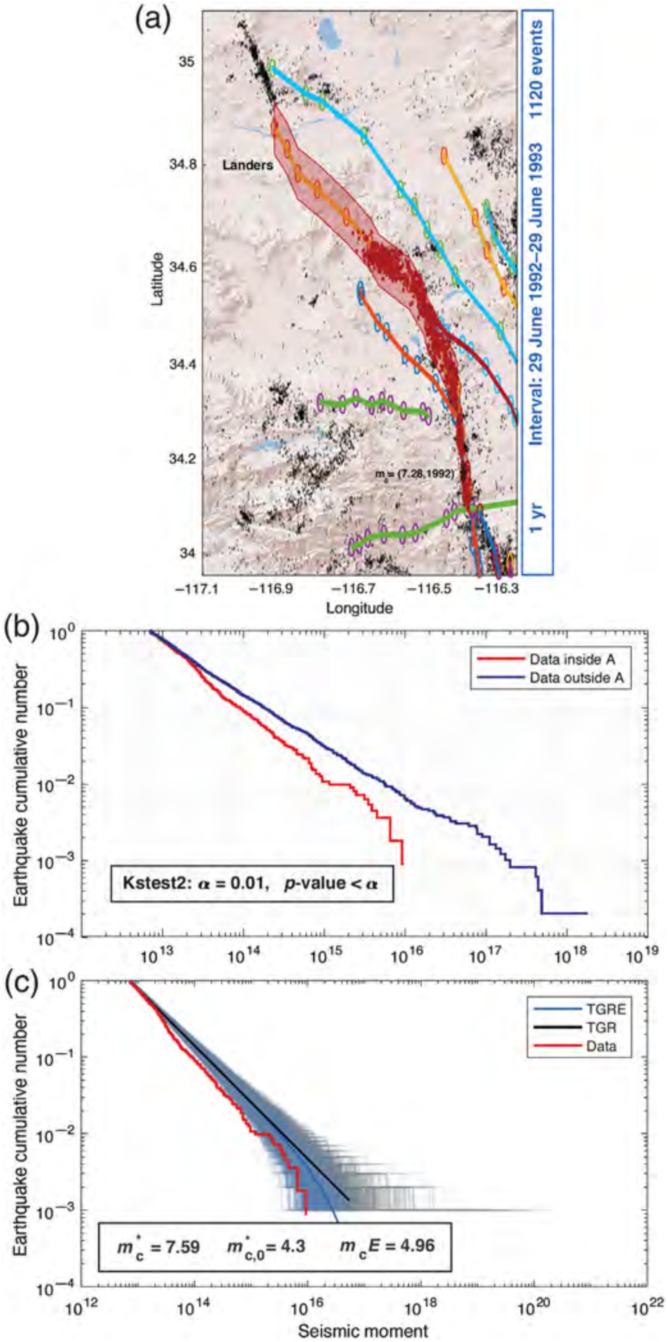


Figure 8. Same as Figure 5, but relative to the temporal interval 29 June 1992–29 June 1993, that is, within 1 yr of the day after the Landers earthquake. The number of events contained in this spatiotemporal window is 1120. The color used for the TGRE model is blue, and $m_c(E) = 4.96$.

DISCUSSION AND CONCLUSIONS

Basic physical principles and empirical evidence suggest that MFD can vary with space and time. To this purpose, in this article, we have proposed the energy-varying seismic moment–frequency model TGRE for earthquake nucleation, which depends on the elastic energy currently available in an area $\mathcal{A}$ of interest. This model acknowledges the elastic rebound

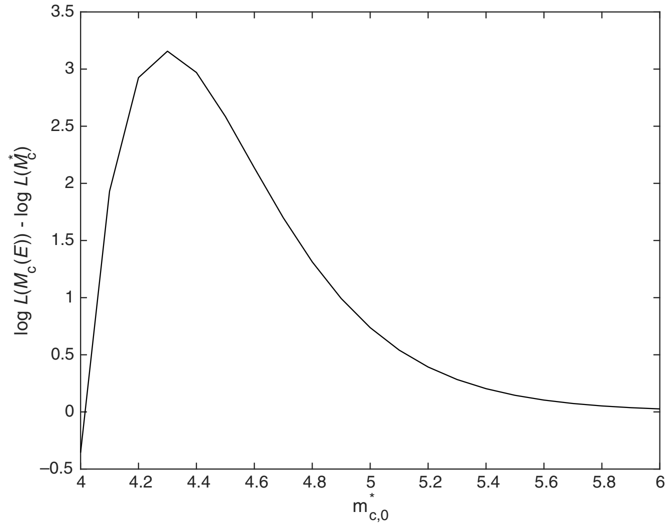


Figure 9. Difference between TGRE and TGR log likelihoods versus the minimum magnitude $m^*_{c,0}$ achieved after the reset.

theory and justifies the observation that the largest triggered earthquakes nucleate always outside the fault section that has just generated a large shock. In a different perspective, the model may also describe quantitatively an intermittent criticality state that is tuned by the available elastic energy. In other words, the state of self-organized criticality—advocated to explain the power law distribution of the seismic moments at large spatiotemporal scales (Bak and Tang, 1989; Sornette and Sornette, 1989)—changes in intermittent criticality when zooming on small space–time windows that have been recently involved in a large earthquake, indicating that a fault system approaches and retreats from a critical state by turns (Ben-Zion *et al.*, 2003; Bowman and Sammis, 2004; Bebbington *et al.*, 2010).

The TGRE distribution is obtained as a modification of Kagan's TGR law in which the corner seismic moment is a time-varying energy function, that is, it is linked to the proxy of the amount of energy available in $\mathcal{A}$. The TGRE model is conceptually simple, and it depends on a few parameters: (1) the corner seismic moment M^*_c, which is loosely related to the strongest event that may nucleate in $\mathcal{A}$; (2) the temporal occurrence of the last large earthquake that reset the elastic energy in $\mathcal{A}$ to a residual value; (3) the rate of the energy recovery, which depends on the recurrence time of the fault(s) involved; and (4) $M^*_{c,0}$, which is the minimum value for the corner seismic moment that is achieved after the occurrence of a resetting earthquake in $\mathcal{A}$. In other words, the TGRE right tail $M_c(E)$ abruptly moves to $M^*_{c,0}$ just after the occurrence of a strong (resetting) event, and then it slowly recovers to the long-term value; in practice, the model inhibits the nucleation of a large triggered earthquake in segments that recently experienced a large shock. An interesting feature of TGRE is that it verifies an invariance condition: because the dimension of the selected space–time window becomes larger, it converges to the TGR law with a limiting corner seismic moment M_c^*.

The TGRE has been designed purposely simple (depending on a few clear physical parameters), acknowledging that understandability (and usability) is inversely proportional to the complexity of a model. Similar to other models, it contains (more or less explicit) subjective choices, but we think that these choices are less subjective than ignoring the empirical evidence that strong triggered earthquakes do not nucleate on a fault that has just ruptured by another strong event, as is assumed in the (T)GR model. This empirical evidence can be hardly explained by space–time variability of the b-value of the GR law, which would lower, not inhibit, the triggering of large earthquakes on a fault that has just slipped.

Despite its simplicity, we have shown that TGRE may explain well the statistically significant difference in the MFDs relative to on- and off-rupture seismicity for the Landers sequence and that the results are stable for possible variations of the parameters. In particular, TGRE outperforms TGR for different values of $m_{c,0}^*$, showing the strongest difference for $m_{c,0}^* = 4.3$. Further studies will be necessary to reduce uncertainties in this value. For now, we notice that the results seem to indicate allowing the corner seismic moment to vary in space and time is more important than the details about the model's parameters choice. That said, we underline that the TGRE reliability (similar to other models) and the comparison with alternative models (e.g., models based on space–time variations of the b-value) have to be evaluated through prospective tests. For this model, prospective tests will be carried out in the framework of the ongoing European Real-time Earthquake Risk Reduction for a Resilient Europe project, which supports the Collaboratory for the Study of Earthquake Predictability network activities in Europe (for the websites, see Data and Resources; Zechar *et al.*, 2010; Schorlemmer *et al.*, 2018).

Finally, we suggest that the implementation of the TGRE may offer some benefits for OEF models because it overcomes one of the conundrums of the best-performing current clustering models (Taroni *et al.*, 2018) in which the likelihood of a large earthquake is exactly where another large earthquake has just occurred. This conundrum has also been identified as one of the main reasons for the instability of the forecasts produced by the UCERF3-ETAS model, which has to impose a space–time variability of the MFD to solve the problem (Field, Milner, *et al.*, 2017). At the same time, TGRE may also provide a different explanation of recent empirical evidence relative to variations of the b-value before and after large earthquakes close to faults (Gulia and Wiemer, 2019). In particular, although it is worth remarking that the meaning of the b-value is questionable for a distribution that is not exponential, such as the TGR, if the corner magnitude gets closer to the completeness threshold (even though the slope remains the same), the b-value necessarily increases (Marzocchi *et al.*, 2020).

More generally, because the use of a proper MFD may have a large impact on the earthquake predictability, we hope that this article will stimulate further thoughts on this issue.

DATA AND RESOURCES

The data used in this study are available at http://www.wgcep.org/ucerf3, https://pubs.usgs.gov/of/2013/1165/, and https://pubs.usgs.gov/of/2007/1437/ (last accessed January 2019). Finally, for the Real-time Earthquake Risk Reduction for a Resilient Europe (RISE) and Collaboratory for the Study of Earthquake Predictability (CSEP) projects, see, respectively, www.rise-eu.org (last accessed April 2020) and https://cseptesting.org (last accessed August 2020).

ACKNOWLEDGMENTS

The authors thank the reviewers and Associate Editor Matthew C. Gerstenberger for their useful comments and suggestions, which significantly improved the quality and understandability of this article. This study was partially supported by the Real-time Earthquake Risk Reduction for a Resilient Europe (RISE) project and funded by the European Union's Horizon 2020 research and innovation program (Grant Agreement Number 821115).

REFERENCES

Allen, T. I., and G. P. Hayes (2017). Alternative rupture-scaling relationships for subduction interface and other offshore environments, *Bull. Seismol. Soc. Am.* **107,** no. 3, 1240–1253, doi: 10.1785/0120160255.

Bak, P., and C. Tang (1989). Earthquakes as a self-organized critical phenomenon, *J. Geophys. Res.* **94,** no. B11, 15,635–15,637, doi: 10.1029/JB095iB05p06853.

Bateman, H. (1953). *Higher Transcendental Functions*, Mc Graw-Hill Book Company, Inc, New York.

Bebbington, M. S., D. S. Harte, and S. C. Jaumé (2010). Repeated intermittent earthquake cycles in the San Francisco Bay region, *Pure Appl. Geophys.* **167,** 801–818, doi: 10.1007/s00024-010-0064-6.

Ben-Zion, Y., M. Eneva, and Y. Liu (2003). Large earthquake cycles and intermittent criticality on heterogeneous faults due to evolving stress and seismicity, *J. Geophys. Res.* **108,** no. B6, doi: 10.1029/2002JB002121.

Bowman, D. D., and C. G. Sammis (2004). Intermittent criticality and the Gutenberg–Richter distribution, *Pure Appl. Geophys.* **161,** no. 9, 1945–1956, doi: 10.1007/s00024-004-2541-z.

Field, E. H., R. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. M. Madugo, *et al.* (2013). Overview of the Uniform California Earthquake Rupture Forecast version 3 (UCERF3) time-independent model, *AGU Fall Meeting*, San Francisco, 9–13 December 2013, Abstract S51F-04.

Field, E. H., T. E. Dawson, K. R. Felzer, A. D. Frankel, V. Gupta, T. H. Jordan, T. Parsons, M. D. Petersen, R. S. Stein, R. J. Weldon II, *et al.* (2009). Uniform California Earthquake Rupture Forecast, version 2 (UCERF2), *Bull. Seismol. Soc. Am.* **99,** no. 4, 2053, doi: 10.1785/0120080049.

Field, E. H., T. H. Jordan, M. T. Page, K. R. Milner, B. E. Shaw, T. E. Dawson, G. P. Biasi, T. Parsons, J. L. Hardebeck, A. J. Michael, *et al.* (2017). A synoptic view of the Third Uniform California Earthquake Rupture Forecast (UCERF3), *Seismol. Res. Lett.* **88,** no. 5, 1259, doi: 10.1785/0220170045.

Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. van der Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017). A spatiotemporal clustering model of the Third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast, *Bull. Seismol. Soc. Am.* **107,** no. 3, 1049, doi: 10.1785/0120160173.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** 193–199, doi: 10.1038/s41586-019-1606-4.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34,** no. 8, 185–188.

Hardebeck, J. L., and T. Okada (2018). Temporal stress changes caused by earthquakes: A review, *J. Geophys. Res.* **123,** no. 2, 1350–1365, doi: 10.1002/2017JB014617.

Ide, S., and G. C. Beroza (2001). Does apparent stress vary with earthquake size? *Geophys. Res. Lett.* **28,** no. 17, 3349–3352, doi: 10.1029/2001GL013106.

Jordan, T., Y. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka, and J. Zschau (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization, *Ann. Geophys.* **54,** 8, doi: 10.4401/ag-5350.

Kagan, Y. Y. (2002a). Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.* **148,** no. 3, 520–541, doi: 10.1046/j.1365-246x.2002.01594.x.

Kagan, Y. Y. (2002b). Seismic moment distribution revisited: II. Moment conservation principle, *Geophys. J. Int.* **149,** no. 3, 731–754, doi: 10.1046/j.1365-246X.2002.01671.x.

Kagan, Y. Y., and F. Schoenberg (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution, *J. Appl. Probab.* **38,** 1, doi: 10.1239/jap/1085496599.

Kagan, Y. Y., P. Bird, and D. D. Jackson (2010). Earthquake patterns in diverse tectonic zones of the globe, *Pure Appl. Geophys.* **167,** 721–741, doi: 10.1007/s00024-010-0075-3.

Kanamori, H. (1977). The energy release in great earthquakes, *J. Geophys. Res.* **82,** no. 20, 2981–2987, doi: 10.1029/JB082i020p02981.

Kanamori, H., and E. E. Brodksy (2004). The physics of earthquakes, *Rep. Progr. Phys.* **67,** no. 8, 1429, doi: 10.1088/0034-4885/67/8/r03.

Kass, R. E., and A. E. Raftery (1995). Bayes factors, *J. Am. Stat. Assoc.* **90,** no. 430, 773–795, available at http://www.jstor.org/stable/2291091 (last accessed September 2020).

King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, University of Michigan Press, Ann Arbor, Michigan, available at http://www.press.umich.edu/titleDetailDesc.do?id=23784 (last accessed April 2019).

Madden, E., and D. Pollard (2012). Integration of surface slip and aftershocks to constrain the 3d structure of faults involved in the M 7.3 Landers earthquake, southern California, *Bull. Seismol. Soc. Am.* **102,** 321–342, doi: 10.1785/0120110073.

Marsan, D., and Y. J. Tan (2020). Maximum earthquake size and seismicity rate from an ETAS model with slip budget, *Bull. Seismol. Soc. Am.* **110,** no. 2, 874–885, doi: 10.1785/0120190196.

Marzocchi, W., I. Spassiani, A. Stallone, and M. Taroni (2020). Erratum to: How to be fooled searching for significant variations of the *b*-value, *Geophys. J. Int.* **221,** no. 1, 351–351, doi: 10.1093/gji/ggaa061.

Marzocchi, W., M. Taroni, and G. Falcone (2017). Earthquake forecasting during the complex Amatrice–Norcia seismic sequence, *Sci. Adv.* **3,** no. 9, doi: 10.1126/sciadv.1701239.

Massey, F. J., Jr. (1951). The Kolmogorov–Smirnov test for goodness of fit, *J. Am. Stat. Assoc.* **46,** no. 253, 68–78, doi: 10.1080/01621459.1951.10500769.

Michael, A. J., S. K. McBride, J. L. Hardebeck, M. Barall, E. Martinez, M. T. Page, N. van der Elst, E. H. Field, K. R. Milner, and A. M. Wein (2019). Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 M_w 7.1 Anchorage, Alaska, earthquake, *Seismol. Res. Lett.* **91,** no. 1, 153–173, doi: 10.1785/0220190196.

Omi, T., Y. Ogata, K. Shiomi, B. Enescu, K. Sawazaki, and K. Aihara (2018). Implementation of a real-time system for automatic aftershock forecasting in Japan, *Seismol. Res. Lett.* **90,** no. 1, 242–250.

Oth, A., D. Bindi, S. Parolai, and D. Di Giacomo (2010). Earthquake scaling characteristics and the scale-(in)dependence of seismic energy-to-moment ratio: Insights from KiK-net data in Japan, *Geophys. Res. Lett.* **37,** no. 19, L19304 , doi: 10.1029/2010GL044572.

Papazachos, B., E. Scordilis, D. Panagiotopoulos, C. Papazachos, and G. Karakaisis (2004). Global relations between seismic fault parameters and moment magnitude of earthquakes, *Bull. Geol. Soc. Greece* **36,** no. 3, 1482–1489, doi: 10.12681/bgsg.16538.

Parsons, T., E. L. Geist, R. Console, and R. Carluccio (2018). Characteristic earthquake magnitude frequency distributions on faults calculated from consensus data in California, *J. Geophys. Res.* **123,** no. 12, 10,761–10,784, doi: 10.1029/2018JB016539.

Reid, H. F. (1911). The elastic-rebound theory of earthquakes, *Univ. Calif. Publ. Bull. Dept. Geol.* **6,** no. 19, 413–444, available at https://ci.nii.ac.jp/naid/10006706254/en/ (last accessed April 2019).

Salisbury, J. B., T. K. Rockwell, T. J. Middleton, and K. W. Hudnut (2012). LiDAR and field observations of slip distribution for the most recent surface ruptures along the central San Jacinto fault, *Bull. Seismol. Soc. Am.* **102,** no. 2, 598–619, doi: 10.1785/0120110068.

Schorlemmer, D., M. Werner, W. Marzocchi, T. Jordan, Y. Ogata, D. Jackson, S. Mak, D. Rhoades, M. Gerstenberger, N. Hirata, *et al.* (2018). The collaboratory for the study of earthquake predictability: Achievements and priorities, *Seismol. Res. Lett.* **89,** no. 4, 1305–1313, doi: 10.1785/0220180053.

Sieh, K., L. Jones, E. Hauksson, K. Hudnut, D. Eberhart-Phillips, T. Heaton, S. Hough, K. Hutton, H. Kanamori, A. Lilje, *et al.* (1993). Near-field investigations of the Landers earthquake sequence, April to July 1992, *Science* **260,** no. 5105, 171–176, doi: 10.1126/science.260.5105.171, available at http://www.jstor.org/stable/2881298 (last accessed March 2020).

Sornette, A., and D. Sornette (1989). Self-organized criticality and earthquakes, *Europhys. Lett.* **9,** no. 3, 197–202, doi: 10.1209/0295-5075/9/3/002.

Spassiani, I. (2020). Stability of the epidemic type aftershock sequence model with tapered Gutenberg–Richter distributed seismic moments, *Bull. Seismol. Soc. Am.* doi: 10.1785/0120200012.

Stallone, A., and W. Marzocchi (2019). Empirical evaluation of the magnitude-independence assumption, *Geophys. J. Int.* **216,** no. 2, 820–839, doi: 10.1093/gji/ggy459.

Taroni, M., W. Marzocchi, D. Schorlemmer, M. Werner, S. Wiemer, J. Zechar, L. Heiniger, and F. Euchner (2018). Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy, *Seismol. Res. Lett.* **89,** no. 4, 1251–1261, doi: 10.1785/0220180031.

Temme, N. M. (1996). *Special Functions. An introduction to the Classical Functions of Mathematical Physics*, John Wiley and Sons Inc., New York, New York, 392 pp.

van der Elst, N. J., and B. E. Shaw (2015). Larger aftershocks happen farther away: Nonseparability of magnitude and spatial distributions of aftershocks, *Geophys. Res. Lett.* **42,** no. 14, 5771–5778, doi: 10.1002/2015GL064734.

Vere-Jones, D. (1978). Earthquake predicion—A statistician's view, *J. Phys. Earth* **26,** no. 2, 129–146, doi: 10.4294/jpe1952.26.129.

Vere-Jones, D. (1988) On the variance properties of the stress release models, *Aust. J. Stat.* **30A,** no. 1, 123–135, doi: 10.1111/j.1467-842X.1988.tb00469.x.

Wang, A.-L., D. Vere-Jones, and X. Zheng (1991). Simulation and estimation procedures for stress release model, in *Stochastic Processes and their Applications*, M. J. Beckmann, M. N. Gopalan, and R. Subramanian (Editors), Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-58201-1, 11–27.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.* **84,** no. 4, 974.

Xiaogu, Z., and D. Vere-Jones (1994). Further applications of the stochastic stress release model to historical earthquake data, *Tectonophysics* **229,** no. 1, 101–121, doi: 10.1016/0040-1951(94)90007-8.

Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2010). The collaboratory for the study of earthquake predictability perspective on computational earthquake science, *Concurr. Comp. Pract. E.* **22,** no. 12, 1836–1847, doi: 10.1002/cpe.1519.

Zheng, X.-G., and D. Vere-Jones (1991). Application of stress release models to historical earthquakes from North China, *Pure Appl. Geophys.* **135,** no. 4, 559–576, doi: 10.1007/BF01772406.

Zhuang, J., M. Werner, S. Hainzl, D. Harte, and S. Zhou (2012). Basic models of seismicity: Temporal models, Community Online Resource for Statistical Seismicity Analysis, 33.

Zöller, G., M. Holschneider, and S. Hainzl (2013). The maximum earthquake magnitude in a time horizon: Theory and case studies, *Bull. Seismol. Soc. Am.* **103,** no. 2A, 860–875, doi: 10.1785/0120120013.

Manuscript received 25 May 2020
Published online 19 January 2021

b-value of what? Complex behavior of the magnitude distribution during and within the 2016–2017 central Italy sequence

Marcus Herrmann (marcus.herrmann@unina.it)
Università degli Studi di Napoli 'Federico II' https://orcid.org/0000-0002-2342-1970
Ester Piegari
Università degli Studi di Napoli 'Federico II'
Warner Marzocchi
University of Naples, Federico II https://orcid.org/0000-0002-9114-1516

Article

Keywords:

Posted Date: January 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1210699/v1

---

b-value of what? Complex behavior of the magnitude distribution during and within the 2016–2017 central Italy sequence

Marcus Herrmann*[1], Ester Piegari[1], Warner Marzocchi[1]

[1] Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse; Università degli Studi di Napoli 'Federico II', Naples, Italy

* Corresponding author: marcus.herrmann@unina.it

December 28, 2021

Abstract:

The Magnitude–Frequency-Distribution (MFD) of earthquakes is typically modeled with the (tapered) Gutenberg–Richter relation. The main parameter of this relation, the b-value, controls the relative rate of small and large earthquakes. Resolving spatiotemporal variations of the b-value is critical to understanding the earthquake occurrence process and improving earthquake forecasting. However, this variation is not well understood. Here we present unexpected MFD variability using a high-resolution earthquake catalog of the 2016–2017 central Italy sequence. Isolation of seismicity clusters reveals that the MFD differs in nearby clusters, varies or remains constant in time depending on the cluster, and features an unexpected b-value increase in the cluster where the largest event will occur. These findings suggest a strong influence of the heterogeneity and complexity of tectonic structures on the MFD. Our findings raise the question of the appropriate spatiotemporal scale for resolving the b-value, which poses a serious obstacle to interpreting and using the MFD in earthquake forecasting.

1 Introduction

Beroza et al. [2021] recently highlighted that current earthquake catalogs achieve a high level of detail that likely contains more information about earthquake occurrence, allows testing of existing hypotheses, and potentially improves earthquake forecasting. One of the main ingredients for earthquake forecasting and seismic hazard models is the Magnitude–Frequency-Distribution (MFD) of earthquakes, which carries information about the proportion between small and large earthquakes. The MFD is typically modeled with the Gutenberg–Richter (GR) relation and its b-value (the slope of the GR relation), which can be used to infer the occurrence rate of large earthquakes from small ones. The b-value is observed to vary in space and time [e. g., *Wiemer and Wyss* 1997; *Hainzl and Fischer* 2002; *Tormann et al.* 2013; *Gulia et al.* 2016; *Shelly et al.* 2016; *Petruccelli et al.* 2019; *Taroni et al.* 2021], which is thought to be primarily related to variations of the stress state in the crust [e. g., *Wyss* 1973; *Scholz* 2015; *El-Isa and Eaton* 2014]. The b-value is also considered as an indicator for other conditions in the crust, which are directly or indirectly related to the stress state, such as faulting style [e. g., *Schorlemmer et al.* 2005; *Petruccelli et al.* 2019], locked or creeping fault patches [e. g., *Wiemer and Wyss* 1997; *Sobiesiak et al.* 2007; *Ghosh et al.* 2008; *Tormann et al.* 2013], material properties [e. g., *Mogi* 1962; *Goebel et al.* 2017], fluid pore-pressure perturbations [e. g., *Hainzl and Fischer* 2002; *Bachmann et al.* 2012; *Shelly et al.* 2016; *Passarelli et al.* 2015], and critical nucleation length [*Dublanchet* 2020], among others [see *El-Isa and Eaton* 2014, and references therein]. b-value variations may therefore have an important role in improving our physical understanding of earthquake occurrence.

Estimating the b-value appears trivial in theory (after all, it is simply the rate parameter of an exponential distribution), but not in practice. Several aspects affect the ability to resolve representative b-value variations in earthquake catalogs, such as:

1. the quality of the data, its spatiotemporal selection, and the various ways of sampling it [e. g., *Tormann et al.* 2013; *Roberts et al.* 2016];
2. the sample size and available magnitude range [e. g., *Wiemer and Wyss* 2002; *Marzocchi and Sandri* 2003; *Roberts et al.* 2016; *Nava et al.* 2016];
3. the used magnitude scale, magnitude binning, and maximum likelihood estimator [*Wiemer and Wyss* 2002; *Marzocchi and Sandri* 2003; *Marzocchi et al.* 2020; *Herrmann and Marzocchi* 2021];
4. the treatment of departures from an exponential-like GR distribution at the upper end (due to truncation or tapering) and lower end (due to the inherent and potentially varying incompleteness) [*Kagan* 2002; *Spassiani and Marzocchi* 2021; *Marzocchi et al.* 2020; *Herrmann and Marzocchi* 2021], e. g., the estimation of the magnitude of completeness, M_c, as the lower magnitude threshold.

Although this list is not exhaustive, these considerations highlight that the outcome of a b-value analysis highly depends on expert judgment and/or subjective choices. A recent scientific discussion between *Gulia and Wiemer* [2021] and *Dascher-Cousineau et al.* [2021] reemphasized that choices have to be specific, meaningful, and reproducible to obtain robust results that contribute to a better understanding of the underlying physical processes. It appears that this field of study requires well-defined schemes and analysis steps. Moreover, choices are critical for real-time applications that need to run automatically, e. g., for operational earthquake forecasting (OEF) purposes [*Jordan et al.* 2011]. Assessing the influence of expert choices and various modeling ideas on the forecasting performance needs community efforts such as the *Collaboratory for the Study of Earthquake Predictability* (CSEP) [*Zechar et al.* 2010; *Schorlemmer et al.* 2018], which tests forecasting models prospectively in a controlled environment.

Here we argue that a complex earthquake sequence with multiple ruptured fault segments can further bias MFD and b-value analysis: If the MFD varies temporally among tectonic structures, an averaged view over the whole sequence or a smoothed view over a finite scale (either in space or time) will neglect or mask those variations and may lead to inappropriate or biased inferences. Instead, an MFD analysis may become more physically meaningful and less ambiguous when accounting for the internal structure and evolution of a sequence. *Igonin et al.* [2018] already showed that the MFD can significantly differ in adjacent but well-defined zones of induced seismicity. In this study, we introduce a new perspective to investigate the spatiotemporal behavior of the MFD and b-value by isolating spatial seismicity clusters of a complex sequence and dividing them into temporal periods. To define clusters, we use the hypocenter density of a seismic sequence; the temporal periods are defined by occurrence time of the largest events.

We use the 2016–2017 central Italy (hereafter 'CI2016') sequence as an example due to its complex tectonic structure, cascading evolution, and the availability of high-resolution catalogs. The CI2016 sequence occurred in the central Apennines, one of Italy's most seismically active areas, and was marked by a cascade of three main events: the $M_\mathrm{w}6.0$ ($M_\mathrm{L}6.0$) Amatrice event on 24 August 2016, the $M_\mathrm{w}5.9$ ($M_\mathrm{L}5.8$) Visso event on 26 October 2016, and the $M_\mathrm{w}6.5$ ($M_\mathrm{L}6.1$) Norcia mainshock on 30 October 2016. On 18 January 2017, four $M_\mathrm{w}5.0$–5.5 events followed near Campotosto. These seven events have been caused by movements on SW-dipping normal faults and they ruptured multiple fault segments, activating a complex fault system [e. g., *Chiaraluce et al.* 2017; *Improta et al.* 2019; *Michele et al.* 2020; *Porreca et al.* 2020; *Tondi et al.* 2020; *Waldhauser et al.* 2021].

The CI2016 sequence is particular in that it features seismicity in a ~1 km-thick subhorizontal detachment at around 10 km depth, which intersects with and confines almost the entire normal fault system above [e. g., *Michele et al.* 2016; *Chiaraluce et al.* 2017; *Vuan et al.* 2017; *Michele et al.* 2020; *Waldhauser et al.* 2021]. Such a feature was already observed in the Apennines at a depth of 15–20 km [*De Luca et al.* 2009], which suggested the presence of a buried subhorizontal thrust related to (the deepest part of) the Apennines build-up. It generally appears as a flat layer, and high-resolution catalogs resolved it as a slightly east-dipping, irregular structure (i. e., with locally varying depth and thickness) [*Chiaraluce et al.* 2017]. *Vuan et al.* [2017] interpreted this feature as a midcrustal shear zone, which decouples the upper and lower crust. They found prior seismicity mostly to occur along this structure, suggesting that it was loaded tectonically and eventually favored the unlocking of the shallower faults through stress transfer. *Waldhauser et al.* [2021] identified partially overlapping fault fragments in this structure.

Magnitude statistics of CI2016 have been investigated in several recent studies. *Montuori et al.* [2016] found that the Amatrice event originated in an area with a high b-value and subsequently reduced the b-value to the north and south, suggesting a high potential for further large events. *Gulia and Wiemer* [2019] found a b-value variation during the course of the sequence, in particular (i) a drop after the Amatrice event (especially in the area to the north where the Norcia mainshock occurred afterward), interpreted as a still impending large earthquake, and (ii) a b-value increase after the Norcia mainshock, interpreted as a substantially reduced chance for a further large event similar to the tectonic background rate. *García-Hernández et al.* [2021] also observed a "marked drop of the b-value" after the Amatrice event (resolved spatially and in depth) and a recovery of the b-value to the background level after the Norcia mainshock; they exclude that these variations are caused by an increased M_c after the main events.

However, those studies did not (i) use a high-resolution catalog, (ii) account for the complexity of the sequence including its depth-dependent structure, and (iii) resolve what happened in the days before the largest event (Norcia). Using a high-resolution catalog, we investigate whether accounting for the complex sequence in an isolated fashion provide a benefit in resolving the spatiotemporal variation of the MFD and b-value. Rather than solely focusing on b-value estimates, we consider it important to exploit more information from the MFD, e. g., by assessing and comparing its exponential-like part and reporting the b-value stability as function of M_c.

2 Results

2.1 Description of clusters

Using the high-resolution catalog of *Tan et al.* [2021], we spatially isolated the five largest seismicity clusters (Cluster 1–5, hereafter abbreviated with C1, C2, etc.) following the procedure described in Methods. Figure 1 shows that the obtained clusters are not randomly distributed, but instead highlight the complex spatial structure of the sequence. For instance, C1 comprises seismicity in the northern part of the subhorizontal structure, parts of the normal fault (Mt. Vettore) that ruptured during the Norcia mainshock, and this mainshock hypocenter itself. C2 represents seismicity in the southern part of the subhorizontal structure, and C3 captures the shallow northern part of the sequence, including the Visso hypocenter. C4 and C5 relate to small-scale structures. These five clusters correspond to the five largest volumes with high hypocenter density (see Supplementary Fig. S1). The Amatrice event does not belong to any of the main clusters because the area around its hypocenter is devoid of events [see also *Improta et al.* 2019; *Michele et al.* 2020; *Tan et al.* 2021]. The Campotosto events were also not assigned to a main cluster.

Figure 2 shows that each cluster has a distinct temporal activity. For instance, C1 was active throughout the sequence until the Campotosto events; C2 was quiet after the Visso event until the Norcia mainshock while C3 was very active in this period. C4 and C5 were mostly active toward the end of the sequence, along with the other clusters in roughly comparable proportions. Supplementary Fig. S2 and Supplementary Note 1 summarize the cluster statistics in terms of size and ratio for each period, making it more apparent that at least ~50% of the events in each period of the sequence belong to a cluster. Moreover, up to two clusters were dominating each period except for the last period.

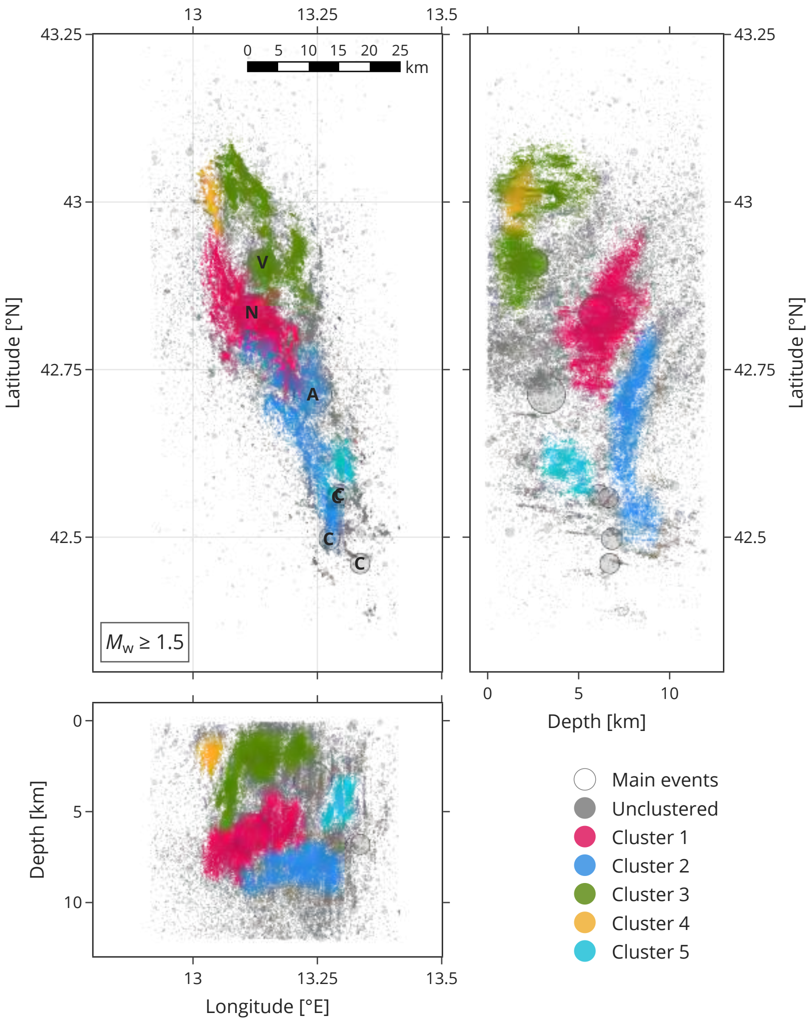


Figure 1: Map view and depth sections of the 2016–2017 central Italy ('CI2016') seismicity with identified clusters (see legend). The depth sections are exaggerated by a factor of three. To better reveal the structure of the individual clusters, the events are plotted ascending by their cluster number on top of 'unclustered' events, neglecting a physically correct appearance. The main events 'Amatrice', 'Visso', 'Norcia', and four 'Campotosto' events are represented by larger circles; in the map view, they are annotated with the respective initial letter (A, V, N, C). Supplementary Fig. S1 shows the event density for the same data.

2.2 Cluster-based MFD analysis using the whole sequence

For the statistical analysis of the MFD, we follow the procedure described in Methods. Table 1 and Fig. 3 indicate differences and similarities in the MFD among the clusters. In particular, Table 1 suggests that C1, C2, and C3 have identical MFD shapes, but that the MFD of C1 and C2 are distinct from the ones of C4 and C5. There is a tendency that C1 differs from C2, although not statistically significant. Figure 3 provides more details about the MFD behavior in terms of the b-value as function of M_c. For instance, the largest clusters C1–C3 (red, blue, and green, respectively) have comparable b-values ($\sim$1.2) at their corresponding $M_c^{\mathrm{Lilliefors}}$, but behave differently for increasing M_c: for $M_w \geq 3.0$, the b-value is much higher in C3 than in C1 or C2. The small-scale clusters C4

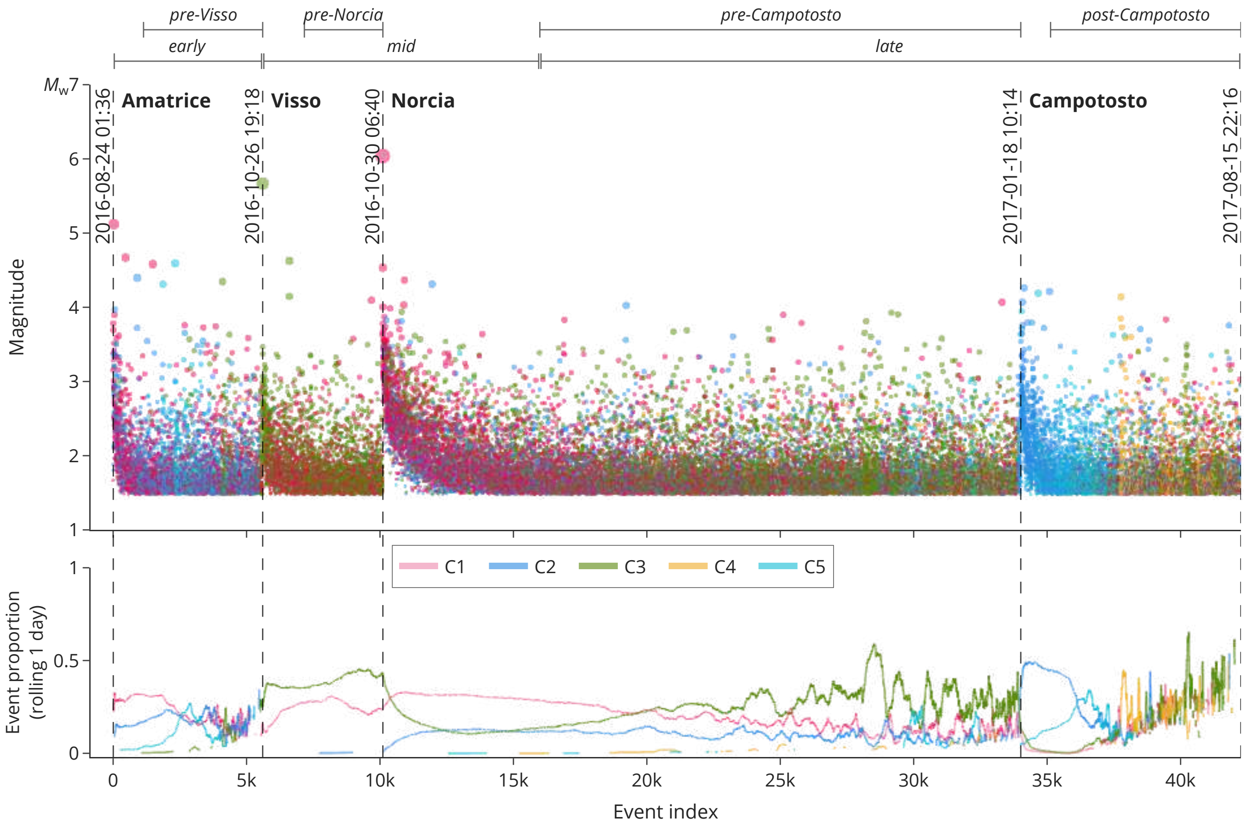


Figure 2: Temporal evolution of CI2016 seismicity colored by cluster association. Note that the x-axis represents the event index of all shown events (i. e., excluding unclustered events). The horizontal whiskers at the top indicate the periods of the temporal subsets. The bottom panel shows the event proportion of each cluster as fraction of the total events (including unclustered events) within a rolling window of the previous 24 hr.

Table 1: Pairwise comparison of the cumulative magnitude distribution of each cluster against the others.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1		0.089	0.26	**2.2e-06**	**1.2e-06**
Cluster 2	0.089		0.59	**0.024**	**0.0084**
Cluster 3	0.26	0.59		0.49	0.36
Cluster 4	**2.2e-06**	**0.024**	0.49		0.51
Cluster 5	**1.2e-06**	**0.0084**	0.36	0.51	

p-values of two-sample Kolmogorov–Smirnov tests (see Methods). Statistically significant p-values are highlighted in bold.

and C5 (yellow and cyan, respectively) show the highest overall b-value. The Lilliefors p-value is useful to judge the reliability of the b-value; a p-value dropping below 0.1 indicates that the b-value for C1 and C3 below M_w2.0 does not relate to a persistent exponentiality with M_c, which can have several reasons (see Supplementary Note 2.2) and necessitates an inspection of the MFD in individual periods, as done in the following subsection.

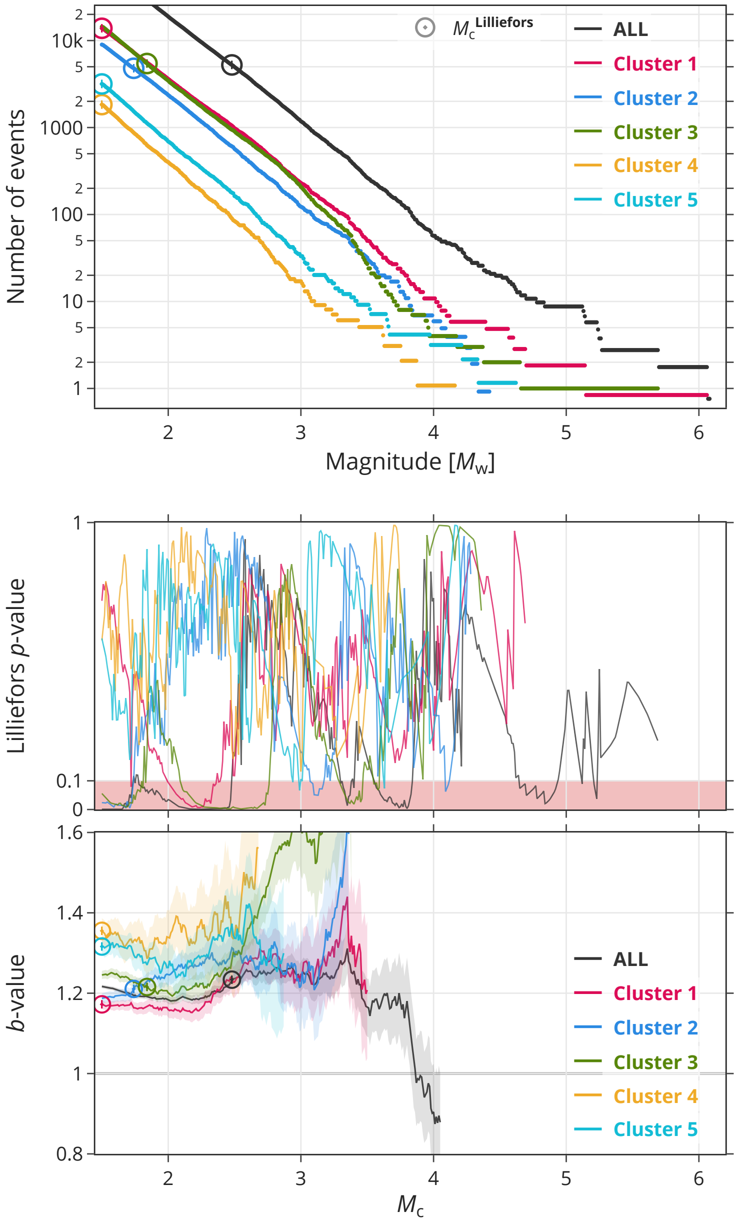


Figure 3: Magnitude statistics for all extracted data (black) and individual clusters (see legend). The top panel shows the data in terms of their magnitude–frequency distribution (MFD). Note that a tiny value is added to each MFD (between -0.1 and 0.1) to avoid visual overlaps at large magnitudes. The middle and bottom panel show, as a function of lower magnitude cutoff, or magnitude of completeness, M_c, the Lilliefors p-value (assuming an exponential distribution as null hypothesis) and the b-value (the slope of the fitted Gutenberg–Richter relation), respectively (see Methods). The $M_c^{\text{Lilliefors}}$ estimates are indicated for each cluster in the top and bottom panels with a circle marker. Supplementary Fig. S3 shows the same analysis using local magnitudes.

For the sake of completeness, we repeated the analysis using local magnitudes, M_L (see Supplementary Fig. S3), which introduces a different MFD behavior for the individual clusters due to a narrower exponential range (see Supplementary Note 2.3).

Table 2: Pairwise MFD comparison of temporal subsets. Like Table 1, but for three periods of Cluster 1, 2, and 3 that exclude short-term incompleteness (STAI).

	Cluster 1			Cluster 2			Cluster 3		
	(pre-V.)	(pre-N.)	(pre-C.)	(pre-V.)	(pre-N.)	(pre-C.)	(pre-V.)	(pre-N.)	(pre-C.)
C1 (pre-Visso)		**0.0037**	0.33	0.59	0.73	**0.03**	**2.1e-05**	0.86	0.79
C1 (pre-Norcia)	**0.0037**		0.23	**0.01**	0.73	**0.044**	**1.3e-09**	**0.0056**	0.052
C1 (pre-Campotosto)	0.33	0.23		0.52	0.42	0.21	**6e-06**	0.62	**0.013**
C2 (pre-Visso)	0.59	**0.01**	0.52		0.78	0.28	**2.5e-06**	0.51	0.067
C2 (pre-Norcia)	0.74	0.72	0.42	0.79		0.51	0.093	0.43	0.26
C2 (pre-Campotosto)	**0.03**	**0.044**	0.21	0.28	0.5		**7.6e-08**	**0.0087**	**0.018**
C3 (pre-Visso)	**2.1e-05**	**1.3e-09**	**5.8e-06**	**2.5e-06**	0.093	**7.6e-08**		**3.1e-06**	**0.012**
C3 (pre-Norcia)	0.86	**0.0056**	0.62	0.51	0.42	**0.0087**	**3.2e-06**		0.098
C3 (pre-Campotosto)	0.79	0.052	**0.013**	0.068	0.26	**0.018**	**0.012**	0.099	

2.3 Cluster-based MFD analysis using temporal subsets

We extend the spatial analysis by a temporal component using three periods that exclude the short-term aftershock incompleteness (STAI) between the main events, namely 'pre-Visso', 'pre-Norcia', and 'pre-Campotosto' (see Methods). Table 2 provides a more granular breakdown of MFD variations above $M_c^{\mathrm{Lilliefors}}$ than Table 1, also temporally within the same cluster. For instance, in C1, only pre-Visso and pre-Norcia are distinct; in C2, no period is distinct, and in C3, pre-Visso is distinct from the other two periods. The MFD in pre-Campotosto is never distinct in any cluster. Comparisons among clusters for the same temporal period show no significant differences between C1 and C2, but when comparing C1 or C2 to C3. (Note that the sample size of C2 in pre-Norcia is very small (26 events), which reduces the power of the KS test to detect potential differences for pairs that include this subset.) The most unique subset is C3 during pre-Visso, which differs from almost all other subsets. Of all 36 pairs, 15 (42 %) are significantly different.

Further investigating the MFDs in terms of a M_c-dependent b-value (Figs. 4 and 5) provides a more nuanced discrimination. The most remarkable observation is that the b-value in C1 is highest before the Norcia mainshock—it has increased after the Visso event from 1.4 to 1.6. After the Norcia mainshock, the b-value remained at a high level (1.5 in the pre-Campotosto period). In C2, the b-value remained high at $\sim$1.45 both before the Visso and after the Norcia mainshock. (This cluster does not have enough events in the pre-Norcia period to estimate a b-value.) In C3, which contains the Visso event, the b-value increased from 1.0 in pre-Visso to 1.4 in pre-Norcia, at which level it stayed also after the Norcia mainshock.

Fig. 4 facilitates a temporal comparison of the MFD among the clusters. In pre-Visso, the b-value is similar in C1 and C2 at around 1.4, and much lower in C3 (1.0). Prior to the Norcia mainshock, the b-value increased both in C1 and C3 (to 1.4−1.6); C2 does not provide enough data. After the Norcia mainshock (i. e., pre-Campotosto), the b-value remains elevated in C1–C3 (1.3−1.5) and C1 and C2 have similar b-values again. After the Campotosto events (see Supplementary Fig. S8, 'post-Campotosto'), the b-value still remains elevated in C1–C3 (1.4−1.5).

For the sake of completeness, we repeated the analysis using M_L (see Supplementary Figs. S4, S5,

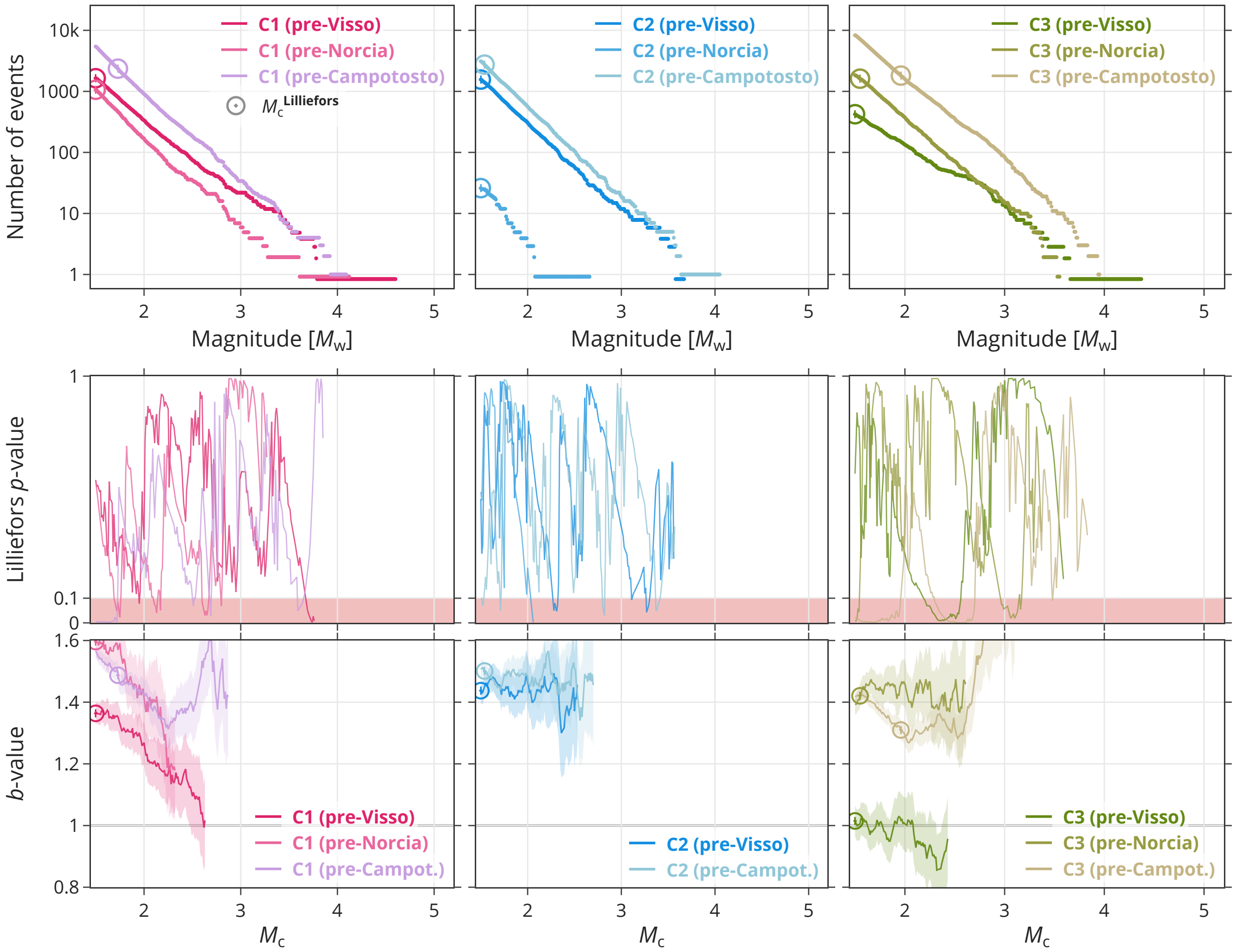


Figure 4: Magnitude statistics in three individual periods of Cluster 1 (left), Cluster 2 (center), and Cluster 3 (right). Like Fig. 3, the top panels show clusters in terms of their magnitude–frequency distribution; the middle and bottom panels show the Lilliefors p-value and b-value as function of M_c, respectively. Supplementary Fig. S4 shows the same analysis using local magnitudes. Supplementary Figs. S7 and S8 compares the periods shown here with periods that include STAI.

and Supplementary Note 2.3), which reproduces our main findings qualitatively with comparable *relative* b-value changes, albeit the b-value behaves differently as function of M_c owing to the scale change. For a comparison using temporal periods that include STAI, see Supplementary Note 2.4 and Supplementary Figs. S6–S8.

3 Discussion

We found that individual earthquake clusters that represent the most active zones of a complex sequence are characterized by a significantly different MFD behavior. In particular, the MFD can experience variations as temporal changes and spatial differences, or remain identical within one cluster throughout the sequence. This observed MFD variability is likely due to fine-scale heterogeneity and complexity of the tectonic structures that were activated in this sequence. In

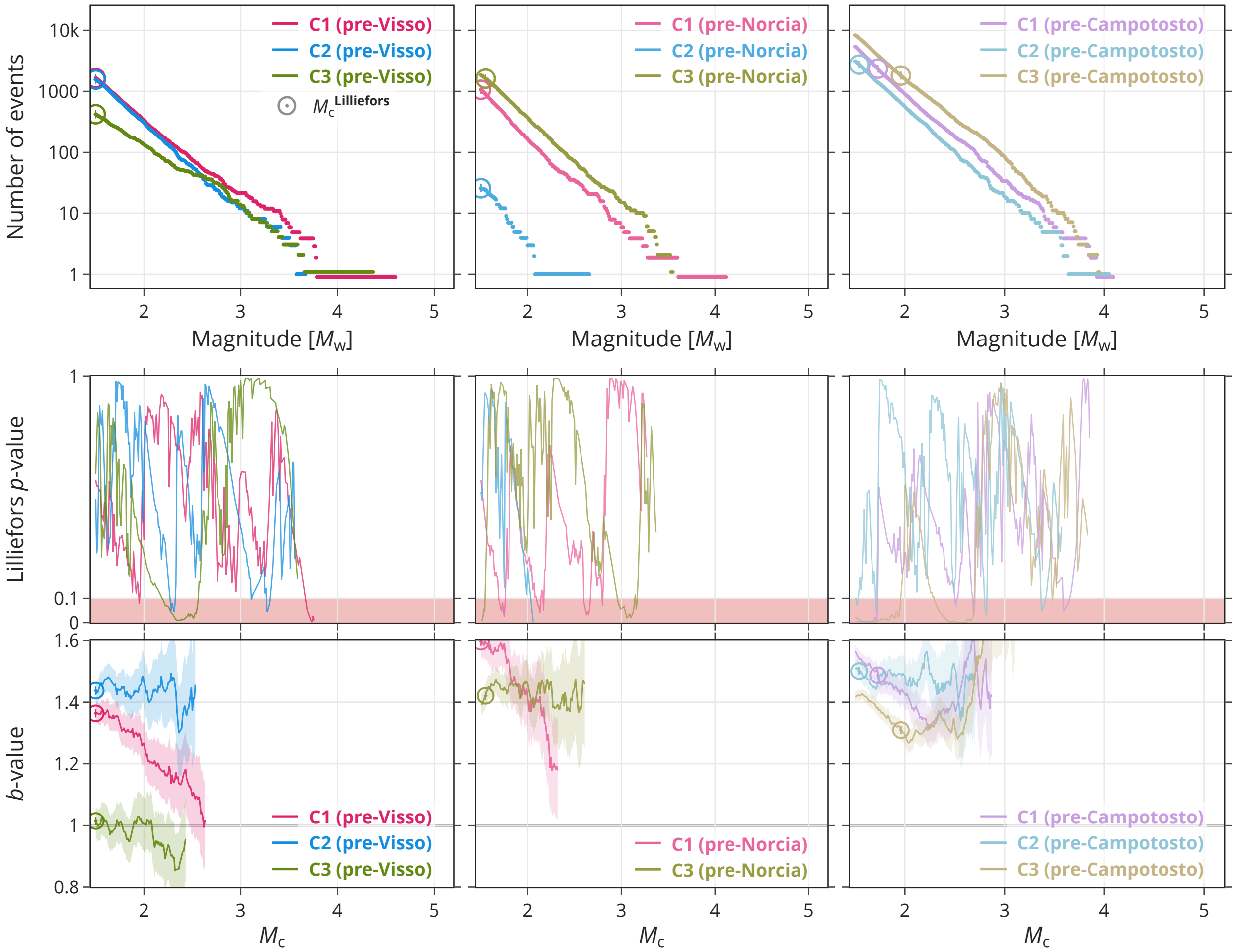


Figure 5: Reordering the magnitude statistics of Fig. 4 temporally by periods: 'pre-Visso' (left), 'pre-Norcia' (center), and 'pre-Campotosto' (right). Like Figs. 3 and 4, the top panels show clusters in terms of their magnitude–frequency distribution; the middle and bottom panels show the Lilliefors p-value and b-value as function of M_c. Supplementary Fig. S5 shows the same analysis using local magnitudes.

the following, we first discuss the observed temporal behavior, followed by a discussion of spatial differences and similarities, an interpretation of our findings, and a summary.

Regarding temporal changes, the most striking observation is the progressive b-value increase in the structure where the strongest earthquake eventually occurred (C1). Apparently, a high b-value did not prevent the nucleation of a large rupture in this structure. This resolved behavior differs from the general observation that the b-value decreases prior to large earthquakes [e. g., *Suyehiro et al.* 1964; *Nanjo et al.* 2012; *Tormann et al.* 2015; *Gulia et al.* 2016; *Gulia and Wiemer* 2019], albeit similar observations to ours do exist [e. g., *Nanjo and Yoshida* 2017]. The increasing b-value in C1 after the Visso event highlights that activity in one cluster may influence the MFD in another one. After the Visso event, the b-value also increased in its own cluster (C3), which corroborates that large events may influence the MFD in their surrounding as already found by [*Gulia et al.* 2018]. The later Norcia mainshock, however, did not alter the MFD in the three main clusters further and the b-value remained high—also after the Campotosto events. This stagnating b-value highlights that the MFD eventually became insensitive to strong seismicity even though it had experienced significant temporal variations in the same structures earlier. This ambivalent character is compounded by the MFD behavior in C2, where the MFD locally remained constant throughout the sequence and apparently unaffected by surrounding seismicity.

When comparing clusters spatially regarding the whole sequence, we found differences in the MFD between the largest clusters (C1 and C2) and the smaller ones (C4 and C5). The former have overall lower b-value estimates, which are due to the stronger influence of STAI as a result of their proximity to larger events. In fact, the b-value estimate can be much lower in periods that include STAI (see Supplementary Note 2.4 and Supplementary Figs. S7–S9). In each of these time periods, we found spatial MFD differences among the largest clusters (C3 differing from C1 and C2). Simultaneously, MFD similarities coexisted among these clusters (C1 and C2), although we do not have evidence for every time period, such as for pre-Norcia when C2 only provides few samples. C1 and C2 have in common that they represent the majority of seismicity in the subhorizontal structure at depth (its northern and southern extension, respectively). Their MFD differs from C3 in each individual period and tend toward a higher b-value, which means that this subhorizontal structure is not only tectonically distinct from the shallower normal faults (see Introduction), but also in terms of the MFD.

Although our study focuses on raising awareness of appropriately resolving MFD and b-value variations, we briefly speculate about the underlying causes for our most remarkable observations in this sequence. The marked MFD variability among the clusters over time likely reflects a heterogeneous stress field and/or a complex fault geometry with significant contributions from the subhorizontal detachment. Moreover, a complex rupture process is suggested by the fact that only some of the main events belong to clusters—a result of the different event densities surrounding these hypocenters. The generally higher b-value in the subhorizontal structure could be caused by the structure's reduced capacity to accumulate stress (i. e., low differential stress). Instead of accumulating stress, it preferentially transfers stress to the shallow fault system, favoring its unlocking [*Vuan et al.* 2017]. Moreover, this subhorizontal thrust is known to release microearthquakes quasi-continuously along its entire length [*Chiaraluce* 2012; *Chiaraluce et al.* 2017], occasionally in minor sequences [*Ciaccio* 2016; *Moschella et al.* 2021], but not hosting larger earthquakes (which should have an extensional mechanism). The very high b-value prior and close to the hypocenter of the Norcia mainshock could be explained with (i) the generally high b-value in the subhorizontal structure, because C1's pre-Norcia seismicity occurred within its north-eastern extension (see Supplementary Fig. S10), whereas its pre-Visso seismicity was located in a shallower part; and (ii) a consequence of the previous two main events (Amatrice and Visso) and their aftershocks generally reducing the differential stress in its surrounding by releasing built-up strain, i. e., stored energy, on the normal faults. Note that the Norcia mainshock nucleated in between the pre-Norcia and the pre-Visso subset of C1 (i. e., the aftershock zones of Amatrice and Visso, respectively, see also *Improta et al.* [2019]), which is consistent with observations that large events tend to nucleate at the rim of seismic clouds [van der *Elst and Shaw* 2015; *Stallone and Marzocchi* 2019] and the cascading stress transfer hypothesis [e. g., *Ellsworth and Bulut* 2018; *Gomberg* 2018].

In summary, our study demonstrated that the spatiotemporal isolation of seismicity clusters resolves a distinct MFD behavior among the most active zones over time, including influences between them. We therefore argue that the MFD highly depends on the observed substructure. Since the most active structures in turn influence the overall MFD behavior of a sequence, a consideration of the activity in individual structures allows us to decompose and analyze the most important contributions of a complex sequence. Our findings point to the problem of choosing an appropriate spatiotemporal scale to resolve the b-value, challenging existing approaches: A too large scale merges potentially different MFD behavior of individual structures and a too fine resolution obscures the tectonic relation and reduces the statistical robustness. The cluster-based approach presented here uses the distribution of the seismicity itself to choose a scale that is physically meaningful and provides robust statistics. Moreover, a spatial scale inferred from a cluster analysis may serve as an appropriate reference volume for the background b-value—provided that the (moment) magnitude estimates are consistent. In Supplementary Note 2 we discuss several more factors and choices that influence and potentially bias the b-value estimate, most importantly related to the sample size, exponentiality, STAI, and magnitude scale. Those aspects are not always carefully addressed when performing b-value analyses. We highlight that the absolute b-value has little meaning not only due to its dependence on the magnitude scale (see Supplementary Note 2.3), but also on the particular conversion relation (see Supplementary Note 2.5 and Supplementary Fig. S9). We hypothesize that a complex and distinct MFD behavior is not unique to the CI2016 sequence, but likely occurs in other regions and sequences. Our method may be beneficial in studying the peculiarities of spatiotemporal MFD variability and improving our understanding of the processes that influence seismicity. Even if the physical mechanisms remain hidden, recognizing that the MFD behaves complex potentially improves spatiotemporal forecast performance. Our approach based on an established clustering algorithm may also help to reduce the amount of expert judgment and subjective choices in MFD analysis, which could facilitate an application in real-time. Future work may focus on a refined identification of spatiotemporal clusters to improve MFD and b-value analysis, possibly by not relying solely on event density.

4 Methods

4.1 High-resolution earthquake catalog of the sequence

We use the high-resolution catalog of *Tan et al.* [2021], which spans from 2016-08-15 to 2017-08-15, and extracted a spatial subset as follows: depth < 12 km; UTM easting: 330–370 km (about longitude 12.94–13.40); UTM northing: 4690–4790 km (about latitude 42.34–43.25). Only events with moment magnitudes $M_\mathrm{w} \geq 1.5$ were used, totaling 76 055 events. The M_w contained in the catalog were converted from local magnitudes, M_L, using the polynomial fit of *Grünthal et al.* [2009], an average European scaling relation based on catalogs of different seismological agencies with most events having $M_\mathrm{L} > 1.5$ and $M_\mathrm{w} \gtrsim 1.5$. For magnitudes of large events to match, *Tan et al.* [2021] calibrated its constant (0.53) to 0.817 (i. e., +0.287).

4.2 Creating spatial earthquake clusters and temporal subclusters

Seismicity was grouped spatially into clusters using DBSCAN [Density-Based Spatial Clustering of Applications with Noise, *Ester et al.* 1996], which groups points based on how closely they are packed together. Points that lie in low-density regions are left as outliers. Because the horizontal extension of the CI2016 sequence is several times larger than the vertical extension, density-connected clouds of hypocenters preferentially extend in horizontal directions. To improve the clustering analysis for such an anisotropic case, data dimensions are usually rescaled beforehand. We therefore rescale the hypocenter coordinates to a uniform extent in each direction, i. e., rescaled into a cube. This procedure increases the local point density in horizontal planes, which facilitates identifying hypocenter clusters with horizontally elongated shapes (see Supplementary Fig. S1). We then applied DBSCAN with the following parameters: $\epsilon = 0.40$, the neighborhood radius and $Z = 200$, the minimum number of points required to form a dense region. This configuration yielded nine clusters, from which we selected the five largest (C1–5, descending by size) and labeled the remaining events as 'unclustered'. Their spatial distribution is shown in Fig. 1 and the data provided as Supplementary Data.

To enable a temporal analysis, each of the largest clusters C1–C3 was divided into three periods (see indicators in Fig. 2):

- '*early*': events before the Visso event;
- '*mid*': events from the Visso event until two days after the Norcia mainshock;
- '*late*': the rest.

Note that there is too few data in C4 and C5 to benefit from the division.

As illustrated in Supplementary Fig. S6, these periods are affected by short-term aftershock incompleteness [STAI, *Kagan* 2004; *Helmstetter et al.* 2006; *Hainzl* 2016; de *Arcangelis et al.* 2018, see also Supplementary Note 2.4]. Supplementary Fig. S6 makes use of equalized plot scales as suggested by *Agnew* [2015] and overlays the event density as suggested by W. Ellsworth (pers. comm., 2021). In this way, Supplementary Fig. S6 informs us about the STAI duration after each main event, leading us to exclude STAI by using a temporal subset of each period for C1, C2, and C3 (see indicators at the top of Fig. 2):

- '*pre-Visso*': like *early*, but excluding the first 0.8 days after the Amatrice event;
- '*pre-Norcia*': like *mid*, but excluding the first 0.6 days after the Visso event and 2.0 days after the Norcia mainshock;
- '*pre-Campotosto*': like *late*, but before the Campotosto event;
- '*post-Campotosto*': like *late*, but after the Campotosto event excluding the first 0.4 days.

4.3 Earthquake statistics

The clusters and their temporal subsets are investigated in terms of their MFD. To quantify MFD differences, we calculate the b-value as function of M_c. The b-value is determined using the bias-free maximum likelihood estimation of *Tinti and Mulargia* [1987] and *Marzocchi and Sandri* [2003] for sample sizes $N \geq 50$. The b-value requires an exponential distribution of the magnitude above M_c to be physically meaningful [*Marzocchi et al.* 2020]. To assess the exponentiality of the MFD, we apply the Lilliefors test [*Marzocchi et al.* 2020; *Herrmann and Marzocchi* 2021], using the implementation of *Herrmann and Marzocchi* [2020], and obtain a p-value as function of M_c, which expresses the probability to observe the MFD assuming that the exponential distribution is the underlying distribution. For a significance level of $\alpha = 0.1$, we derive the lowest magnitude level for which the MFD can be considered exponential, referred to as $M_c^{\text{Lilliefors}}$. We always refer to the b-value at $M_c^{\text{Lilliefors}}$.

As an alternative to quantify MFD differences, we use the two-sample Kolmogorov–Smirnov (KS) test and compare the MFD of clusters or their temporal subsets pairwise. For each pair, the largest $M_c^{\text{Lilliefors}}$ is used as lower magnitude cutoff. The KS test returns a p-value as a measure for the strength of evidence against the null hypothesis that the two MFDs come from the same parent distribution. We interpret a p-value < 0.05 as a statistically significant difference.

We do not explore whether the MFD can be characterized by a tapered GR distribution, and therefore neglect variations of the corner magnitude, e. g., due to released energy close to faults [*Spassiani and Marzocchi* 2021]. While the b-value correlates with the largest magnitude [*Marzocchi et al.* 2020], the KS test has reduced sensitivity toward the tails of the distributions. We assume that distinct b-values or significant p-values reflect differences or changes of the entire exponential part of the MFD.

4.4 Prior seismicity

The high-resolution catalog of *Tan et al.* [2021] only contains 15 events with $M_w \geq 1.5$ before the start of the sequence (i. e., before the Amatrice event). For a comparison of seismicity during the sequence with prior seismicity, we have initially considered HORUS [*Lolli et al.* 2020, `horus.bo.ingv.it`] as a temporally extensive catalog that provides M_w magnitudes. Those magnitudes were converted from M_L using a magnitude regression that differs from the conversion relation used in the CI2016 catalog of *Tan et al.* [2021]. In fact, a comparative MFD analysis for CI2016 seismicity shows that the b-value differs considerably between both catalogs (0.2 units at $M_c^{\text{Lilliefors}}$, see Supplementary Note 2.5 and Supplementary Fig. S9). Therefore, the two M_w scales are not consistent with each other. This inconsistency does not allow a reliable comparison of the b-value among these two catalogs (e. g., against a reference “background” b-value based on HORUS). We therefore did not consider HORUS data in our MFD analyses.

Data availability

All data generated or analyzed during this study are included in this article (and its supplementary information files).

Acknowledgments

This study was supported by the *‘Real-time Earthquake Risk Reduction for a Resilient Europe’* (RISE) project, funded by the European Union’s *Horizon 2020* research and innovation program

(Grant Agreement No. 821115). We thank the creators and maintainers of `plotly`'s Python graphing library (`www.plotly.com/python`).

Author contributions

M.H. performed the statistical analyses, created the figures, and wrote the manuscript. E.P. performed the clustering analysis and reviewed the manuscript. W.M. lead the project and reviewed the manuscript. All authors designed the study and discussed the results.

References

Agnew, D. C. (2015). "Equalized Plot Scales for Exploring Seismicity Data". In: *Seismological Research Letters* 86.5, pp. 1412–1423. DOI: 10.1785/0220150054.

Bachmann, C. E., S. Wiemer, B. P. Goertz-Allmann, and J. Wössner (2012). "Influence of pore-pressure on the event-size distribution of induced earthquakes". In: *Geophysical Research Letters* 39.9. DOI: 10.1029/2012GL051480.

Beroza, G. C., M. Segou, and S. Mostafa Mousavi (2021). "Machine learning and earthquake forecasting—next steps". In: *Nature Communications* 12.1, p. 4761. DOI: 10.1038/s41467-021-24952-6.

Chiaraluce, L. (2012). "Unravelling the complexity of Apenninic extensional fault systems: A review of the 2009 L'Aquila earthquake (Central Apennines, Italy)". In: *Journal of Structural Geology* 42, pp. 2–18. DOI: 10.1016/j.jsg.2012.06.007.

Chiaraluce, L. *et al.* (2017). "The 2016 Central Italy Seismic Sequence: A First Look at the Mainshocks, Aftershocks, and Source Models". In: *Seismological Research Letters* 88.3, pp. 757–771. DOI: 10.1785/0220160221.

Ciaccio, M. G. (2016). "Instrumental seismicity of the Amatrice earthquake epicentral area: A review". In: *Annals of Geophysics* 59.FASTTRACK5, pp. 1–8. DOI: 10.4401/ag-7283.

Dascher-Cousineau, K., T. Lay, and E. E. Brodsky (2021). "Reply to "Comment on 'Two Foreshock Sequences Post Gulia and Wiemer (2019)' by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky" by Laura Gulia and Stefan Wiemer". In: *Seismological Research Letters* 92.5, pp. 3259–3264. DOI: 10.1785/0220210059.

De Luca, G., M. Cattaneo, G. Monachesi, and A. Amato (2009). "Seismicity in central and Northern Apennines integrating the Italian national and regional networks". In: *Tectonophysics*. DOI: 10.1016/j.tecto.2008.11.032.

De Arcangelis, L., C. Godano, and E. Lippiello (2018). "The Overlap of Aftershock Coda Waves and Short-Term Postseismic Forecasting". In: *Journal of Geophysical Research: Solid Earth* 123.7, pp. 5661–5674. DOI: 10.1029/2018JB015518.

Dublanchet, P. (2020). "Stress-Dependent b Value Variations in a Heterogeneous Rate-and-State Fault Model". In: *Geophysical Research Letters*. DOI: 10.1029/2020GL087434.

Ellsworth, W. L. and F. Bulut (2018). "Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks". In: *Nature Geoscience* 0. DOI: 10.1038/s41561-018-0145-1.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press, pp. 226–231.

García-Hernández, R., L. D'Auria, J. Barrancos, G. D. Padilla, and N. M. Pérez (2021). "Multiscale Temporal and Spatial Estimation of the b-Value". In: *Seismological Research Letters* 92.6, pp. 3712–3724. DOI: 10.1785/0220200388.

Ghosh, A., A. V. Newman, A. M. Thomas, and G. T. Farmer (2008). "Interface locking along the subduction megathrust from b -value mapping near Nicoya Peninsula, Costa Rica". In: *Geophysical Research Letters* 35.1, p. L01301. DOI: 10.1029/2007GL031617.

Goebel, T. H., G. Kwiatek, T. W. Becker, E. E. Brodsky, and G. Dresen (2017). "What allows seismic events to grow big?: Insights from b-value and fault roughness analysis in laboratory stick-slip experiments". In: *Geology* 45.9, pp. 815–818. DOI: 10.1130/G39147.1.

Gomberg, J. (2018). "Unsettled earthquake nucleation". In: *Nature Geoscience* 11.7, pp. 463–464. DOI: 10.1038/s41561-018-0149-x.

Grünthal, G., R. Wahlström, and D. Stromeyer (2009). "The unified catalogue of earthquakes in central, northern, and northwestern Europe (CENEC)—updated and expanded to the last millennium". In: *Journal of Seismology* 13.4, pp. 517–541. DOI: 10.1007/s10950-008-9144-9.

Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). "The Effect of a Mainshock on the Size Distribution of the Aftershocks". In: *Geophysical Research Letters* 45.24, pp. 13, 277–13, 287. DOI: 10.1029/2018GL080619.

Gulia, L., T. Tormann, S. Wiemer, M. Herrmann, and S. Seif (2016). "Short-term probabilistic earthquake risk assessment considering time-dependent b values". In: *Geophysical Research Letters* 43.3, pp. 1100–1108. DOI: 10.1002/2015GL066686.

Gulia, L. and S. Wiemer (2019). "Real-time discrimination of earthquake foreshocks and aftershocks". In: *Nature* 574.7777, pp. 193–199. DOI: 10.1038/s41586-019-1606-4.

Gulia, L. and S. Wiemer (2021). "Comment on "Two Foreshock Sequences Post Gulia and Wiemer (2019)" by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky". In: *Seismological Research Letters* 92.5, pp. 3251–3258. DOI: 10.1785/0220200428.

Hainzl, S. (2016). "Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs". In: *Journal of Geophysical Research: Solid Earth* 121.9, pp. 6499–6509. DOI: 10.1002/2016JB013319.

Hainzl, S. and T. Fischer (2002). "Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia". In: *Journal of Geophysical Research: Solid Earth* 107.B12, ESE 5–1–ESE 5–9. DOI: 10.1029/2002JB001865.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). "Comparison of short-term and time-dependent earthquake forecast models for southern California". In: *Bulletin of the Seismological Society of America* 96.1, pp. 90–106. DOI: 10.1785/0120050067.

Herrmann, M. and W. Marzocchi (2020). *Mc-Lilliefors: A completeness magnitude that complies with the exponential-like Gutenberg-Richter relation*. DOI: 10.5281/zenodo.4162497.

Herrmann, M. and W. Marzocchi (2021). "Inconsistencies and Lurking Pitfalls in the Magnitude–Frequency Distribution of High-Resolution Earthquake Catalogs". In: *Seismological Research Letters* 92.2A, pp. 909–922. DOI: 10.1785/0220200337.

Igonin, N., M. Zecevic, and D. W. Eaton (2018). "Bilinear Magnitude-Frequency Distributions and Characteristic Earthquakes During Hydraulic Fracturing". In: *Geophysical Research Letters*. DOI: 10.1029/2018GL079746.

Improta, L. *et al.* (2019). "Multi-segment rupture of the 2016 Amatrice-Visso-Norcia seismic sequence (central Italy) constrained by the first high-quality catalog of Early Aftershocks". In: *Scientific Reports* 9.1, p. 6921. DOI: 10.1038/s41598-019-43393-2.

El-Isa, Z. H. and D. W. Eaton (2014). "Spatiotemporal variations in the b-value of earthquake magnitude-frequency distributions: Classification and causes". In: *Tectonophysics* 615-616, pp. 1–11. DOI: 10.1016/j.tecto.2013.12.001.

Jordan, T. H. *et al.* (2011). “Operational earthquake forecasting: State of knowledge and guidelines for utilization”. In: *Annals of Geophysics* 54.4. DOI: 10.4401/ag-5350.

Kagan, Y. Y. (2002). “Seismic moment distribution revisited: I. Statistical results”. In: *Geophysical Journal International* 148.3, pp. 520–541. DOI: 10.1046/j.1365-246x.2002.01594.x.

Kagan, Y. Y. (2004). “Short-Term Properties of Earthquake Catalogs and Models of Earthquake Source”. In: *Bulletin of the Seismological Society of America* 94.4, pp. 1207–1228. DOI: 10.1785/012003098.

Lolli, B., D. Randazzo, G. Vannucci, and P. Gasperini (2020). “The Homogenized Instrumental Seismic Catalog (HORUS) of Italy from 1960 to Present”. In: *Seismological Research Letters* 91.6, pp. 3208–3222. DOI: 10.1785/0220200148.

Marzocchi, W. and L. Sandri (2003). “A review and new insights on the estimation of the b-value and its uncertainty”. In: *Annals of geophysics* 46.December, pp. 1271–1282. DOI: 10.4401/ag-3472.

Marzocchi, W., I. Spassiani, A. Stallone, and M. Taroni (2020). “How to be fooled searching for significant variations of the b-value”. In: *Geophysical Journal International* 220.3, pp. 1845–1856. DOI: 10.1093/gji/ggz541.

Michele, M., L. Chiaraluce, R. Di Stefano, and F. Waldhauser (2020). “Fine-Scale Structure of the 2016–2017 Central Italy Seismic Sequence From Data Recorded at the Italian National Network”. In: *Journal of Geophysical Research: Solid Earth* 125.4. DOI: 10.1029/2019JB018440.

Michele, M. *et al.* (2016). “The Amatrice 2016 seismic sequence: a preliminary look at the mainshock and aftershocks distribution”. In: *Annals of Geophysics* 59.Fast Track 5, pp. 1–8. DOI: 10.4401/ag-7227.

Mogi, K. (1962). “Magnitude-Frequency Relation for Elastic Shocks Accompanying Fractures of Various Materials and Some Related Problems in Earthquakes (2nd Paper)”. In: *Bulletin of the Earthquake Research Institute* 40.4, pp. 831–853. DOI: 10.15083/0000033759.

Montuori, C. A., M. Murru, and G. Falcone (2016). “Spatial variation of the b-value observed for the periods preceding and following the 24 August 2016, Amatrice earthquake (ML6.0) (central Italy)”. In: *Annals of Geophysics* 59.FastTrack5. DOI: 10.4401/ag-7273.

Moschella, M., M. G. Ciaccio, and D. Latorre (2021). “Minor earthquake sequences in the Amatrice-Norcia epicentral area (Central Italy)”. In: *Tectonophysics* 809.September 2020, p. 228858. DOI: 10.1016/j.tecto.2021.228858.

Nanjo, K. Z., N. Hirata, K. Obara, and K. Kasahara (2012). “Decade-scale decrease in b value prior to the M 9-class 2011 Tohoku and 2004 Sumatra quakes”. In: *Geophysical Research Letters* 39.20, 2012GL052997. DOI: 10.1029/2012GL052997.

Nanjo, K. Z. and A. Yoshida (2017). “Anomalous decrease in relatively large shocks and increase in the p and b values preceding the April 16, 2016, M7.3 earthquake in Kumamoto, Japan”. In: *Earth, Planets and Space* 69.1, p. 13. DOI: 10.1186/s40623-017-0598-2.

Nava, F. A., V. H. Márquez-Ramírez, F. R. Zúñiga, L. Ávila-Barrientos, and C. B. Quinteros (2016). *Gutenberg–Richter b-value maximum likelihood estimation and sample size*. DOI: 10.1007/s10950-016-9589-1.

Passarelli, L. *et al.* (2015). “Aseismic transient driving the swarm-like seismic sequence in the Pollino range, Southern Italy”. In: *Geophysical Journal International* 201.3, pp. 1553–1567. DOI: 10.1093/gji/ggv111.

Petruccelli, A., D. Schorlemmer, T. Tormann, A. P. Rinaldi, S. Wiemer, P. Gasperini, and G. Vannucci (2019). “The influence of faulting style on the size-distribution of global earthquakes”. In: *Earth and Planetary Science Letters* 527, p. 115791. DOI: 10.1016/j.epsl.2019.115791.

Porreca, M. *et al.* (2020). “3D geological reconstruction of the M. Vettore seismogenic fault system (Central Apennines, Italy): Cross-cutting relationship with the M. Sibillini thrust”. In: *Journal of Structural Geology* 131.March 2019, p. 103938. DOI: 10.1016/j.jsg.2019.103938.

Roberts, N. S., A. F. Bell, and I. G. Main (2016). “Mode switching in volcanic seismicity: El Hierro 2011-2013”. In: *Geophysical Research Letters* 43.9, pp. 4288–4296. DOI: 10.1002/2016GL068809.

Scholz, C. H. (2015). “On the stress dependence of the earthquake b value”. In: *Geophysical Research Letters* 42.5, pp. 1399–1402. DOI: 10.1002/2014GL062863.

Schorlemmer, D., S. Wiemer, and M. Wyss (2005). “Variations in earthquake-size distribution across different stress regimes.” In: *Nature* 437.7058, pp. 539–42. DOI: 10.1038/nature04094.

Schorlemmer, D. *et al.* (2018). “The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities”. In: *Seismological Research Letters* 89.4, pp. 1305–1313. DOI: 10.1785/0220180053.

Shelly, D. R., W. L. Ellsworth, and D. P. Hill (2016). “Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm”. In: *Journal of Geophysical Research: Solid Earth* 121.3, pp. 1776–1795. DOI: 10.1002/2015JB012719.

Sobiesiak, M., U. Meyer, S. Schmidt, H.-J. Götze, and C. M. Krawczyk (2007). “Asperity generating upper crustal sources revealed by b value and isostatic residual anomaly grids in the area of Antofagasta, Chile”. In: *Journal of Geophysical Research* 112.B12, B12308. DOI: 10.1029/2006JB004796.

Spassiani, I. and W. Marzocchi (2021). “An energy-dependent earthquake moment–frequency distribution”. In: *Bulletin of the Seismological Society of America* 111.2, pp. 762–774. DOI: 10.1785/0120200190.

Stallone, A. and W. Marzocchi (2019). “Empirical evaluation of the magnitude-independence assumption”. In: *Geophysical Journal International* 216.2, pp. 820–839. DOI: 10.1093/gji/ggy459.

Suyehiro, S., T. Asada, and M. Ohtake (1964). “Foreshocks and Aftershocks Accompanying a Perceptible Earthquake in Central Japan”. In: *Papers in Meteorology and Geophysics* 15.1, pp. 71–88. DOI: 10.2467/mripapers1950.15.1_71.

Tan, Y. J. *et al.* (2021). “Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence”. In: *The Seismic Record* 1.1, pp. 11–19. DOI: 10.1785/0320210001.

Taroni, M., J. Zhuang, and W. Marzocchi (2021). “High-Definition Mapping of the Gutenberg–Richter b-Value and Its Relevance: A Case Study in Italy”. In: *Seismological Research Letters* 92.6, pp. 3778–3784. DOI: 10.1785/0220210017.

Tinti, S. and F. Mulargia (1987). “Confidence intervals of b values for grouped magnitudes”. In: *Bulletin of the Seismological Society of America* 77.6, pp. 2125–2134. URL: https://pubs.geoscienceworld.org/ssa/bssa/article/77/6/2125/119039.

Tondi, E., D. Jablonská, T. Volatili, M. Michele, S. Mazzoli, and P. P. Pierantoni (2020). “The Campotosto linkage fault zone between the 2009 and 2016 seismic sequences of central Italy: Implications for seismic hazard analysis”. In: *GSA Bulletin* 133.7, pp. 1679–1694. DOI: 10.1130/B35788.1.

Tormann, T., B. Enescu, J. Woessner, and S. Wiemer (2015). “Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake”. In: *Nature Geoscience* 8.2, pp. 152–158. DOI: 10.1038/ngeo2343.

Tormann, T., S. Wiemer, S. Metzger, A. J. Michael, and J. L. Hardebeck (2013). “Size distribution of Parkfield’s microearthquakes reflects changes in surface creep rate”. In: *Geophysical Journal International* 193.3, pp. 1474–1478. DOI: 10.1093/gji/ggt093.

Van der Elst, N. J. and B. E. Shaw (2015). “Larger aftershocks happen farther away: Nonseparability of magnitude and spatial distributions of aftershocks”. In: *Geophysical Research Letters* 42.14, pp. 5771–5778. DOI: 10.1002/2015GL064734.

Vuan, A., M. Sugan, L. Chiaraluce, and R. Di Stefano (2017). “Loading Rate Variations Along a Midcrustal Shear Zone Preceding the M w 6.0 Earthquake of 24 August 2016 in Central Italy”. In: *Geophysical Research Letters* 44.24, pp. 12, 170–12, 180. DOI: 10.1002/2017GL076223.

Waldhauser, F., M. Michele, L. Chiaraluce, R. Di Stefano, and D. P. Schaff (2021). “Fault Planes, Fault Zone Structure and Detachment Fragmentation Resolved With High-Precision Aftershock Locations of the 2016–2017 Central Italy Sequence”. In: *Geophysical Research Letters* 48.16, pp. 1–10. DOI: 10.1029/2021GL092918.

Wiemer, S. and M. Wyss (1997). “Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times?” In: *Journal of Geophysical Research: Solid Earth* 102.B7, pp. 15115–15128. doi: 10.1029/97jb00726.

Wiemer, S. and M. Wyss (2002). “Mapping spatial variability of the frequency-magnitude distribution of earthquakes”. In: *Advances in Geophysics*. Ed. by R. Dmowska and B. Saltzman. Vol. 45. Elsevier, pp. 259–302. doi: 10.1016/S0065-2687(02)80007-3.

Wyss, M. (1973). “Towards a Physical Understanding of the Earthquake Frequency Distribution”. In: *Geophysical Journal International* 31.4, pp. 341–359. doi: 10.1111/j.1365-246X.1973.tb06506.x.

Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2010). “The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science”. In: *Concurrency and Computation: Practice and Experience* 22.12, pp. 1836–1847. doi: 10.1002/cpe.1519.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- datasuppTan2021extractMw1.5withClusterID.zip
- supplementclean.pdf

New physical implications from revisiting foreshock activity in southern California

Ester Manganiello[1], Marcus Herrmann[1], and Warner Marzocchi[1]

[1]Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli 'Federico II', Naples, Italy

Corresponding author: E. Manganiello, ester.manganiello@unina.it

Key Points:

- We compare the foreshock activity in southern California with the prediction of the best-performing earthquake clustering model.
- Sequences with an anomalous excess of foreshocks are associated mostly with moderate mainshocks and preferentially with high heat flow.
- The prevalence of anomalous foreshock sequences in zones of high heat flow does not support the pre-slip nucleation model.

Abstract

Foreshock analysis promises new insights into the earthquake nucleation process and could potentially improve earthquake forecasting. Well-performing clustering models like the Epidemic-Type Aftershock Sequence (ETAS) model assume that foreshocks and general seismicity are generated by the same physical process, implying that foreshocks can be identified only in retrospect. However, several studies have recently found higher foreshock activity than predicted by ETAS. Here, we revisit the foreshock activity in southern California using different statistical methods and find anomalous foreshock sequences, i.e., those unexplained by ETAS, mostly for mainshock magnitudes below 5.5. The spatial distribution of these anomalies reveals a preferential occurrence in zones of high heat flow, which are known to host swarm-like seismicity. Outside these regions, the foreshocks generally behave as expected by ETAS. These findings show that anomalous foreshock sequences in southern California do not indicate a pre-slip nucleation process, but swarm-like behavior driven by heat flow.

Plain Language Summary

Many studies have observed that large earthquakes are preceded by smaller events, called foreshocks. If they have distinctive characteristics that make them recognizable in an ongoing sequence in real time, they can significantly improve the forecasting capability of large earthquakes. To investigate the nature of foreshocks, we compare real seismicity with the expectation of the most skilled earthquake clustering model, which assumes that foreshocks do not have any distinctive characteristics with respect to general seismicity. We find that discrepancies between reality and expectation mostly affect foreshock sequences that anticipate moderate mainshocks with magnitudes below 5.5. We show that those anomalous foreshock sequences tend to occur where the heat flow is high, which are already known for the occurrence of swarm-like sequences. Outside these regions, the observed foreshock activity is explained well by the clustering model. These findings indicate that anomalous foreshock sequences are not diagnostic of impending large earthquakes but are influenced by the heat flow.

1 Introduction

It is well known that many large earthquakes are preceded by smaller events (e.g., 1999 M7.6 Izmit, Turkey (Bouchon et al., 2011; Ellsworth & Bulut, 2018), 2009 M6.1 L'Aquila, Italy (Chiaraluce et al., 2011), 2011 M9.0 Tohoku, Japan (Kato et al., 2012), 2019 M7.1 Ridgecrest, USA (Meng & Fan, 2021)), which are (a posteriori) called foreshocks. The role of foreshocks in earthquake predictability can be epitomized by two still debated conceptual hypotheses about earthquake nucleation: the "pre-slip model" versus the "cascade model" (Ellsworth & Beroza, 1995; Gomberg, 2018). According to the former, foreshocks are diagnostic precursors, because they are triggered by an aseismic slip that precedes large earthquakes; in the latter model, foreshocks are like any other earthquake, which trigger one another, with one of them eventually becoming exceedingly larger (the mainshock).

Notwithstanding the still active debate on these hypotheses, seismologists are not yet able to recognize foreshocks in real-time, tacitly implying that foreshocks are not different from the rest of seismicity, indirectly supporting the cascade model. This view is further supported by the fact that the current best performing short-term earthquake forecasting model (Taroni et al., 2018)—the Epidemic-Type Aftershock Sequences (ETAS; Ogata, 1988) model—assumes that foreshocks, mainshocks, and aftershocks are undistinguishable and governed by the same process. ETAS belongs to the class of branching point process models known in the statistical literature as Hawkes or self-exciting point processes: every earthquake can trigger other earthquakes according to established empirical relations, with their magnitudes being independent from past seismicity. In essence, ETAS implicitly acknowledges the cascade model and its good forecasting performance makes ETAS an appropriate null hypothesis.

Instead, if foreshocks are dominated by mechanisms other than earthquake triggering, as the pre-slip model expects, they could be distinguished from general seismicity and potentially increase the probability for a larger earthquake to follow. Several studies recently investigated foreshock sequences of southern California and found that they deviate from expectations of the classical ETAS model with spatially invariant parameters. For example, Seif et al. (2019), Petrillo and Lippiello (2021), and Moutote et al. (2021) find, albeit at varying degrees, a higher foreshock activity in real seismicity than in synthetic catalogs simulated with ETAS. Hence, ETAS appears to be unable to predict all the observed seismicity, which may suggest that foreshocks are distinct from general seismicity and governed by different mechanisms. These findings provide hope that foreshocks are distinguishable and could pave the way to significantly improved earthquake predictability.

Here we reexamine foreshock activity in southern California and investigate the existence and main characteristics of foreshock sequences that cannot be explained by ETAS, i.e., anomalous foreshock sequences. In other words, we look for evidence against the cascade model. To make the results comparable to previous analyses, we use an ETAS model with spatially invariant triggering parameters. We perform two different statistical tests and consider the potential influence of subjective choices, such as the method to identify mainshocks and their foreshocks. To fathom the main characteristics of possible anomalous foreshock sequences, we investigate different magnitude classes and analyze the spatial correlation with heat flow as a physical parameter. With our findings, we aim to contribute to improving earthquake forecasting and the understanding of earthquake nucleation processes.

2 Data and Methods

We use the relocated earthquake catalog for southern California catalog (Hauksson et al., 2012, see Data Availability Statement), selecting all earthquakes with $M \geq 2.5$ from 1-1-1981 to 31-12-2019 except nuclear events (i.e., at the Nevada Test site) from the catalog, totaling 47'574 events.

Because there is no absolute and precise procedure to identify mainshocks, foreshocks, and aftershocks, the way of analyzing a catalog and distinguishing these events is unavoidably subjective (Molchan & Dmitrieva, 1992; Zaliapin et al., 2008). To mitigate this subjective choice, we analyze the catalog using two quite different techniques: the Nearest-Neighbor (NN) clustering analysis proposed by Baiesi and Paczuski (2004) and elaborated by Zaliapin et al. (2008), and the spatiotemporal windows (STW) method (Agnew and Jones, 1991; Marzocchi and Zhuang, 2011; Seif et al., 2019).

The NN method operates in a space-time-magnitude domain based on the NN distance η_j, i.e., the space-time-magnitude distance between event j and all earlier events i that is minimal. The event i with the shortest distance to event j is called NN, or parent, event. By assigning a parent event to each event j, all events become associated with another. To identify individual families (i.e., sequences) or single events, we use the same threshold $\eta_0 = 10^{-5}$ as Zaliapin et al.

(2008), which effectively removes event associations with too large η_j. For each sequence, we refer to the event with the largest magnitude as the mainshock and all associated events that occur before it as its foreshocks. We only consider sequences with foreshocks and ignore those that have no foreshocks.

For the STW method, we initially consider all events with magnitude $M \geq 4$ as possible mainshocks. Then, we exclude events that are (i) preceded by a larger event within a spatiotemporal window of 10 km and 3 days; (ii) preceded by an event with $M > 5$ within 100 km and 180 days; and (iii) not preceded by at least one event within 10 km and 3 days. For the remaining mainshocks, all preceding events within a window of 10 km and 3 days are considered foreshocks.

To simulate synthetic catalogs, we use the ETAS model of K. Felzer (Felzer et al., 2002, see Data Availability Statement and supporting information Text S1 and Table S1) with spatially invariant triggering parameters given by Hardebeck et al. (2008, see Table S2). Using an available ETAS model reduces potential influences from subjective parameter choices. We verify its overall reliability by comparing the number of events in the real catalog with the distribution of simulated events in the synthetic catalogs (see Text S2 and Figures S1 and S2), finding that the ETAS model is consistent with the observation.

Once the mainshocks and their foreshocks have been identified in both the real and 1000 synthetic catalogs, we compare their foreshock statistics using two approaches named TEST1 and TEST2. The two tests are described in detail below; both use the cascade model, which is implied by ETAS, as null hypothesis but emphasize different aspects of the problem. TEST1 involves the average number of observed foreshocks per sequence, whereas TEST2, which has been inspired by the work of Seif et al. (2019), involves the frequency of observing a certain number of foreshocks per sequence. We apply both tests to various mainshock magnitude classes $C_{\mathrm{M}} = \{4.0 \leq m_{\mathrm{M}} < 4.5, 4.5 \leq m_{\mathrm{M}} < 5.0, 5.0 \leq m_{\mathrm{M}} < 5.5, 5.5 \leq m_{\mathrm{M}} < 6.0, m_{\mathrm{M}} \geq 6.0\}$ and foreshock magnitude thresholds $T_{\mathrm{F}} = \{m_{\mathrm{F}} \geq 2.5, m_{\mathrm{F}} \geq 3.0, m_{\mathrm{F}} \geq 3.5, m_{\mathrm{F}} \geq 4.0\}$; these choices are based on Seif et al. (2019), but we add the class $4.0 \leq m_{\mathrm{M}} < 4.5$ to C_{M}. Although we report statistical test results, we do not formally account for applying the tests multiple times; the results are therefore meant to indicate possible patterns of (apparently) anomalous foreshock activity.

In TEST1, the null hypothesis under test $H_0^{(1)}$ is that the average number of foreshocks in the real catalog is not larger than the corresponding quantity in the synthetic catalogs. For each mainshock magnitude class $\mathrm{c} \in C_{\mathrm{M}}$ and each foreshock magnitude threshold $\mathrm{t} \in T_{\mathrm{F}}$, we count the number of mainshocks (with foreshocks), $N_{\mathrm{M}}^{\mathrm{real}}$, and the number of foreshocks $N_{\mathrm{F}}^{\mathrm{real}}$ in the real catalog; $N_{\mathrm{F}}^{\mathrm{real}}$ is normalized by $N_{\mathrm{M}}^{\mathrm{real}}$ to obtain $\widehat{N}_{\mathrm{F}}^{\mathrm{real}}$. We calculate the same quantity for each synthetic catalog and build its empirical cumulative distribution function (eCDF); if $\widehat{N}_{\mathrm{F}}^{\mathrm{real}}$ is above the 99th percentile of the eCDF, we reject $H_0^{(1)}$ at a significance level of 0.01.

In TEST2, the null hypothesis under test $H_0^{(2)}$ is that for each number of foreshocks, $N_{\mathrm{F}} > 0$, the frequency of observed cases is not larger than the frequency in synthetic catalogs. For each $c \in C_{\mathrm{M}}$ and each $t \in T_{\mathrm{F}}$, we count the number of mainshocks that have a certain N_{F} and normalize it by $N_{\mathrm{M}}^{\mathrm{real}}$. In this way, we obtain the probability mass function (PMF) for the real catalog as a function of N_{F}. Then, we apply the same procedure to each synthetic catalog and obtain 1000 synthetic PMFs, for which we calculate the 99th percentile at each N_F. Finally, at each N_F, we reject $H_0^{(2)}$ at a significance level of 0.01 if the corresponding PMF value of the real catalog is larger than the 99th percentile (i.e., when the real catalog contains more foreshock sequences with this specific N_F than expected by ETAS). In essence, TEST2 seeks anomalies at every N_F, whereas TEST1 could be seen as a cumulative version of TEST2.

Based on the results of the tests, we can label each foreshock sequence as 'anomalous' or 'normal' using an intuitive approach: for TEST1, if the null hypothesis is rejected for a certain class, all foreshock sequences with a N_F larger than the 99th percentile of the eCDF in that class are labeled as 'anomalous' (and 'normal' otherwise); for TEST2, if the null hypothesis is rejected for a specific N_F, all sequences with this N_F are labeled as 'anomalous' (and 'normal' otherwise). Effectively, a foreshock sequence in $c \in C_\mathrm{M}$ is labeled 'anomalous' if it is 'anomalous' in at least one class $t \in T_\mathrm{F}$. For TEST1, we argue that the approach is conservative, because comparing a single sequence against the average behavior of foreshock sequences may lead to wrongly label more actual normal foreshock sequences as 'anomalous' (i.e., false positives) than wrongly labeling anomalous foreshock sequences as 'normal' (i.e., false negatives). To investigate this aspect, we perform an alternative analysis by building two eCDFs of N_F (i.e., without normalizing by N_M): one for the real catalog (eCDF$^\mathrm{real}$) and one for all synthetic catalogs combined (eCDF$^\mathrm{ETAS}$). If the 99th percentile of eCDF$^\mathrm{real}$ is larger than the corresponding percentile of eCDF$^\mathrm{ETAS}$ in a certain class, we label each foreshock sequence as 'anomalous' whose N_F is above the 99th percentile of eCDF$^\mathrm{ETAS}$.

To investigate the physical interpretation of possible anomalous foreshock sequences in the real catalog, we analyze their spatial distribution. Specifically, taking inspiration from Zaliapin and Ben-Zion (2013), we create a map by interpolating heat flow measurements (see Data Availability Statement) with a radial smoothing approach ($r = 20$ km) to acknowledge areas without data. For each foreshock sequence, we extract the interpolated heat flow value closest to the mainshock location if it is within *r,* otherwise we discard the sequence. Then we test if the distribution of extracted heat flow values is significantly different for normal and anomalous foreshock sequences. If pre-slip is responsible for anomalous foreshock sequences, we should not find any difference, i.e., a spatial pattern. We employ two statistical tests: the two-sample Kolmogorov-Smirnov test (null hypothesis: the two distributions have the same parent distribution), and the paired Wilcoxon test (null hypothesis: the two distributions have the same median). In essence, the Kolmogorov-Smirnov test is sensitive to any kind of difference between both distributions, whereas the Wilcoxon test is sensitive to one distribution having higher values than the other.

3 Results

3.1 Testing for anomalous foreshock activity

Figure 1 shows the results of TEST1 using NN to identify mainshocks and their foreshocks; the results using STW are reported in supporting information Figure S3. Each subplot shows a comparison of the eCDF based on synthetic catalogs with the observed value from the real catalog for each class in C_M and T_F. As shown in Figure 1 and Figure S3, TEST1 rejects $H_0^{(1)}$, i.e., identifies anomalous foreshock sequences, exclusively for mainshock magnitudes $m_\mathrm{M} < 5.5$. Of a total of 152 foreshock sequences, we find 61 (40%) to be anomalous; with the STW method we find 143 foreshock sequences of which 34 (23%) are anomalous (all with $m_\mathrm{M} <$

5.5.). Using instead the alternative analysis without normalizing by N_{M} (Figure S4), we find 19 (12.5%) to be anomalous, which suggests that TEST1 overestimated the number of anomalies due to using averages, as anticipated in Data and Methods.

Figure 2 shows the results of TEST2 for each class in C_{M} and T_{F} using the NN method; the results using the STW method are reported in supporting information Figure S5. Most PMF values of the real catalog are not anomalous because they are below the 99th percentile of synthetic PMF values. We find 21 of 152 (14%) foreshock sequences to be anomalous, most of which are again associated with m_{M} < 5.5 (only three have larger m_{M}) . Using the STW method we find 10 of 143 (7%) foreshock sequences to be anomalous.

For comparison, Figure 2 also reports the results obtained by applying the approach of Seif et al. (2019), which tests a similar yet different null hypothesis than TEST2. Specifically, they treat all synthetic catalogs as one single compound catalog. In this way, the PMF is normalized with a much larger number of mainshocks than a single catalog (e.g., like the real catalog); for an increasing number of synthetic catalogs, the PMF decreases progressively observation (i.e., lowering the detectable minimum frequency) and moves further away from the real. In other words, our TEST2 honors the fact that a finite earthquake catalog must have a lower detectable frequency of foreshocks in the PMF; this lower frequency depends on the number of mainshocks that have foreshocks, which in turn depends on the length of the earthquake catalog (the lowest frequency is 1 out of the number of mainshocks that have foreshocks). In addition, the approach of Seif et al. (2019) normalizes the PMF by the total number of mainshocks that have foreshocks (N_{M}, as we do in TEST2) *and* no foreshocks, which further reduces the PMF by another 0.5–1 order of magnitude depending on $c \in C_{\mathrm{M}}$.

We repeated TEST1 and TEST2 at a 0.05 significance level (i.e., 95th percentile), which was originally used by Seif et al. (2019), see supporting information (Text S3 and Figures S6 and S7).

3.2 Correlating foreshock sequences with the heat flow

To investigate the physical cause of anomalous foreshock sequences we inspect the correlation of their locations with the local heat flow. We choose this property because previous papers suggested that the heat flow relates to statistical properties of seismic sequences (e.g., Enescu et al., 2009, Chen & Shearer, 2016; Ross et al., 2021; Zaliapin & Ben-Zion, 2013).

Figures 3a and 4a overlay the locations of normal and anomalous foreshock sequences identified by TEST1 and TEST2, respectively, on a heat flow map. Figures 3b and 4b show the corresponding eCDFs of the interpolated heat flow observed at the locations of normal and anomalous foreshock sequences. In both cases, anomalous foreshock sequences tend to occur more frequently at locations of higher heat flow than normal sequences. This trend is confirmed by the *p*-values of the two-sample Kolmogorov-Smirnov and paired Wilcoxon tests (see annotations in Figures 3b and 4b), which are below 0.05, indicating that the two samples come from different parent distributions with different means. Figures 3 and 4 are based on the NN method to identify mainshocks and their foreshocks; the results based on the STW method confirm our findings (see supporting information Figures S8 and S9), as do the results based on a 0.05 significance level (Figures S10 and S11). Moreover, TEST1-based results are stable even if we use the alternative procedure to identify anomalous foreshock sequences using eCDFs without normalizing by N_{M} (see Figure S12).

We verify the stability of our results using foreshock anomalies identified by Petrillo and Lippiello (2021). The authors provided us locations of their identified normal and anomalous foreshock sequences (G. Petrillo, pers. comm., 2022), letting us apply our analysis on a dataset that is completely independent from our assumptions and modeling choices. The results shown in Figure S13 confirm our findings of a preferential occurrence of foreshock anomalies in zones of high heat flow.

Finally, we add a word of caution on the interpretation of the results, that is, the spatial coverage of heat flow data compared to the earthquake activity is rather incomplete in northern Mexico. For instance, several anomalous foreshock sequences occur in this area but cannot be included in the heat flow analysis due to the lack of heat flow measurements. In addition, the available heat flow measurements in northern Mexico are not consistent with the Geothermal map of North America (Blackwell & Richards, 2004), which indicates a generally high heat flow (> 100 $\mu W/m^2$) in this area along the San Andreas Fault.

4 Discussion & Conclusion

We have found that foreshocks have the same characteristics of general seismicity as expected by ETAS, except for some cases. Our finding is in general agreement with previous studies of foreshock activity, all of which found (with some important differences not discussed here) higher foreshock activity than expected (Chen & Shearer, 2016; Moutote et al., 2021; Petrillo & Lippiello, 2021; Seif et al., 2019). However, our results additionally show that foreshock anomalies are mostly associated with mainshock magnitudes below 5.5—independently from the two tests and the two procedures to identify mainshocks and their foreshocks. Moreover, these anomalies are located preferentially (and statistically significant) in zones of high heat flow. The combination of these two findings suggests that sequences with anomalous foreshock activity behave more like seismic swarms. In fact, independent studies (e.g., Enescu et al., 2009, Chen & Shearer, 2016; Ross et al., 2021; Zaliapin & Ben-Zion, 2013) have shown that swarm-like activity is common in those areas where we have found anomalous foreshock sequences.

Our results do not allow us to further elucidate why foreshock anomalies correlate with high heat flow. The anomalies may be driven by specific physical mechanisms (e.g., actual seismic swarms mostly driven by fluids) or still relate to a cascade model that is not spatially uniform. The latter may be better described by an ETAS model with spatially varying triggering parameters. In fact, Enescu et al. (2009) and Nandan et al. (2017) show that some parameters of a spatially varying ETAS model (which mostly depend on the more abundant aftershocks) correlate with the heat flow in southern California. Such a more elaborated clustering model implies more active foreshock sequences where the heat flow is high, which agrees with our empirical findings based on the analysis of (less abundant) foreshocks.

Conversely, foreshock sequences located in zones of lower heat flow predominantly behave as expected, i.e., in agreement with the null hypothesis given by the ETAS model. Since it is reasonable to assume that a pre-slip model should not be severely affected by heat flow, our results do not indicate the pre-slip model as a major candidate to explain the anomalous foreshock behavior in southern California. It goes without saying that our results do not prove the cascade model as the truth, but that they do not bring any evidence against it and in favor of the pre-slip model.

Our results also highlight the importance of analyzing seismic sequences in zones of high heat flow in more detail, especially to understand the physical reasons of anomalous foreshock sequences: Are they related to seismic swarms with an implicit limitation to the mainshock magnitude? Or are they related to different clustering processes than those driving tectonic sequences? The difference is crucial, in particular regarding the forecasting of large earthquakes.

Our findings raise an urgent need to find (quasi-)real-time methods to discriminate swarm-like from ETAS-like sequences. Such a discrimination method could lead to significant improvements in earthquake forecasting, because being able to identify a swarm-like sequence as such could markedly reduce the forecast probability for a large earthquake. We note that an interesting attempt in this direction has been made by Zaliapin and Ben-Zion (2013), who found that swarm-like sequences have a different topologic tree structure (i.e., an internal clustering hierarchy, which connects background and triggered earthquakes). Unfortunately, this method can currently only be used retrospectively, limiting its applicability in earthquake forecasting. We envision other possible parameterizations of the topologic tree structure that may facilitate its use from a forecasting perspective.

Acknowledgments

We thank Luc Moutote and an anonymous reviewer for their very useful suggestions, which significantly helped improving the article. We also thank G. Petrillo for providing us with independent data. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115, *Real-Time Earthquake Risk Reduction for a Resilient Europe* (RISE).

Open Research

The southern California catalog of Hauksson et al. (2012) was obtained from https://scedc.caltech.edu/data/alt-2011-dd-hauksson-yang-shearer.html, version “1981–2019” (last accessed April 2021). Heat flow data were obtained from the following sources: National Geothermal Data System (http://geothermal.smu.edu/static/DownloadFilesButtonPage.htm, last accessed May 2021) using the data sets ‘*Aggregated Well Data*’, ‘*Heat Flow Observation in Content Model Format*’, ‘*SMU Heat Flow Database of Equilibrium Log Data and Geothermal Wells*’, and ‘*SMU Heat Flow Database from BHT Data*’; and RE Data Explorer (https://www.re-explorer.org/re-data-explorer/download/rede-data, last accessed May 2021) for northern Mexico. The ETAS simulator of K. Felzer was obtained from https://web.archive.org/web/20200712004939/https://pasadena.wr.usgs.gov/office/kfelzer/AftSimulator.html, last accessed February 2022). The alternative dataset of normal and anomalous foreshock locations was provided by G. Petrillo (pers. comm., 2022). The methods to perform our foreshocks analyses are available as MATLAB code at DOI: 10.5281/zenodo.6376221 and https://gitlab.com/ester.manganiello/foreshock-analyses.

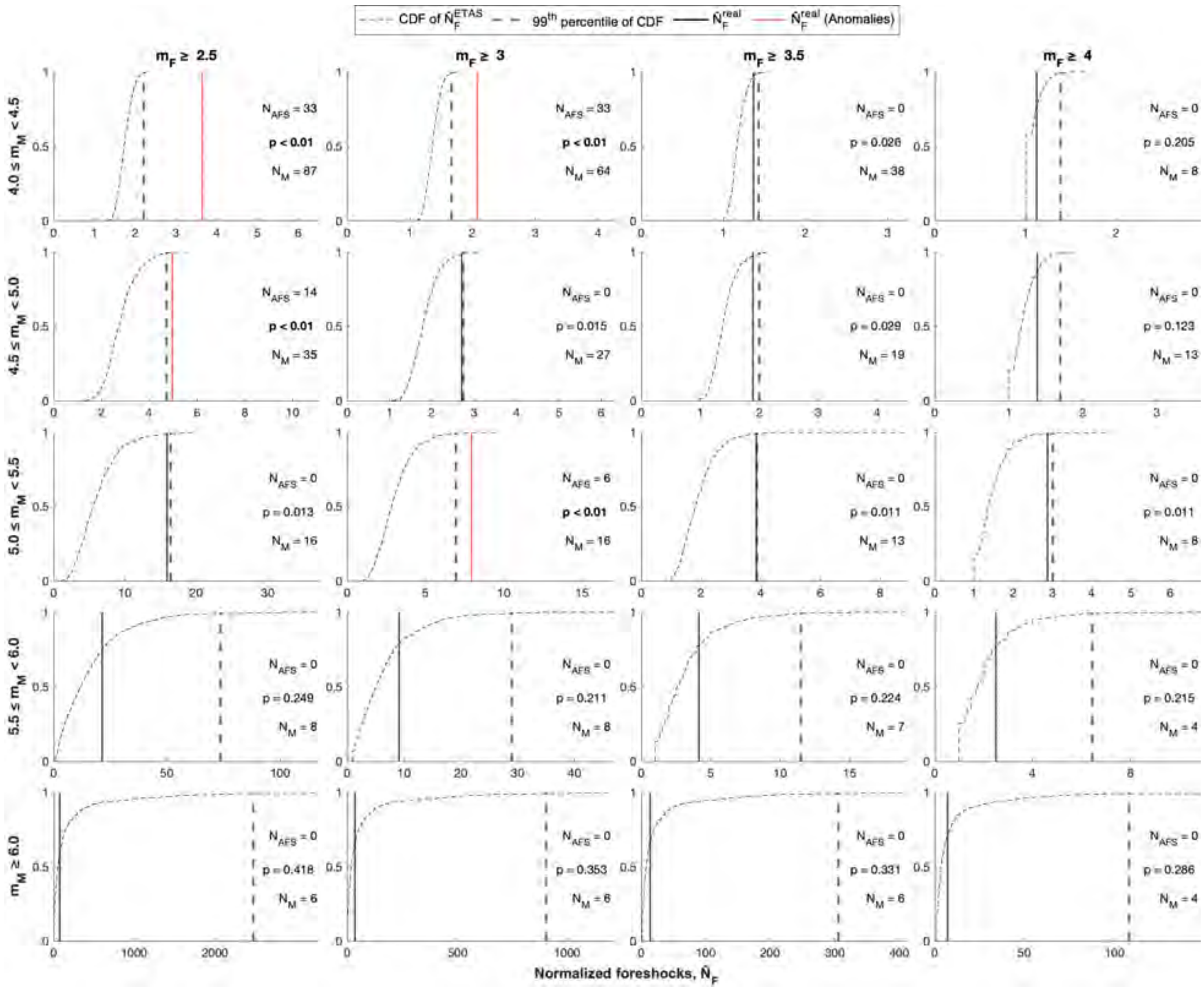


Figure 1. Results of TEST1 for various classes of the mainshock magnitude m_{M} (rows) and thresholds for the foreshock magnitude m_{F} (columns). Each subplot displays the number of normalized foreshocks $\widehat{N}_{\mathrm{F}}$ for the real catalog (vertical lines; red if anomalous, black otherwise) and the empirical Cumulative Distribution Functions (eCDFs, dashed curves) with its 99th percentile (dashed vertical lines) for 1000 synthetic catalogs. Each subplot also reports the number of anomalous foreshock sequences, N_{AFS}, the *p*-value for TEST1, and the number of mainshocks, N_{M}. The results are based on the NN method; supporting information Figure S3 shows results based on the STW method. Note that each subplot uses a different N_{F}-axis range to account for the varying data range.

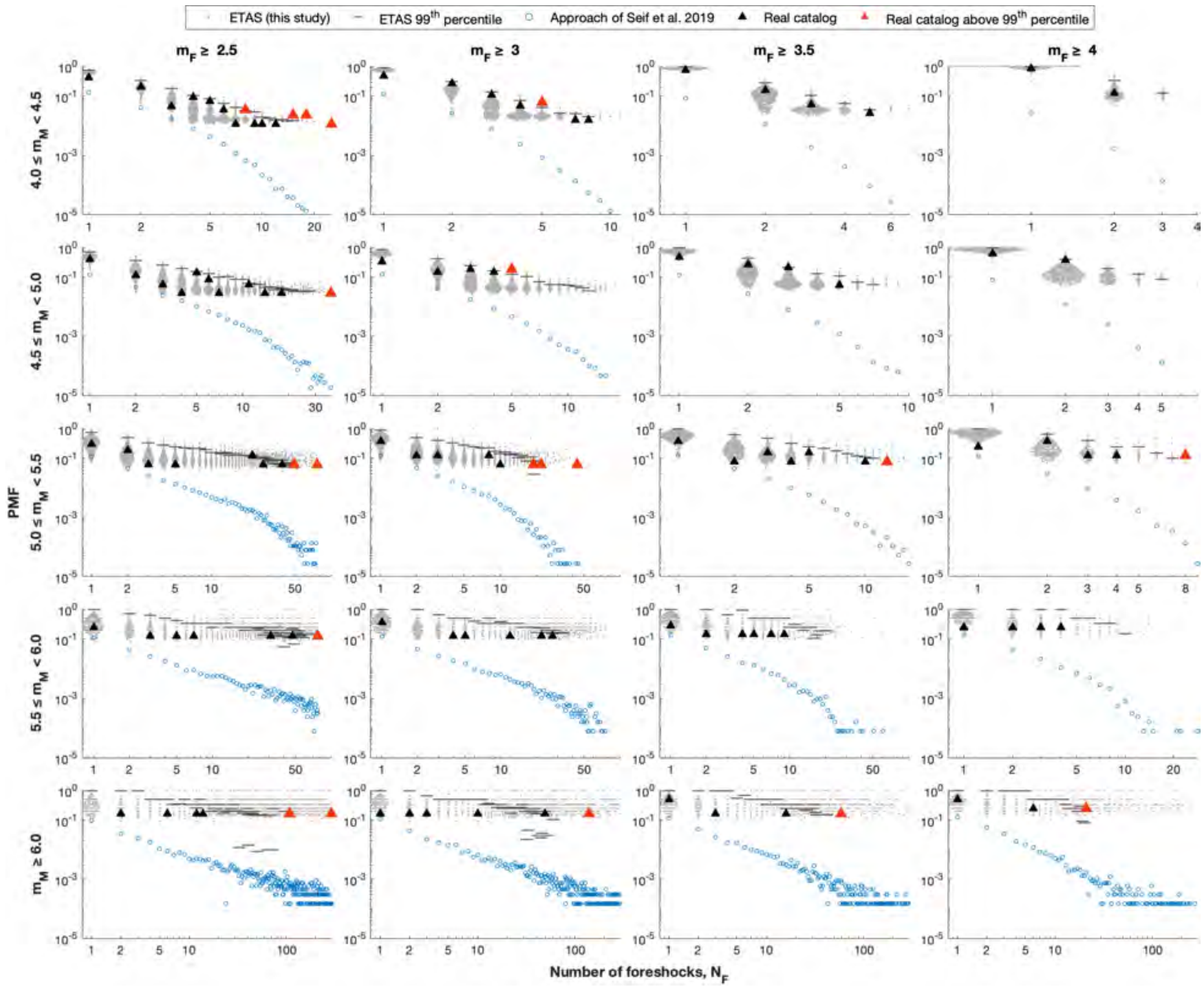


Figure 2. Results of TEST2 showing probability mass functions (PMFs) of the number of foreshocks N_{F} for various classes of m_{M} (rows) and m_{F} (columns). The PMFs are shown for (i) the real catalog (triangles), (ii) all synthetic catalogs (small gray dots as swarm distributions) with their 99$^{\text{th}}$ percentile (gray horizontal bars), and (iii) when considering all synthetic catalogs as a single compound catalog (blue open circles, using the approach of Seif et al., 2019). Triangles become red when they are located above the 99$^{\text{th}}$ percentile of (ii). The results are based on the NN method to identify mainshocks and their foreshocks; supporting information Figure S5 shows results based on the STW method. Note that each subplot uses a different N_{F}-axis range.

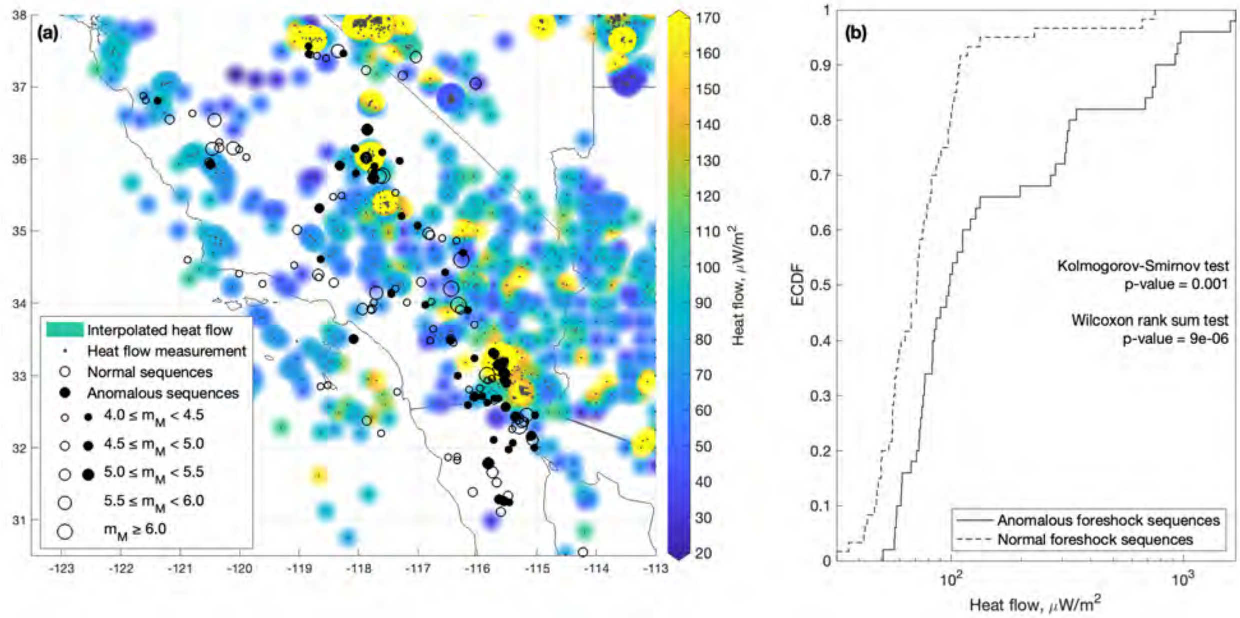


Figure 3. Correlating foreshock sequences with the heat flow. (a) Locations of normal (empty circles) and anomalous foreshock sequences (filled circles) identified with TEST1 overlayed on a heat flow map. The circles sizes scales with m_{M} (see legend). The interpolated heat flow map is based on sampled heat flow measurements (small gray dots, see Data and Methods section); (b) eCDFs of heat flow values at locations of normal (dashed curve) and anomalous foreshock sequences (solid curve); both eCDFs are compared using two statistical tests (see annotation with corresponding *p*-values). The results are based on the NN method; supporting information Figure S8 shows results based on the STW method.

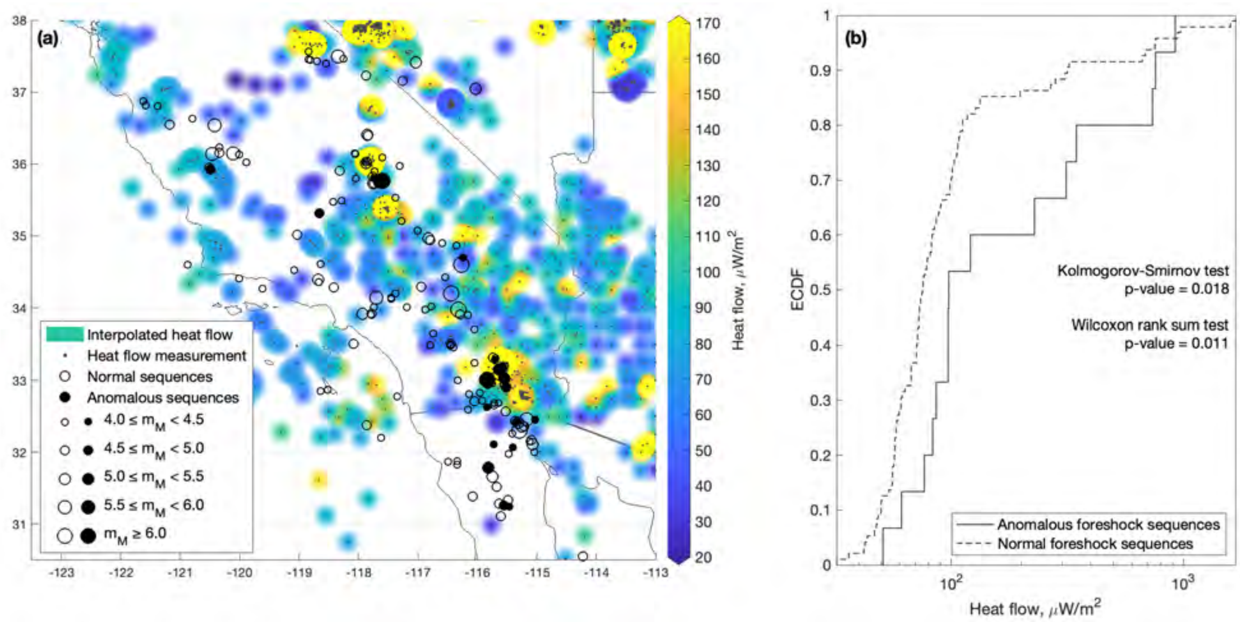


Figure 4. Like Figure 3 but with foreshock sequences labeled as ‘anomalous’ or ‘normal’ using TEST2. Supporting information Figure S9 shows results based on the STW method.

References

Agnew, D. C., & Jones, L. M. (1991). Prediction probabilities from foreshocks. *Journal of Geophysical Research, 96*(B7), 11959. https://doi.org/10.1029/91JB00191

Baiesi, M., & Paczuski M. (2004). Scale-free networks of earthquakes and aftershocks. *Physical Review E*, *69*, 066106. https://doi.org/10.1103/PhysRevE.69.066106.

Blackwell, D. D., & Richards, M. (2004). Geothermal Map of North America, AAPG Map, scale 1:6,500,000, Product Code 423. url: https://www.smu.edu/-/media/Site/Dedman/Academics/Programs/Geothermal-Lab/Graphics/Geothermal_MapNA_7x10in.gif

Bouchon, M., Karabulut, H., Aktar, M., Özalaybey, S., Schmittbuhl, J., & Bouin, M. P. (2011). Extended Nucleation of the 1999 Mw 7.6 Izmit Earthquake. *Science*, *331*(6019), 877–880. https://doi.org/10.1126/science.1197341

Chen, X. & Shearer, P. M. (2016). Analysis of Foreshock Sequences in California and Implications for Earthquake Triggering. *Pure Applied Geophysical*, *173*(1), 133–152. https://doi.org/10.1007/s00024-015-1103-0

Chiaraluce, L., Chiarabba, C., De Gori, P., Di Stefano, R., Improta, L., Piccinini, D., Schlagenhauf, A., Traversa, P., Valoroso, L., & Voisin, C. (2011). The 2009 L'Aquila (central Italy) seismic sequence. *Bollettino di Geofisica Teorica ed Applicata*, *52*(3), 367-387. https://doi.org/10.4430/bgta0019

Ellsworth, W. L., & Beroza, G. C. (1995). Seismic Evidence for an Earthquake Nucleation Phase. *Science*, *268*, 851-855. https://doi.org/10.1126/science.268.5212.851

Ellsworth, W. L., & Bulut, F. (2018). Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks. *Nature Geoscience*, *11*(7), 531–535. https://doi.org/10.1038/s41561-018-0145-1

Enescu, B., Hainzl, S., and Ben-Zion, Y. (2009). Correlations of Seismicity Patterns in Southern California with Surface Heat Flow Data. *Bulletin of the Seismological Society of America*, *99* (6), 3114-3123. https://doi.org/10.1785/0120080038

Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., & Rice, J. R. (2002). Triggering of the 1999 M_w 7.1 Hector Mine earthquake by aftershocks of the 1992 M_w 7.3 Landers earthquake. *Journal of Geophysical Research*, *107*(B9), 2190. https://doi.org/10.1029/2001JB000911

Gomberg, J. (2018). Unsettled earthquake nucleation. *Nature Geoscience*, 11(7), 463–464. https://doi.org/10.1038/s41561-018-0149-x

Hardebeck, J. L., Felzer, K. R., & Michael, A. J. (2008). Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *Journal of Geophysical Research*, *113*, 3B08310. https://doi.org/10.1029/2007JB005410

Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2019). *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244. https://doi.org/10.1785/0120120010

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of Slow Slip Leading Up to the 2011 Mw 9.0 Tohoku-Oki Earthquake. *Science*, *335*(6069), 705-708. https://doi.org/10.1126/science.1215141

Marzocchi, W., & Zhuang, J. (2011). Statistics between mainshocks and foreshocks in Italy and Southern California. *Geophysical Research Letters*, *38*(9), 2011GL047165. https://doi.org/10.1029/2011GL047165

Meng, H., & Fan, W. (2021). Immediate foreshocks indicating cascading rupture developments for 527 M 0.9 to 5.4 Ridgecrest earthquakes. *Geophysical Research Letters*, *48*, e2021GL095704. https://doi.org/10.1029/2021GL095704

Molchan, G. M., & Dmitrieva, O. E. (1992). Aftershock identification: methods and new approaches. *Geophysical Journal International*, *109*(3), 501–516. https://doi.org/10.1111/j.1365-246X.1992.tb00113.x

Moutote, L., Marsan, D., Lengliné, O., & Duputel, Z. (2021). Rare occurrences of non-cascading foreshock activity in Southern California. *Geophysical Research Letters*, *48*, e2020GL091757. https://doi.org/10.1029/2020GL091757

Nandan, S., Ouillon, G., Wiemer, S., and Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to California, *Geophysical Research Letters Solid Earth*, *122*, 5118–5143. https://doi.org/10.1002/2016JB013266

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association, 83 (401), 9-27. https://doi.org/10.1080/01621459.1988.10478560

Petrillo, G., & Lippiello, E. (2021). Testing of the foreshock hypothesis within an epidemic like description of seismicity. *Geophysical Journal International*, *225*, 1236–1257. https://doi.org/10.1093/gji/ggaa611

Ross, Z. E, Cochran, E. S., Trugman, D. T., & Smith, J. D. (2021). 3D fault architecture controls the dynamism of earthquake swarms. *Science*, *368*(6497), 1357-1361. https://doi.org/10.1126/science.abb0779

Seif, S., Zechar, J. D., Mignan, A., Nandan, S., & Wiemer, S. (2019). Foreshocks and Their Potential Deviation from General Seismicity. *Bulletin of the Seismological Society of America*, *109* (1), 1–18. https://doi.org/10.1785/0120170188

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective CSEP Evaluation of 1-Day, 3-Month, and 5-Yr Earthquake Forecasts for Italy. *Seismological Research Letters*, *89*, 1251-1261. https://doi.org/10.1785/0220180031

Zaliapin, I., & Ben-Zion, I. (2013). Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth*, *118*, 2847-2864. https://doi.org/10.1002/jgrb.50179

Zaliapin, I., & Ben-Zion, I. (2013). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust. *Journal of Geophysical Research: Solid Earth*, *118*, 2865-2877. http://doi.org/10.1002/jgrb.50178

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering Analysis of Seismicity and Aftershock Identification. *Physical Review Letters*, *101*, 018501. https://doi.org/10.1103/PhysRevLett.101.018501

AGU ADVANCING EARTH AND SPACE SCIENCE

JGR Solid Earth

RESEARCH ARTICLE
10.1029/2020JB020226

Key Points:
- Spatially varying seismicity can be efficiently modeled as a Log-Gaussian Cox process which include deterministic and stochastic components
- LGCPs can be analyzed with integrated nested Laplace approximations to compare seismicity models and identify useful model components
- These models find maps of strain rate and distance to nearest fault useful for constraining spatial seismicity

Supporting Information:
- Supporting Information S1

Correspondence to:
K. Bayliss,
Kirsty.Bayliss@ed.ac.uk

Citation:
Bayliss, K., Naylor, M., Illian, J., & Main, I. G. (2020). Data-driven optimization of seismicity models using diverse data sets: Generation, evaluation, and ranking using inlabru. *Journal of Geophysical Research: Solid Earth*, *125*, e2020JB020226. https://doi.org/10.1029/2020JB020226

Received 19 MAY 2020
Accepted 12 OCT 2020
Accepted article online 14 OCT 2020

Data-Driven Optimization of Seismicity Models Using Diverse Data Sets: Generation, Evaluation, and Ranking Using Inlabru

Kirsty Bayliss[1], Mark Naylor[1], Janine Illian[2], and Ian G. Main[1]

[1]School of GeoSciences, University of Edinburgh, Edinburgh, UK, [2]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

Abstract Recent developments in earthquake forecasting models have demonstrated the need for a robust method for identifying which model components are most beneficial to understanding spatial patterns of seismicity. Borrowing from ecology, we use Log-Gaussian Cox process models to describe the spatially varying intensity of earthquake locations. These models are constructed using elements which may influence earthquake locations, including the underlying fault map and past seismicity models, and a random field to account for any excess spatial variation that cannot be explained by deterministic model components. Comparing the alternative models allows the assessment of the performance of models of varying complexity composed of different components and therefore identifies which elements are most useful for describing the distribution of earthquake locations. We demonstrate the effectiveness of this approach using synthetic data and by making use of the earthquake and fault information available for California, including an application to the 2019 Ridgecrest sequence. We show the flexibility of this modeling approach and how it might be applied in areas where we do not have the same abundance of detailed information. We find results consistent with existing literature on the performance of past seismicity models that slip rates are beneficial for describing the spatial locations of larger magnitude events and that strain rate maps can constrain the spatial limits of seismicity in California. We also demonstrate that maps of distance to the nearest fault can benefit spatial models of seismicity, even those that also include the primary fault geometry used to construct them.

Plain Language Summary Recently, many statistical models for earthquake occurrence have been developed with the aim of improving earthquake forecasting. Several different underlying factors might control the location of earthquakes, but testing the significance of each of these factors with traditional approaches has not been straightforward and has restricted how well we can combine different successful model elements. We present a new approach using a point process model to map the spatial intensity of events. This method allows us to combine maps of factors which might affect the location of earthquakes with a random element that accounts for other spatial variation. This allows us to rapidly compare models with different components to see which are most helpful for describing the observed locations. We demonstrate this approach using synthetic data and real data from California as a whole and the 2019 Ridgecrest sequence in particular. Slip rates are found to be beneficial for explaining the spatial distribution of large magnitude events, and strain rates are found useful for constraining spatial limits of observed seismicity. Constructing a fault distance map can also improve models where many events cannot be directly linked to an individual fault.

1. Introduction

For a variety of reasons, including the lack of clear, reliable precursors and the inherent nonlinearity and complexity of the underlying process, the deterministic prediction of individual earthquakes remains an elusive goal (Jordan et al., 2011). Instead, the focus has shifted to forecasting the probability of occurrence of a population of events in space and time, in an attempt to determine the degree of predictability of earthquakes (Field, 2007; Jordan & Jones, 2010; Vere-Jones, 1995). In order to make reliable forecasts, it is necessary to understand as much as possible about the spatiotemporal behavior of earthquakes and the underlying processes that drive them.

Statistical point process models have been used to describe earthquake occurrence for many years (Ogata, 1998; Vere-Jones, 1970; Vere-Jones & Davies, 1966). The aim of these models is to describe the occurrence of earthquakes as a series of points in time, space, or both, with an appropriate “mark” such as earthquake magnitude. With the creation of the Epidemic-Type Aftershock Sequence (ETAS) model and development of robust methods to estimate parameter values (Ogata, 1988, 2011; Ogata & Zhuang, 2006; Veen & Schoenberg, 2008), point process models have been applied extensively in statistical seismology. For example, the ETAS model is widely used for catalog simulation and model testing (Helmstetter, 2003; Helmstetter & Sornette, 2002a, 2002b, 2003, Helmstetter et al., 2005; Nandan et al., 2017; Seif et al., 2017), as well as being used in aftershock forecasting models (Marzocchi et al., 2014; Rhoades et al., 2018; Taroni et al., 2018).

The uniform California earthquake rupture forecast (UCERF) is a fault-based model for forecasting seismicity in the state of California. The most recent implementations of this model (UCERF3) include a time-independent model that assumes that the process is statistically stationary and memoryless (Field et al., 2014), a long-term time-dependent model incorporating memory of large past events (Field et al., 2015), and a short-term time-dependent forecast model (Field et al., 2017) which makes use of the ETAS model to forecast aftershock activity. These models are used in parallel in hazard assessment for the state of California for different applications and consist of four main components: a fault model for the physical locations and architecture of known faults, slip rate models that estimate the slip and creep estimates for each individual fault from geodetic and geological data, event rate models which describe the long-term rate of earthquakes throughout the area, and a probability model that describes the likelihood of an earthquake occurring within a specified time period. The time-dependent models also include long-term renewal and short-term clustering processes, while the time-independent model is applied to probabilistic seismic hazard assessment (PSHA). Typically, different model configurations are selected by use of a logic tree, where each branch is given a weighting determined by expert judgement through workshops. This approach allows the construction of models containing different information which will then be included in a full rupture forecast model and allows the inclusion of uncertainties in model parameters. In the case of UCERF3, this results in a logic tree with a total of 5,760 branches, requiring the use of high-power computing facilities to produce the resulting forecast models.

The idea of hybrid forecasting models is not a new one. Marzocchi et al. (2012) suggest a Bayesian method for combining models and ranking them based on their respective Bayes factor. Rhoades et al. (2014) proposed multiplicative combinations of models from the RELM project to improve the forecasting ability of models compared to a smoothed seismicity model. Rhoades et al. (2015) expanded on this approach and applied it to events in New Zealand, in which they included a covariate that accounted for the distance to a mapped fault. In each of these examples there is a requirement for individual forecasts to be developed before combination and for the user to determine a weight for the individual model components, an issue which is highlighted particularly in Marzocchi et al. (2012). Nevertheless, there remains a significant component of epistemic uncertainty due to lack of data, notably on the occurrence rates of large magnitude events.

In this paper we address the question: How can we be sure which components of the model are most useful in describing the seismicity in a data-driven approach, that is, without recourse to expert judgement? Seismology currently lacks both a straightforward method for the objective combination of useful earthquake data and a robust framework for the rapid evaluation of seismicity models with a full description of the associated uncertainty. Here we present a possible solution—the application of integrated nested Laplace approximations (INLA) for the spatial modeling of earthquake data by modeling seismicity as a log-Gaussian Cox process (LGCP). Once the most useful model components are identified, these can be applied in a straightforward way to prospective forecasting models. We begin by outlining the theory underlying the INLA method and justifying its use in a seismological context. We then demonstrate the ease with which models describing the longer-term spatial distribution of seismicity can be constructed and compared using synthetic data and data used in the UCERF3 model. We highlight how INLA can be used to straightforwardly assess the contribution of different components of a model, how well each component can describe the observed spatial distribution of events, and how the inclusion of a random field in the analysis can help us identify what our model is lacking.

2. Theory

The earliest point process models for earthquakes were discussed by Vere-Jones (1970), who lamented that the mathematical methods required to evaluate these models did not exist at the time. Early models by Ogata (1988) demonstrated how point processes could be used for the temporal modeling of earthquake sequences, by combining the Poisson rate of independent or "background" earthquakes and the Omori law for dependent aftershock events to account for temporal clustering. Though the ETAS model has seen many improvements over the years, the question of the most suitable spatial model for a spatiotemporal ETAS model still remains. An isotropic inverse power law distribution is often used (Ogata & Zhuang, 2006; Werner et al., 2011) though more general spatial kernels were proposed by Ogata (1998) and sophisticated alternatives that include fault data and Coulomb stress changes have also been suggested (Bach & Hainzl, 2012). By stacking global data, Huc and Main (2003) showed that an inverse power law with an exponential tail was the most appropriate global average, implying a correlation length for triggering similar to the seismogenic thickness. An alternative point process approach would be to consider the spatial distribution of events first and then extend such a model to be fully spatiotemporal. Given that the spatial intensity of earthquake occurrence is known to vary and to be associated with underlying subsurface conditions which cannot be directly observed, a method that allows a quantitative description of the stochastic nature of earthquakes in space is required. Here we propose to solve this problem by adopting a LGCP which models a spatially varying intensity process as a function of deterministic and stochastic effects. Such models can be rapidly constructed and evaluated using INLA. Below, we outline the theory behind LGCPs and the INLA method for fitting them. INLA is a computationally efficient way to construct models, so we can construct models with different combinations of potentially useful components and compare their performance with appropriate methods.

2.1. LGCPs and INLA

LGCP are a popular class of model for spatial variability in ecology, as they allow some observed spatial pattern to be described by some deterministic location effects and a "random field" component which describes any remaining spatial variability. Observations can then be modeled with a spatially varying intensity function that describes a continuous stochastic process as a function of combined stochastic and deterministic effects (Diggle et al., 2013). Where current point process models for seismicity began by describing the temporal distribution of earthquakes, LGCPs aim to model the spatial variation of events using an inhomogeneous Poisson process, which can be extended to a fully spatiotemporal marked point process. In the case of earthquake data, our "deterministic" model components will inevitably still contain uncertainty and can range from observed data such as fault maps to spatial data models, such as smoothed seismicity or strain rate models.

The INLA approach is a computationally efficient alternative to Markov chain Monte Carlo (MCMC) approaches to Bayesian model fitting. The INLA method is incredibly flexible and has been widely applied to point process data sets in ecology (Dutra Silva et al., 2017; Illian et al., 2012; Sadykova et al., 2017; Yuan et al., 2017) and to some natural hazard examples including tornado (Gómez-Rubio et al., 2015), wildfire (Díaz-Avalos et al., 2016), and landslide modeling (Lombardo et al., 2018).

2.1.1. LGCP

Sometimes called the doubly stochastic Poisson process, the Cox process (Cox, 1955) is a generalization of the Poisson process where the process intensity λ is itself stochastic in nature. Vere-Jones (1970) proposed using the Cox process to describe the spatial distribution of earthquakes, if clustering were removed. A LGCP is a special case of the Cox process where the rate of events is determined by some underlying random field, which is assumed to be Gaussian in nature and to vary spatially. The construction of these models allows the user to specify underlying factors that might describe the observed spatial distribution, for example, how underlying soil characteristics might affect the location of a certain tree species (Illian et al., 2012). Multiple underlying factors can be included in the spatially varying intensity function to account for the observed spatial distribution of events. A random field can then also be added, which describes the remaining spatial structure which cannot be captured by more deterministic effects.

For a homogeneous Poisson process with intensity λ, the intensity can be modeled as

$$\lambda = e^{\beta_0} \tag{1}$$

for any positive intensity such that β_0 is an intercept term representing the mean of the intensity. For an inhomogeneous Poisson process with spatially varying intensity, we can also add a Gaussian random field

10.1029/2020JB020226

$\zeta(s)$ which accounts for spatial correlation in the observed points. This gives us

$$\lambda(s) = e^{\beta_0+\zeta(s)} \tag{2}$$

where we can consider the exponentiated term a linear predictor such that $\eta = \beta_0+\zeta(s)$. Again, β_0 represents a mean intensity, but the random field $\zeta(s)$ captures the fluctuations about β_0. A log Gaussian Cox process is then a model where the point intensity λ can be described as follows:

$$\lambda(s) = e^{\eta(s)} \tag{3}$$

The observed spatially varying intensity can also be modeled as

$$log(\lambda(s)) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m(s) + \zeta(s) \tag{4}$$

where β_m are linear covariates that may influence the spatially varying intensity of the points. These explain the observed spatial variation in the intensity of the process, with the spatially varying Gaussian random field $\zeta(s)$ accounting for the fluctuations in intensity that the deterministic covariates cannot fully explain. More complicated, nonlinear functions can also be included in the linear predictor. In this way, $\zeta(s)$ describes variation of point intensity that is not accounted for by other model components and therefore highlights the spatial areas in which the model components are not sufficient to describe observed spatial patterns of intensity.

2.1.2. INLA

To fit an LGCP model in a Bayesian manner, it is necessary to estimate the posterior distributions of the model parameters $\boldsymbol{\theta}$. Traditionally, MCMC methods may have been used; however, these are time-consuming to fit and prohibit the use of complex models. Where MCMC takes many samples from a posterior distribution, the INLA method (Rue et al., 2009) makes use of a series of approximations to estimate posterior distributions. Without the need for many iterations and the concerns of convergence associated with MCMC, INLA is a computationally efficient alternative for Bayesian analysis for models such as the LGCP where a latent Gaussian component is assumed. INLA is less generally applicable than MCMC because of the requirement for the latent Gaussian structure. Taylor and Diggle (2014) argue that an MCMC approach allows for a more flexible model and better analysis of joint posteriors than INLA. Teng et al. (2017) discuss different Bayesian approaches to the analysis of LGCP models, including different implementations of the INLA model, including the approach with stochastic partial differential equations (SPDEs) that we use in this paper. In this paper we have chosen to use INLA due to the speed and ease of application facilitated by recent software developments, allowing rapid model configuration and comparisons.

The basic idea is described in Rue et al. (2009). Some observed data x_i can be described by a parameter vector $\boldsymbol{\theta}$, and each element of $\boldsymbol{\theta}$ can be described by some hyperparameters $\boldsymbol{\psi} = \{\psi_i \dots \psi_k\}$ in a hierarchical Bayesian model. The INLA method aims to evaluate the marginal posteriors for each element of the parameter vector $\boldsymbol{\theta}$, which can be written as

$$p(\theta_i|x) = \int p(\theta_i, \psi|x)\mathrm{d}\psi = \int p(\theta_i|\psi, x)p(\psi|x)\mathrm{d}\psi \tag{5}$$

and for each element of the hyperparameter vector, $\boldsymbol{\psi}$, which can be written as

$$p(\psi_k|x) = \int p(\psi|x)\mathrm{d}\psi_{-k} \tag{6}$$

where $d\psi_{-k}$ is all other components of ψ except k.

It is necessary to calculate the joint posterior of the hyperparameters $p(\boldsymbol{\psi}|x)$ to calculate the posterior marginal distributions of the hyperparameters (Equation 6) and also to calculate $p(\theta_i|\boldsymbol{\psi}, x)$ in order to solve Equation 5 for the posteriors of each of the parameters θ_i. INLA does this by using Laplace approximations, nested because they are required for both $p(\boldsymbol{\psi}|x)$ and $p(\theta_i|\psi, x)$. INLA makes a Gaussian approximation of $p(\psi|x)$ which can be written $\tilde{p}(\psi|x)$ and a simplified Laplace approximation using a Taylor expansion of the approximation of $\tilde{p}(\theta_i|\psi, x)$. The main limitation of such an approach is the use of the Laplace approximation, which assumes that a smooth, peaked posterior distribution can be approximated by a Gaussian.

The posterior marginals can then be approximated with

$$p(\theta_i|x) \approx \int \tilde{p}(\theta_i|\psi, x)\tilde{p}(\psi|x)\mathrm{d}\psi \tag{7}$$

which can be solved numerically through a finite weighted sum. This is a suitable approximation when the posterior marginals are roughly Gaussian in nature but can also accommodate less Gaussian posteriors, as discussed at length in Rue et al. (2009) and Rue et al. (2017).

2.1.3. Gaussian Fields and Matérn Correlation

Gaussian fields are a useful mathematical concept that can be used to model underlying or latent processes. In the LGCP framework outlined here, a Gaussian field ζ is used to model spatial variation not accounted for by the deterministic model components, β_m, as in Equation 4. In this way, the Gaussian random field models the spatial structure by accounting for any spatial correlation between events. The combination of the random field and deterministic covariates models the intensity of the LGCP. We define the Gaussian field ζ as

$$\zeta(s) \sim GaussianField(0, \Sigma) \tag{8}$$

where the mean $= 0$ and the covariance is Σ. Calculating the covariance can be tricky, so instead of calculating all the values independently, a standard correlation function can be used to describe the correlation between points, and the area over which such correlation extends. A Matérn correlation function can be used to define the covariance such that

$$\Sigma = Cov_M = \sigma^2 Corr_M \tag{9}$$

where σ^2 is some variance parameter and Cov_M and $Corr_M$ are the Matérn covariance and correlation respectively. The Matérn correlation is specified as

$$Corr_M = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa\|s_i - s_j\|)^\nu K_\nu(\kappa\|s_i - s_j\|) \tag{10}$$

where s_i and s_j are the spatial positions of observations i and j and $\|s_i - s_j\|$ describes the Euclidean distance between these points. K_ν is a Bessel function, and the correlation has parameters κ and ν. Assuming that $\nu = 1$, the equation simplifies to have dependence only on κ and the distance between points. The Matérn correlation function describes the distances over which points within the model have some correlation, such that if the parameter κ is smaller, there is more long-range spatial dependency.

With this approach, the parameters required to describe the underlying Gaussian random field are simply σ^2 and κ. This will still be time-consuming to compute, unless we make the assumption that variation within the random field will only be on a local scale. If we can make the assumption that the underlying field is Markovian, such that only neighboring points will have nonzero correlation, the correlation matrix becomes sparse. Such an assumption approximates the random field with a Gaussian Markov random field (GMRF). Lindgren et al. (2011) provided an explicit link between Gaussian random fields and GMRFs that allows Gaussian random fields with Matérn covariance to be approximated by GMRFs even in cases where the spatial correlation structure is long range. The Matérn correlation structure is an extremely popular and flexible correlation structure used widely in a variety of spatial modeling applications (Guttorp & Gneiting, 2006).

2.1.4. SPDE and Mesh Construction

The INLA approach can be used with continuous domain random field models as described by Lindgren et al. (2011) and Simpson et al. (2012), leading to the application of the method to a range of complex spatial and spatiotemporal models (Blangiardo & Cameletti, 2013; Gómez-Rubio et al., 2015; Lindgren & Rue, 2015). Essentially, this requires defining a mesh on which to construct the point process model, such that the random field can be evaluated at each mesh vertex. Lindgren et al. (2011) detailed how random field models can be described by solutions of sets of SPDEs. The parameters of the SPDEs are directly linked to the parameters of the Matérn correlation, so solving the SPDEs gives the required parameters for the Matérn covariance of the underlying GMRF. The SPDEs can be solved using a finite element approach, where the area can be represented by a mesh, with basis function representations used to calculate the value at each mesh vertex. Thus, the SPDE approach allows the mapping of a random field from discrete points to a continuous field by the use of a correlation matrix that describes how the data points interact, or specifically the range of interaction of the points, which will be reflected in the resulting spatial intensity. This therefore allows us to consider a spatially continuous field rather than discrete point information, which in an earthquake context allows us to infer something about areas which have not experienced earthquakes within any point data set and not just the areas which have. A continuous field model also allows us to see where the model performs well in terms of the deterministic covariates describing the intensity, compared to areas where the intensity is mostly described by the random field component. Simpson et al. (2015) proved that the SPDE method converged well for LGCPs and with minimal error in the posterior distribution due to the method.

The mesh for the SPDE calculations is constructed using a restrained refined Delauney triangulation of a point data set using the inla.mesh.2d function. The mesh boundaries are determined by the extent of the point locations, with a coarser mesh extending slightly outside of this area to reduce edge effects. A more complex mesh may provide greater resolution but will require more computational power, so a compromise is required which will provide reasonable resolution at an acceptable cost. The mesh can be constructed from the point locations themselves, but this results in a finer mesh in areas of clustering and a coarser mesh in areas with few events, when the spatial structure of most interest is likely to be somewhere in between these two extremes. As such we have chosen to specify the mesh domain as the spatial extent of the points rather than construct the mesh on the points themselves. This also makes models on different time periods more comparable where the point locations are likely to be different. A further consideration when constructing the mesh is the range of the Matérn correlation describing the random field. The correlation range in the remaining spatial structure must be greater than the length of the mesh edges so that the resulting intensities are reliable.

We construct and run the following models using the r package inlabru (Bachl et al., 2019) to fit LGCP models to the observed points using INLA. The inlabru package provides a user-friendly approach for using INLA for point process models, building on the R-INLA package (Lindgren et al., 2011; Martins et al., 2013; Rue et al., 2009). inlabru makes use of the sp package (Bivand et al., 2013; Pebesma & Bivand, 2005), using spatial data frames to handle data, and R packages raster (Hijmans, 2019), rgdal (Bivand et al., 2019), and rgeos (Bivand & Rundel, 2019) are also used for data wrangling. All maps in this paper are constructed with the use of ggmap (Kahle & Wickham, 2013) and tmap (Tennekes, 2018), with color schemes from RColorBrewer (Neuwirth, 2014). The process for fitting LGCP models in inlabru is straightforward. A model is constructed for the random field component, based on a user-defined mesh. An equation describing the model components is defined, and an LGCP is fitted to this model. The LGCP fits can then be compared using DIC or by predicting the intensities that would be returned by the LGCP model. The predicted intensities can be a combination of all model components or include only some of the modelers choosing. In this way, the effect of adding different model components can be compared by studying changes to the predicted intensity of the model as a whole and also by looking at the random field alone, which will identify spatial variability that the deterministic components cannot explain.

2.2. Model Comparison and DIC

The deviance information criterion, or DIC, was developed by Spiegelhalter et al. (2002) as an alternative to the commonly used AIC. DIC is a measure that can be used to compare different models with varying numbers of parameters, designed as analogous to AIC for use with hierarchical models reflecting the trade-off between the "goodness of fit" and model "complexity." The DIC is the expected deviance, penalized by the effective number of model parameters, which is a measure of the difference in posterior mean deviance and the deviance of the posterior means of the individual parameters. This penalizes more complex models similar to AIC, therefore preferring the simplest models that can explain the resulting data. DIC is designed specifically for hierarchical models where the model is structured in such a way that there is structural dependence between parameters, as is the case when we include parameter priors. The DIC is calculated within the INLA and inlabru software (and similarly within other software used for Bayesian hierarchical modeling such as BUGS), making it a popular choice for model comparison in these contexts (Spiegelhalter et al., 2014).

In this case the DIC is calculated at the posterior mean of the latent field and the posterior mode of the hyperparameters. The deviance D is defined by

$$D(\theta) = -2\sum_{i} \log p(x_i|\theta) \tag{11}$$

The effective number of parameters pD is calculated using the trace of the prior precision matrix Q multiplied by the posterior covariance matrix Q^*

$$pD = n - tr\{Q(\theta)Q^*(\theta)^{-1}\} \tag{12}$$

where n is the number of observations and the total model DIC is then

$$DIC = \overline{D(\theta)} - pD \tag{13}$$

Rue et al. (2009) describe the full details of how DIC is calculated within the INLA software.

The Collaboratory for the study of earthquake predictability (CSEP, e.g. Michael & Werner, 2018) aims to improve earthquake forecasting by testing earthquake forecasts in real time, with the future aim of extending this approach to full hazard models (Schorlemmer et al., 2018). CSEP tests earthquake forecasts using a variety of different tests, ranging from simple tests of the number of forecast events (the N-test) to comparisons of forecast likelihoods (L-test) and residual comparisons, with different testing centers using different testing approaches for the models they are assessing. The DIC is also a likelihood-based model assessment tool, making it similar in some ways to the CSEP L-test, except in this case we are applying the tests to spatial data patterns rather than prospective or pseudo-prospective data at this time. We also compare the number of events expected by each of our constructed models—the abundance of points obtained by integrating over our mesh intensity models—which is somewhat analogous to the N-test used by CSEP. Nevertheless, the optimization presented here is a necessary but not sufficient criterion to developing a true prospective forecasting model. In future work we will develop pseudo-prospective and ultimately prospective forecast model tests to allow comparison with competing models in the CSEP framework.

Throughout this work, we use DIC as a tool for model comparison as a first test and highlight other methods of comparing models and model outputs. Our aim is not to discriminate between models based solely on their DIC, but the DIC and number of events predicted by the LGCP model make a useful first pass for testing and comparing models.

3. Data Types

A huge benefit of the inlabru approach is in the ability to combine different data types within a model. The earthquakes themselves are described as points, and the deterministic components of the model can be included as lines, polygons, maps, or raster images with discrete or continuous variables. Constructed LGCP models can include any combination of these components, and the output of one model can be straightforwardly included in the next. Continuous variables can be included with the use of a function that returns the value of the variable at a given point in space. Categorical information can be added for data which takes the form of discrete layers. In the inlabru terminology these are termed "factor" covariates, and we demonstrate their use in constructing a binary "fault factor" below. We begin by outlining the different data sets that can be included in an inlabru model with application to several data sets for California. Further information on each of the data types is included in the supporting information.

3.1. Earthquake Catalog: Spatiotemporal Point Data Set

We make use of various subsections of the UCERF3 catalog which consists of events above a fixed magnitude threshold. An LGCP model aims to model spatially varying intensity. The simplest possible LGCP model is one where we assume no known underlying spatial structure such that the intensity is a function of a Gaussian random field only. The smoothed seismicity model in Figure 1 is constructed in such a way, using a Matérn covariance for the random field. We see that the intensity model is behaving as we would expect, with high intensity in areas with a greater number of events.

3.2. Fault Maps: Polygon Data Set

Given that we know that the spatial distribution of observed earthquakes is related to underlying fault systems, we can include fault polygons in the model to see how well the fault locations account for the spatial distribution of events. The polygons are buffered as in the UCERF3 model. This is also the basis for our fault distance map which simply returns the distance to the nearest map fault for any point within the area of interest.

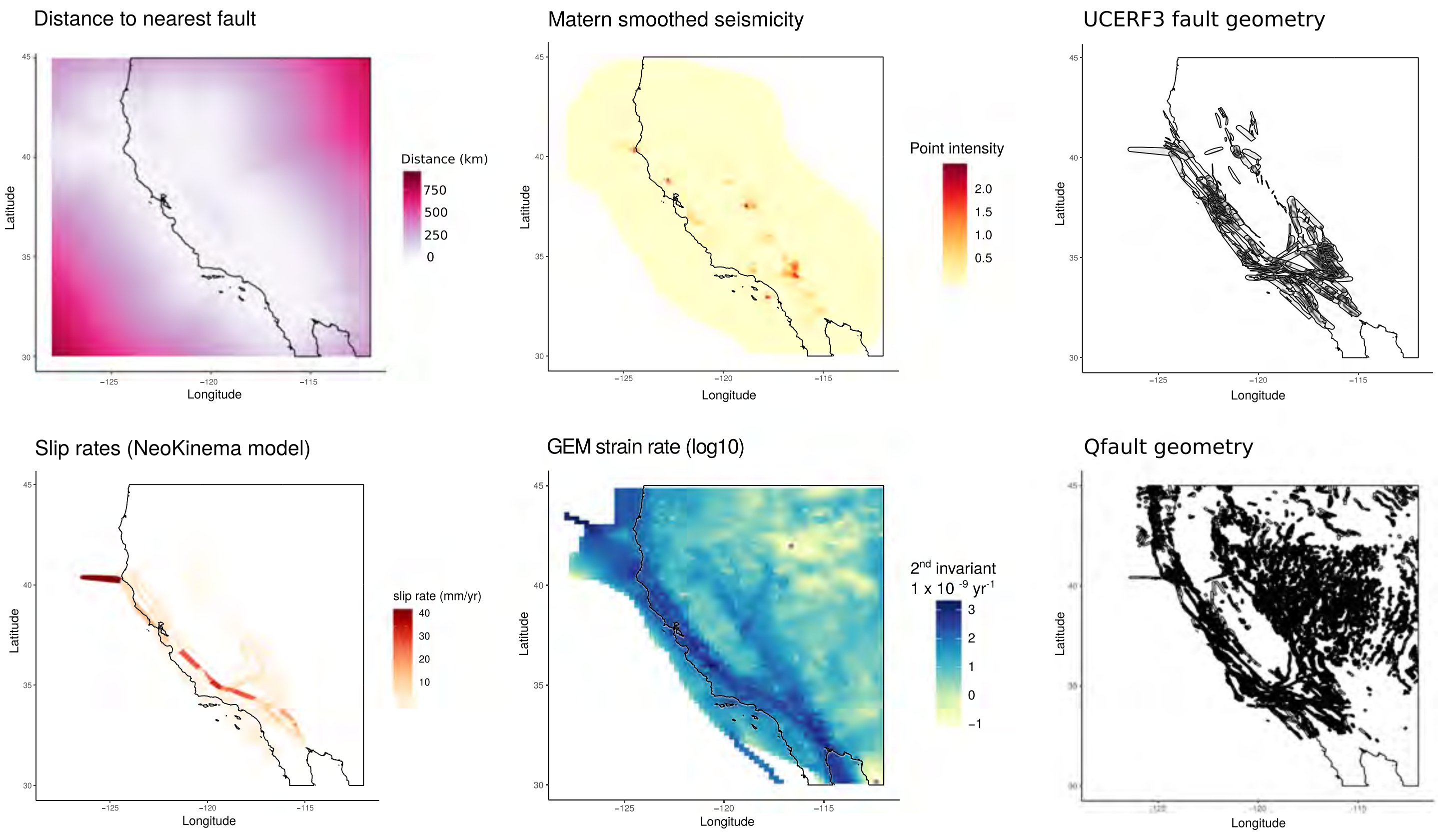


Figure 1. Input model components used within this work. Slip rates and fault geometry from UCERF3, Matérn smoothed seismicity created from UCERF3 $M2.5+$ data set, distance to nearest fault calculated from UCERF3 geometry. Strain rates from GEM strain rate model, where we have used the $\log_{10}$ of these values. Qfault geometry from the USGS Quaternary fault model, cropped around the study area.

Alternatively, we consider two further covariates related to fault geometry. Since a significant number of events within the catalog do not fall within the fault polygons, we can construct a fault distance map which shows the distance from the nearest fault. As an alternative to the UCERF3 fault geometry, we also make use of the USGS Quaternary faults model (https://earthquake.usgs.gov/hazards/qfaults/background.php), which we shall refer to as "QFaults." Figure 1 shows the different inputs for the models in this paper, including the two different fault geometries, the fault distance model and the other spatial components described below.

3.3. Slip Rate Data: Spatial Covariate or Fault Mark

There are four possible slip rate models considered within the UCERF3 logic tree. To work with these, we can construct either factor or continuous maps of slip rates for the different fault polygons, where a factor map requires discrete levels of data. In this paper, we use continuous slip rate values for each individual fault, such that the $\log_{10}$ slip rate is returned for any given point. Off-fault, the slip rate will be zero, so that we have essentially attached a value to each fault polygon only. Within the slip rate models, the value of the slip rate at any given point will be returned instead of the binary classification used to identify if a fault is present. The slip rates for the NeoKinema model used throughout this work are shown in Figure 1.

3.4. Past Seismicity: Continuous Spatial Covariate

We use a subset of the UCERF3 data to construct a Matérn-smoothed past seismicity model, by fitting an LGCP model to the point data alone and predicting the model intensity for events that occurred before our time period of interest. This allows us to use the observed past seismicity as a spatial covariate in future models, where we have smoothed the past seismicity by assuming that the intensity is a function of a random field only. We can therefore construct a past seismicity map using any subset of the data such that the past seismicity does not include any events in the model itself. A past seismicity model for all $M2.5+$ created using Matérn smoothing is shown in Figure 1.

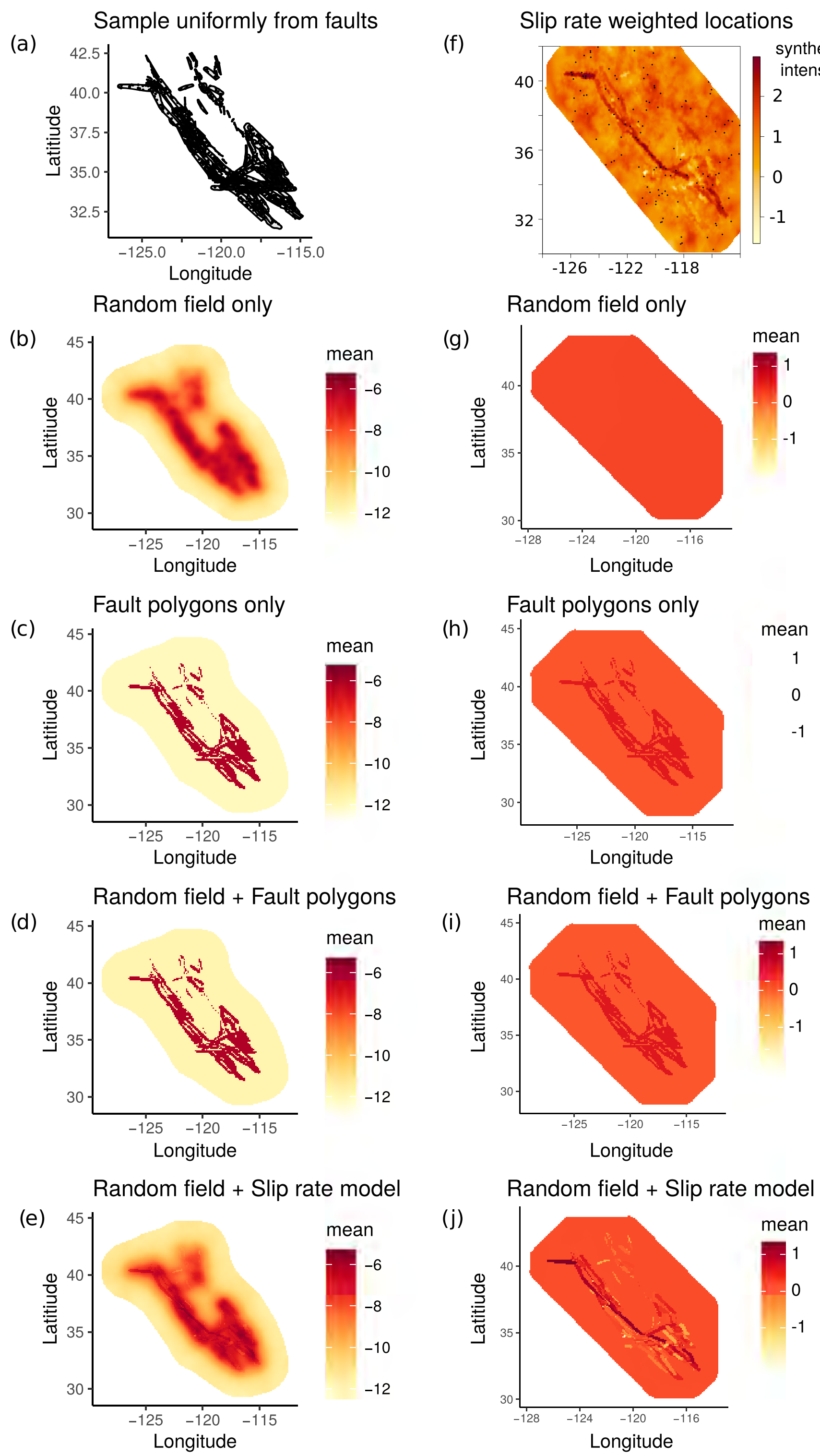


Figure 2. Inversion of two synthetic data sets: event locations randomly sampled from the full fault polygon set (a–e) and event locations sampled from a model of random field + fault slip rates according to the NeoKinema slip rate model (f–j). The top row shows the locations of synthetic events within the fault polygons. Plot (f) shows an example of the synthetic intensity model, where the fault polygons contribute to different intensities and the scale bar shows the log intensity of the synthetic model. Subsequent rows show predicted intensities for models of random field only (Row 2, b and g), models which only include fault polygons (Row 3, c and h), models with random field and fault polygons (Row 4, d and i), and models with fault slip rates attached to the fault polygons and with included random field (bottom row, e and j). The scale bar shows the log posterior mean intensity.

Table 1
DIC Estimates for Models of Synthetic Data Uniformly Sampled From Fault Polygons

Model	DIC	ΔDIC	Abundance
Random field + fault geometry	7,178	0	498 ± 21
Fault geometry	7,182	4	501 ± 23
Random field + slip rate	7,785	607	510 ± 23
Random field only	7,785	607	513 ± 23

3.5. Strain Rate: Continuous Spatial Covariate

We make use of the GEM strain rate model (Kreemer et al., 2014) which is a global strain rate model constructed with the use of deforming cells in areas of high strain. Since the UCERF3 data are not global and instead from a small area with a good catalog, the combined model of past seismicity and strain rate may perform less well than the past seismicity alone over the short timescales considered here, especially considering the resolution of the strain rate map. Over longer timescales the strain rate map may prove more useful by adding information. Given the abundance of model inputs we have access to in California, the strain rates may or may not be beneficial to the model, but the INLA method allows us to compare the effect of including a strain rate component with the addition of more detailed fault information. This allows us to assess if the strain rate might contribute meaningfully to models for areas which lack fault slip rates. The GEM strain rate model for California is shown in Figure 1.

4. Inversion of Synthetic Catalogs to Demonstrate the Method

To demonstrate the capabilities of the inlabru method, we construct two synthetic data sets based on the fault geometry of UCERF3. Using the R package sp (Bivand et al., 2013; Pebesma & Bivand, 2005), we randomly sample a chosen number of points from the fault polygons. The first model uniformly samples from within the fault polygons, while the second weights the number of points from each fault polygon by the slip rates. We set the number of randomly sampled events from polygons to 500, but for the slip rate weighted synthetic data the number of events varies and is much smaller. To properly assess the models, we construct 50 random data sets for each model. To each of these synthetic data sets, we then fit five models: a random field model, a model with fault polygons only, a model with fault polygons and random field, a fault model with slip rates, and a fault model with slip rates and random field.

4.1. Inversion of Events Randomly Distributed on Fault Network

Events are randomly sampled from the buffered fault polygons using the spsample function. Figure 2(a) shows one random catalog generated in this way and the resulting intensity predictions from the four constructed models (b-e). Table 1 shows the resulting model DIC values and predicted number of events from the LGCP fit as mean ± standard deviation for the data set in Figure 2. The fault geometry model significantly improves upon the random field only model by accounting for much of the spatial distribution, but because the distribution within the fault polygons is random, the random field + fault geometry model performs the best. The model with slip rates performs better than the random field alone, as the slip rates are related to the fault locations, but as the slip rates do not describe the observed pattern of events in space, the DIC of the slip rate + random field model is higher than that of the fault geometry model. If we compare the number of events predicted by each of the four models, all of the models predict a reasonable number of events, given that 500 synthetic events have been used. The density plot in Figure 3 shows that the performance of the different models varies significantly with different randomly sampled events. The model with fault geometry overlaps almost entirely with the fault geometry and random field model, while the slip rate and random field and random field only models also overlap significantly. The slip rates alone do poorly in all 50 randomly sampled catalogs.

4.2. Inversion of Events Randomly Distributed on Fault Network Weighted by Slip Rate

For our second synthetic model, we construct a Log-Gaussian Cox model where the intensity consists of a Gaussian random field and the fault slip rates, where the slip rate component is similar in magnitude

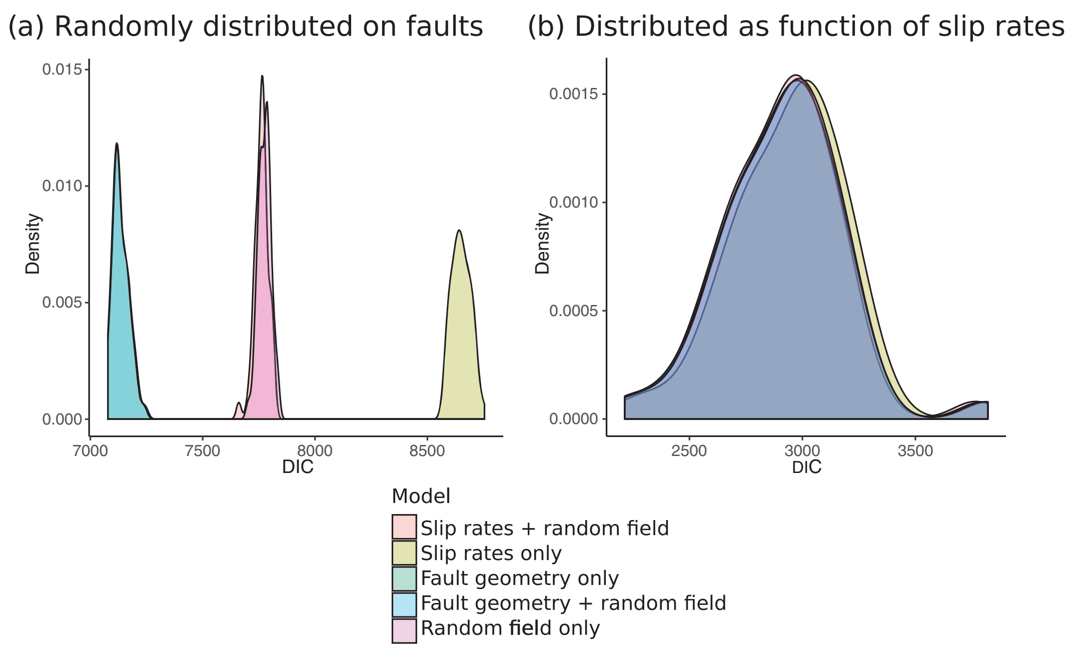


Figure 3. Densities of model DIC over 50 random samples from (a) the fault polygons and (b) weighted by the slip rates.

to the random field. This model is shown in Figure 2f. The points are sampled using a built-in inlabru function (sample.lgcp), which samples around 160 events in each realization, but the exact number varies. An example of the DIC results is shown in Table 2, with these results corresponding to the intensities shown in Figure 2. The intensity range is extended by the inclusion of the slip rate model (j), making the narrower ranges of models g–i appear almost uniform on the same scale. The number of modeled events in this sample was 155, with all but the slip rate only model giving a reasonable estimate of the number of events. It should not be surprising that the fault geometry model also works well, given that the slip rates are each associated with faults.

In this case our model is constructed based on the fault slip rates and a random component, so we would expect the fault slip rates + random field model to perform best. The densities in Figure 3 show how the resulting DICs overlap significantly in this model, with the slip rate + random field model outperforming the fault model + random field model very slightly. The fault geometry and slip rate models without a random field also perform well on some occasions, because the slip rate component of the intensity is larger than the random field and the slip rates are only associated with faults. This results in all five models having similar DIC values, because all five of the models go some way to explaining the observed spatial distribution of events. In this case the small sample size may also contribute to the similar performance of each of the models, with the low number of samples making it difficult to identify which model performs best.

In both cases, the resulting DICs show a preference for the models including the underlying processes used to generate the data set. In the randomly sampled model, the fault geometry model and fault geometry with

Table 2
DIC Estimates for Models of Synthetic Data Sampled as a Function of Slip Rates

Model	DIC	ΔDIC	Abundance
Random field + slip rate	3,143	0	156 ± 14
Random field + fault geometry	3,145	2	156 ± 10
Fault geometry only	3,145	2	157 ± 12
Random field only	3,155	12	155 ± 12
Slip rates only	3,191	48	135 ± 13

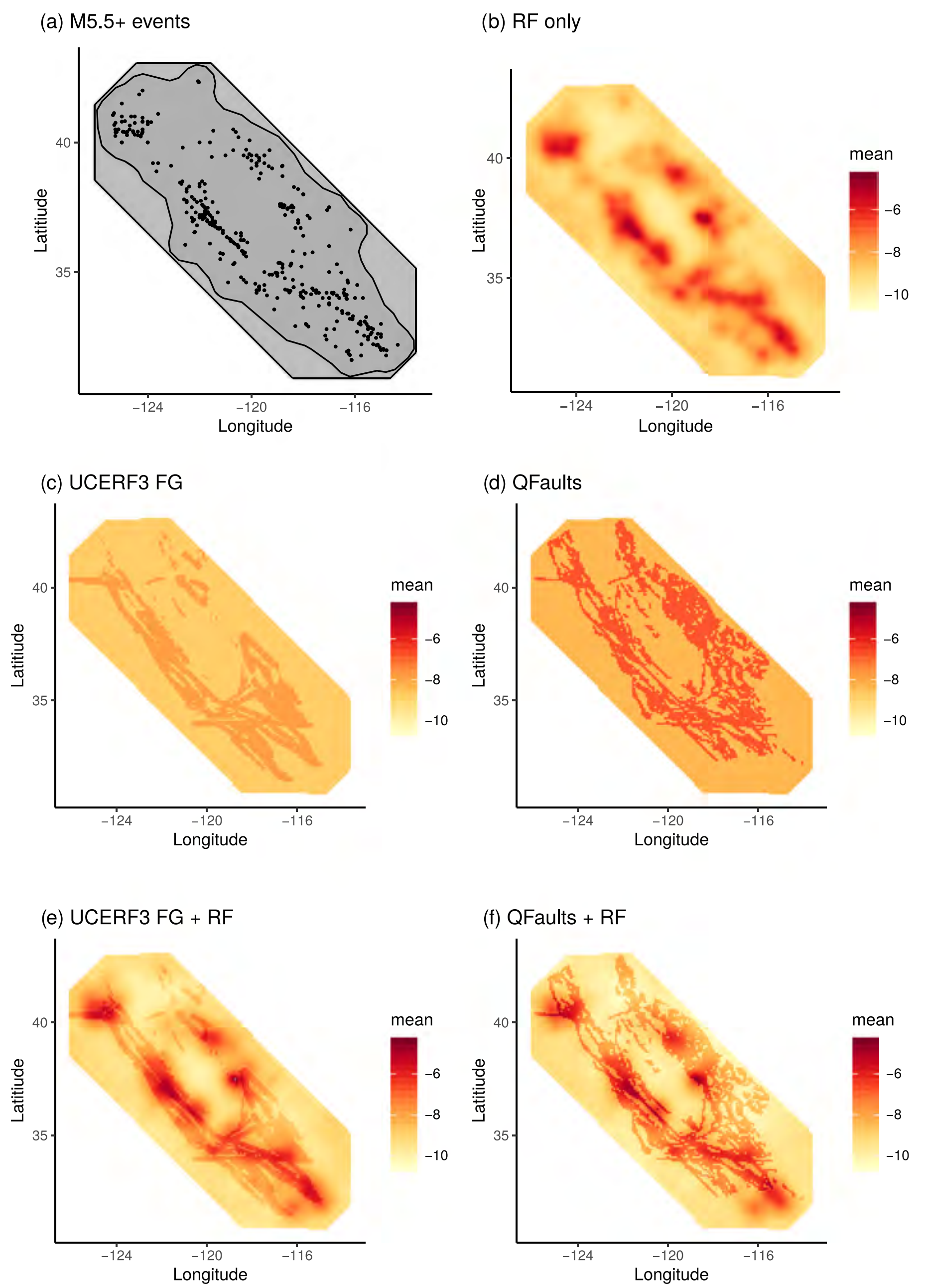


Figure 4. Inversion of fault models for two different fault geometries: UCERF3 fault geometry 3.2 and the Quaternary fault model (Qfaults). The top row shows the mesh used and locations of $M5.5+$ events (a) and the inversions for the model which includes only the random field (RF, b). The middle row (c, d) shows the resulting intensities for models of each of the two fault geometries (FG), where these have been included as a binary factor (events are on a fault or not). The bottom row (e, f) shows resulting model intensities for models which include fault polygons for each geometry and a random field. The scale bar shows the log posterior mean intensity.

random field are clearly the favored models, with overlap resulting from the random generation of events. For the slip rate weighted points there is more variation, arising from the way in which the synthetic data sets are constructed allowing more variation in the resulting catalogs. It is clear, however, that the inlabru method is able to identify models which perform well and are consistent with the underlying patterns used to create the data sets in these synthetic examples. Such consistency is a necessary (but not sufficient) criterion in assessing forecasting power Murphy (1993). We therefore feel confident in moving forward and applying the method to real data where the true model is not known a priori.

5. Inversion of California Data Set

5.1. UCERF Fault Models

We first consider the fault geometries shown in Figure 1, using M5.5+ events. Figure 4 shows the resulting intensities for five different models for real data in California: one with only a random field (top), one with each of the fault geometries only, and one for each of the fault geometries with an included random field. The fault polygons are included in the model as a binary factor covariate, such that the model checks if a fault is present at any point in space but does not consider any of the other fault information at this point. Because of this binary approach, all faults have an equal weighting within the model, so hence why the middle row shows all faults in the same color. Including a random field in the models along with the fault geometry (bottom) allows the model to account for events which do not occur on a fault polygon or where there is a significantly high number of events that the fault polygons cannot explain.

The DICs for each of these models are shown in the top part of Table 3. The random field and fault geometry maps both have a lower DIC than the random field alone, suggesting that the inclusion of the fault maps improves the model's ability to account for the locations of the events. The fault geometries alone are not as good for describing the spatial location of events, while including the random field allows the models to account for extra spatial variation. We can also use the models to predict the number of events expected, given the fitted LGCP. These are shown as abundances in Table 3, where the mean and standard deviations are reported. There are 385 events in the UCERF3 M5.5+ catalog, so the random field and fault geometry models predict very good numbers of events. The random field alone also predicts the correct number of events within one standard deviation, as do the fault geometry only models. By including a random field, we are accounting for the extra spatial variation that the fault map is missing, as a consequence of the incompleteness of the fault map, the clustering of events, or some combination of both.

The fault polygons clearly improve the intensity model when combined with a random field, but to what degree? To investigate this, we applied models to each of the fault geometry buffers (see Figure S2 in supporting information), for both UCERF3 fault geometries, resulting in a total of 16 models, with half including a random field. The DICs for these models are reported in the supporting information. Regardless of the buffer applied, the fault geometry and random field models perform better than the random field alone, with the fault geometry only models all performing worse. The combined buffer polygons perform better than the unbuffered polygons, or the uniform 1 km buffer models, but the slightly better model appears to be the dip-dependent buffer, as this already does a good job of describing the spatial distribution of M5.5+ events. Fault geometry 3.2 performs better than 3.1 because it accounts for the spatial locations of the events slightly better—240 of the 385 M5.5+ events occur within the fault polygons of FG3.2, compared to 237 for fault geometry 3.1. To simplify our model testing, we use fault geometry 3.2 for all further fault models, with the combined buffer applied. This allows us to use the UCERF3 slip rates, even though the QFault geometry is the better performer according to Table 3.

We can also add slip rates for each fault according to one of four slip rate models in UCERF3. A comparison of the four slip rate models demonstrates that the NeoKinema slip rate model performs best of the four models in terms of DIC (see supporting information Figure S3). Using the M5.5+ data, we see that the DIC for the slip rate + random field model is lower than that of the random field only or random field + fault geometry models, showing that the slip rates benefit the model.

5.2. Combining Components

The true power of the inlabru approach is in the ability to construct models with different elements and compare their performance in terms of accounting for the spatial distribution of observed earthquakes. We construct 23 models containing combinations of the elements discussed above and shown in Figure 1, using the NeoKinema fault slip rates and fault geometry 3.2. The past seismicity model and distance to fault maps are created using events from 1984 to 2004 with $M \geq 2.5$, while all of the above models are for earthquakes occurring between 2004 and 2011 in the UCERF3 catalog. The DIC for these models is much higher than when using the M5.5+ catalog due to the greater number of involved events. The mesh used for each model is constructed based on the entire UCERF3 catalog to provide adequate spatial coverage, though the mesh used for the past seismicity is extended slightly to avoid artifacts at the edges of the mesh in the later models. In the following discussion, past seismicity refers to a map of Matérn-smoothed past seismicity (see Figure 1)

10.1029/2020JB020226

Table 3
DIC Results for Combined Models With M5.5+ (Top) and M2.5+ (Bottom), Where Δ DIC Compares DIC for the "Best" Model and δDIC Compares DIC for the Next Best Model

M	Model	DIC	ΔDIC	δDIC
5.5	Random field + Qfault geometry	5,849	0.0	0.0
	Random field only	5,919	70	70
	Random field + UCERF3 fault geometry	5,920	71	1
	QFault geometry	6,296	447	376
	UCERF3 Fault geometry	6,322	473	26
2.5	Strain rate + past seismicity + fault distance + slip rate + random field	54,961	0.0	0
	Fault geometry + past seismicity + slip rate + random field	55,176	215	215
	Past seismicity + strain rate + random field	55,221	260	45
	Past seismicity + strain rate + slip rate + random field	55,239	278	18
	Fault geometry + fault distance + past seismicity + random field	55,255	294	16
	Fault geometry + past seismicity + random field	55,347	386	92
	Fault distance + past seismicity + random field	55,673	712	326
	Fault distance + past seismicity + slip rate + random field	55,685	724	12
	Fault geometry + strain rate + slip rate + random field	55,727	766	54
	Fault geometry + strain rate + random field	55,758	797	31
	Past seismicity + random field	55,979	1,018	221
	Past seismicity + slip rate + random field	55,992	1,031	13
	Fault geometry + fault distance + random field	56,185	1,224	193
	Strain rate + slip rate + random field	56,231	1,270	46
	Strain rate + random field	56,236	1,275	5
	Fault geometry + random field	56,373	1,412	137
	Fault distance + random field	56,472	1,511	236
	Fault distance + slip rate + random field	56,479	1,518	7
	Slip rate + random field	56,812	1,851	333
	Random field only	56,821	1,860	9
	Fault distance only	109,791	54,830	52,970
	Past seismicity only	112,361	57,400	2,570
	Slip rate only	112,697	57,736	336

and the fault factor refers to a binary factor covariate which returns 1 in the event that any given point is within a fault polygon and 0 elsewhere, therefore representing a fault map.

Models with random field components perform significantly better than those without. The fault distance is a more helpful inclusion for models with fault geometry than it is for models including slip rate. The fault distance model provides similar information to the fault geometry model but allows continuous variation with distance from the mapped fault location and so should be useful to both model types. This is potentially a consequence of the poor behavior of the categorical fault factor rather than the inherent utility of the fault distance map per se. The fault distance map is also limited by the resolution that can be achieved within the map, currently around 2.5 km. This resolution should be sufficient for the particular catalog used here, where the majority of events outwith fault polygons have distances of ≥2.5 km from the nearest fault polygon but may be more challenging if many events are at very small fault distances.

The combined model with past seismicity + fault distance + slip rate + strain rate + random field performs best in terms of DIC. The model DICs tell us that past seismicity is more helpful than the fault distance when combined with slip rate + random field. Past seismicity is better able to account for areas of spatial clustering than fault distance maps, but the fault distance maps can account for some fault and event location uncertainty, so are helpful for the model in terms of improving the DIC. This suggests that the fault distance map adds extra useful information to the model, even when the fault map used to create it is included.

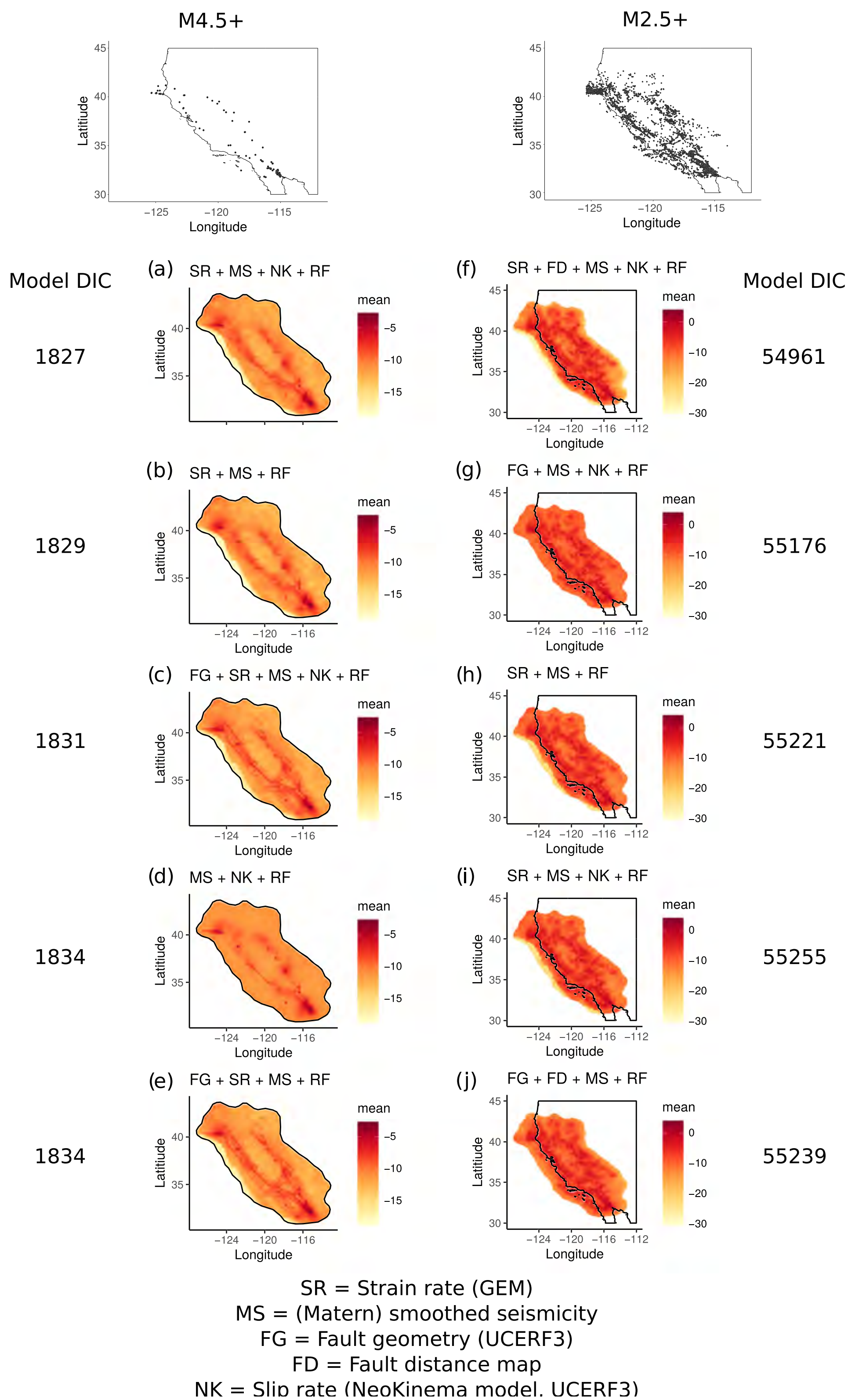


Figure 5. The five top performing models by DIC for $M4.5+$ seismicity (left, a–e) and $M2.5+$ seismicity (right, f–j). The scale bar shows the log posterior mean intensity. Intensity scales are different for the two data sets given the different number of events. Model component performance also varies depending on the data set, with the fault slip rates proving more useful in the $M4.5+$ seismicity leading to the prevalence of fault structures in the intensity maps.

The $M2.5+$ model contains significant clusters of events. To test if the model component performance is different at other magnitude cutoffs, we consider a catalog of $M4.5+$ events, again using the split catalog. The $M4.5+$ catalog contains far fewer events, with only 127 earthquakes. Many of these are related to the $M7.2$ 2010 El Mayor-Cucupah earthquake in Baja California. The location of these events and the five best models by DIC for each of these two data sets are shown in Figure 5. The two different data sets have different scales because of the different number of events, though the resulting intensity patterns are similar. The past seismicity + strain rate model performs well for both data sets in terms of its ranking. The slip rates (NK) perform better for the $M4.5+$ data set due to the higher slip rates on faults in the Baja California region, where the majority of $M4.5+$ events occur. This leads to the fault geometry being a stronger constraint on the resulting intensity as opposed to in the $M2.5+$ catalog.

As well as looking at the performance of different models spatially and their DIC, we can examine the posteriors of the different model components to see how much each component contributes to the intensities. This allows us to consider the contribution of individual components to the models and how this changes with different component combinations. This is especially useful when we are using components that may account for similar spatial patterns.

In the case of the $M2.5+$ model with the lowest DIC, the past seismicity and strain rate components contribute most significantly to the intensity with a posterior mean and standard deviation of 5.6 ± 0.2 and 3.76 ± 0.15, respectively, while the fault distance and slip rate components have a much smaller contribution at -0.076 ± 0.005 and 0.009 ± 0.007. The full posterior distributions can be found in supporting information Figure S4. Table 3 shows that though the contributions of the fault distance and slip rate are small, they do improve the DIC of the model overall.

In both data sets, the smoothed seismicity always improves the model DIC by accounting for some of the significant spatial clustering of events and by highlighting areas where many events have occurred before. As the past seismicity includes smaller events, it can therefore account for very high intensity in areas which have experienced large earthquake sequences in the past. The strain rate model also consistently performs well, both on its own and in combination with other components. The model of strain rate and past seismicity performs very well (Table 3), which is consistent with the Strader et al. (2018) assessment of the GEAR1 forecast. In this case, we can see from the output model intensity that the strain rate model is important for constraining the spatial limits of seismicity in a way that other model components cannot. This suggests that the strain rate can contribute helpful information to forecasts, and the global availability of the strain rate map means that this can be added to models for other regions where detailed information on faults may be unavailable or have higher uncertainty. Though the fault geometry and fault distance maps are not particularly useful on their own, they improve models when they are included by accounting for some of the smaller spatial structure that is not defined by better performing components. The fault distance map may account for some of the uncertainty in the fault geometry, but it also improves models which otherwise only include the fault geometry, because it can account for the lack of events in the northeast and southwest of the map. The fault distance map is likely to be more useful in areas where the fault map is less complete or highly uncertain. It may also perform better when fault buffers are smaller or not constrained by fault dips. The slip rates proved to have variable performance, proving useful for $M \geq 5.5$ models (see also supporting information Figure S1) and in $M4.5$ models but less useful for smaller events (Table 3).

Component performances are ranked according to their individual performance and their performance in combination with other components considering both data sets. The availability of the model components is also considered, as are any alternatives that could be considered in model construction if the specific component is unavailable or as a comparison. While the past seismicity is catalog dependent, it can be calculated for any region by fitting a random field to the point data. The fault geometry and slip rates are highly dependent on the area of study, but the fault distance map can potentially help to explain some of the spatial distribution of events where these components are incomplete or unavailable. The random field from models with fault geometry and fault distance components could help to identify areas in which the fault map is incomplete. The GEM strain rate model is a global strain rate model and is therefore available for any region; however, the deforming cells at oceanic plate boundaries are poorly constrained which may lead to some model artifacts. This can be seen in our California example (Figure 1) as the thin blue stripe in the south of the map, which is the result of the plate boundary imposed by the modeling.

Table 4
Fault Model Rankings at Different Magnitude Cutoffs

Model	$M5.5+$	$M5.0+$	$M4.5+$	$M4.0+$	$M3.5+$
UCERF faults	3	3	2	1	1
QFaults	1	1	1	2	3
NK slip rates	2	2	3	3	4
Randomized slip rates	5	5	4	4	2
Random field only	4	4	5	5	5

This provides us with a foundation for building models for any region, by identifying which components work well and which components can be used in their place should they be unavailable, which is summarized in Figure S5 of supporting information. Extending this approach by considering other data types and performance in other regions could prove valuable for constructing earthquake forecasts in future work.

5.3. Fault Geometry, Slip Rates, and Their Effects at Different Magnitudes

We have considered so far the UCERF3 fault geometry and slip rate data, but the above comparisons of model component combinations at different magnitudes demonstrate that the slip rates may have different importance for different data sets. We test this by comparing slip rate and fault geometry models at different magnitude cutoffs. We also use a set of randomized slip rates, where the slip rates are randomly reassigned to faults within the geometry, to test how much the usefulness of the slip rates is influenced by the actual slip rate values and how much is a function of the models ability to consider different faults more or less important. Finally, we also include the QFault model as an alternative fault geometry. This combination of models allows us to assess the performance of slip rates and fault geometry as a function of the magnitude cutoffs and to further explore why different model components perform well or otherwise.

Table 4 ranks each of the five models at several different magnitude thresholds, highlighting changes to model ranking as a result of the changes in data set. Lower magnitude thresholds will result in catalogs with more significant clustering, but as mentioned with the $M4.5+$ catalog above, there may still be significant clustering even at higher magnitude cut-offs. Longer catalogs could help to minimize the effect of recent large sequences. The randomized slip rates will perform better when they better account for the observed seismicity at different cutoffs. In this case the randomized slip rates do not perform better than modeled slip rates at magnitudes $M4+$. This is consistent with the results for $M2.5+$ seismicity above and suggests that the slip rates are more useful for describing the locations of large events. For events $M4.5+$, the slip rates perform better than the fault geometry alone, but in each of these models the Quaternary fault model performs better than the UCERF3 fault geometry. Figure 1 demonstrates that the Quaternary fault geometry includes many faults to the northeast which can better account for seismicity in this region. The UCERF3 fault geometry performs better than the Quaternary faults at lower magnitudes due to the changing number and distribution of events. For larger magnitude catalogs, events not related to the UCERF3 geometry faults will be significant as a function of catalog size, but this will be less true for smaller magnitude events (and therefore larger catalogs). This may also be related to the buffering of the fault geometries being different, with the fixed buffering of the Quaternary faults performing more poorly than the dip-dependent buffering in the UCERF geometry. The changes in model ranking at different magnitude cutoffs highlight the effect of clustering and the inclusion of smaller events on model component performance and the complexity involved in constructing such models.

In summary, models including the fault maps and strain rate data are always improvements on the random field seismicity model alone and add more information for larger magnitude thresholds. This is important for applications to seismic hazard, which is normally dominated by intermediate to large magnitude events, that is, if the frequency-moment distribution takes the form of a pure power law or a power law with an exponential taper (e.g., Main, 1995). In this case there may be a minimum threshold for being likely to be felt at a particular intensity likely to cause damage, often set at magnitude 5 for design of Nuclear Power Plants (Bommer & Crowley, 2017). In such cases, the analysis presented in Table 4 shows that the fault map and strain rate data provide critical constraints in describing the seismicity.

6. Discussion

6.1. Smoothed Seismicity

On short forecasting timescales, we should expect local clustering of earthquakes to dominate so we would expect recent smoothed seismicity models to be informative. On longer timescales we would expect the entire seismogenic region to be sampled so a longer sample may better account for longer-term trends in seismicity that are better captured by fault and strain rate maps. For time-independent forecasting, as long a sample as possible should be used, so that the effect of short-term clustering is reduced and as many large events as possible are captured within the past seismicity model.

The earliest models submitted to the Regional Likelihood Earthquake Model testing (RELM, a precursor to CSEP) were mostly constructed on the basis of some form of smoothed seismicity (Field, 2007), though some also included strain rates or geological information. The preliminary results for California found that the smoothed past seismicity model of Helmstetter et al. (2007) provided the best result of all submitted models (Schorlemmer et al., 2007; Zechar et al., 2013) whether aftershocks were included or not. Smoothed seismicity has also been found to perform well for forecasting when combined with strain rate (Bird et al., 2015; Strader et al., 2018) and when updated regularly as part of 1 day forecasts in New Zealand (Rhoades et al., 2018). Our results above demonstrated the utility of the past seismicity in spatial seismicity models, both individually and in combination with other components. We saw that including the past seismicity with other components always improved model DIC compared to models with the same components and no past seismicity. We are therefore quite confident that our results using the inlabru approach are consistent with findings well documented elsewhere in the literature. Further, we have demonstrated that a Matérn smoothing of the spatial intensity is also suitable for describing past seismicity (see supporting information and Figure S7 for explicit comparison). The improved performance of the Matérn smoothed seismicity in the inlabru model may be a result of the alternative gridding—the Matérn smoothing is based upon a Delauney triangulated mesh constructed from earthquake locations, rather than a uniform grid. We propose that this may well influence the effect of the smoothed seismicity within the model. In particular, this may be interesting in settings where gridded smoothed seismicity has proved less useful to forecast models, such as in the Italian CSEP tests (Taroni et al., 2018) which found recent smoothed seismicity to perform less well than a model which included more historic seismicity and fault locations. The inlabru framework would also allow the ranking of models containing each of these components as combining the components in one model is straightforward.

6.2. Slip Rates in California

Our results demonstrated that the slip rate performance is quite variable but generally better for large events which are more likely to be independent. This would suggest that slip rates could be a valuable constraint for stationary, time-independent hazard models. To explore the full effect of slip rates on the model DIC, we remove each fault polygon from the model sequentially and record the DIC for each model to calculate a ΔDIC for each fault. We do this for two different model types to see the effect of including a random field and consider only events with $M \geq 5.5$.

Figure 6 shows the results for a fault model only (left) and for a model that includes a random field (right). High positive ΔDIC values suggest that a fault is important to the model, as removing it causes the total model DIC to increase relative to the model with all faults. Conversely, faults with negative ΔDIC values suggest that the overall model is improved by their absence. The models with fault slip rates alone return positive ΔDIC values for 102 faults, while the models with random field return positive ΔDIC values for 286 of 320 faults. This suggests that the random field model finds all faults more valuable to the model than a model that includes slip rate alone, such that removing any particular individual fault will have a similar effect on DIC, with a few exceptions. Removing even the largest faults has little effect on DIC, while removing some of the smaller faults is more significant due to the greater number of events associated with them, as shown in Figure S8.

The models that include the random field are more likely to have an increased DIC when faults are removed than the slip rate only model (Figure S8). Removing the faults from the model in this way does not affect DIC for the models for 95 models without random field component, but none of the faults have $\Delta\mathrm{DIC} = 0$ when a random field is included in the model, suggesting that all faults have a contribution to the total DIC in the model with random field. Further, the faults with large ΔDIC for the models including a random field have

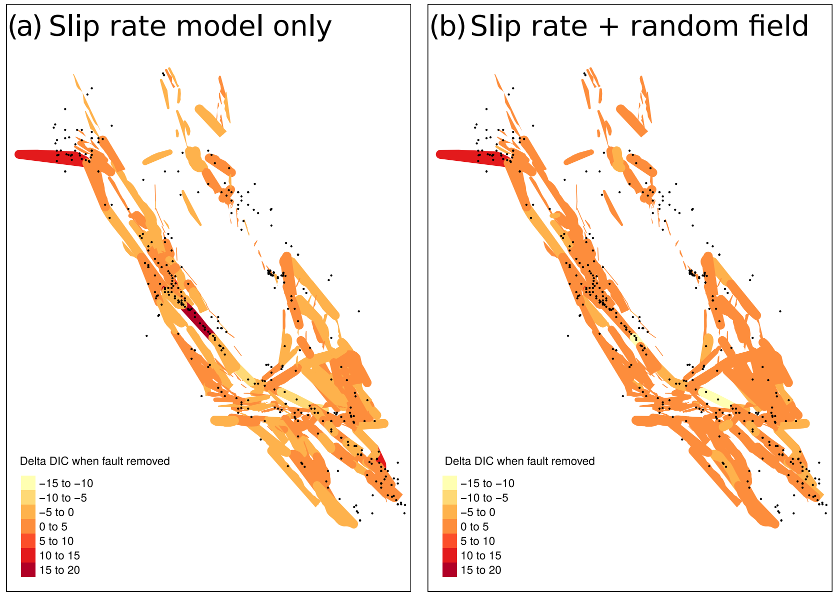


Figure 6. Δ*DIC* values for each fault in the UCERF3 catalog, where the color of the fault reflects the change in total model DIC when it is removed from the fault polygons. The locations of the *M*5.5+ events are shown in black, with the left panel (a) showing the results for a model of slip rate fault polygons only and the model on the right (b) including a random field component.

a smaller ΔDIC than in models without a random field, such that removing any one individual fault will have a smaller effect on total DIC for a model with random field. It is also worth noting that the total DIC is always lower for models with a random field than models without. The random field is able to explain more of the spatial distribution than the slip rate model alone, such that removing any individual fault does not have a significant impact on resulting DIC. This is promising for areas where the fault map is less complete. These findings demonstrate that the slip rates in California cannot fully account for the observed spatial distribution of earthquakes. The slip rate or size of an individual fault does not correlate with the number of events occurring within the fault polygon.

6.3. The 2019 Ridgecrest Sequence

The July 2019 Ridgecrest sequence in southern California allowed us to apply the inlabru method to a recent event sequence. We used a catalog of 1,459 *M*2.5+ events that occurred between 15 June and 15 July 2019 retrieved from the USGS earthquake database. We consider events that fall within an area of −117 to −118 longitude and 35–37 latitude. Within our Ridgecrest catalog, 985 events are not within a buffered UCERF3 fault polygon, with the remaining 474 events linked to five different fault polygons. The *M*7.1 event on 6 July and the *M*6.4 event 2 days previously are not within any UCERF3 buffered fault polygon, with the largest event directly linked to a fault polygon being a magnitude 5.5 event on the Airport Lake fault. This motivated us to try the USGS Quaternary fault model as an alternative fault map.

We constructed four models for seismicity during this time period (Figure 7), each including a random field and one spatial covariate: UCERF3 and Quaternary fault geometries, the GEM strain rate and a past seismicity model based on the entire UCERF3 catalog, and subsequent events (i.e., events from May 2012 to January 2019 as well as the UCERF3 events). All four models were compared to a random field only model and proved to improve the model DIC, highlighting that each of these covariates was helpful in describing the spatial patterns of seismicity.

The past seismicity and strain rate models perform best according to the DICs recorded in Table 5, though the predicted intensity models show the grid pattern that comes from the spatial resolution of the input covariates. This gridding appears to benefit the models, with high intensity areas falling within grid squares with higher past seismicity. The area of the Ridgecrest events is a generally higher strain rate area than most of Southern California. In both cases, increased resolution of the data may improve the performance, but we can see that the current resolution is enough to improve the model. The UCERF3 geometry

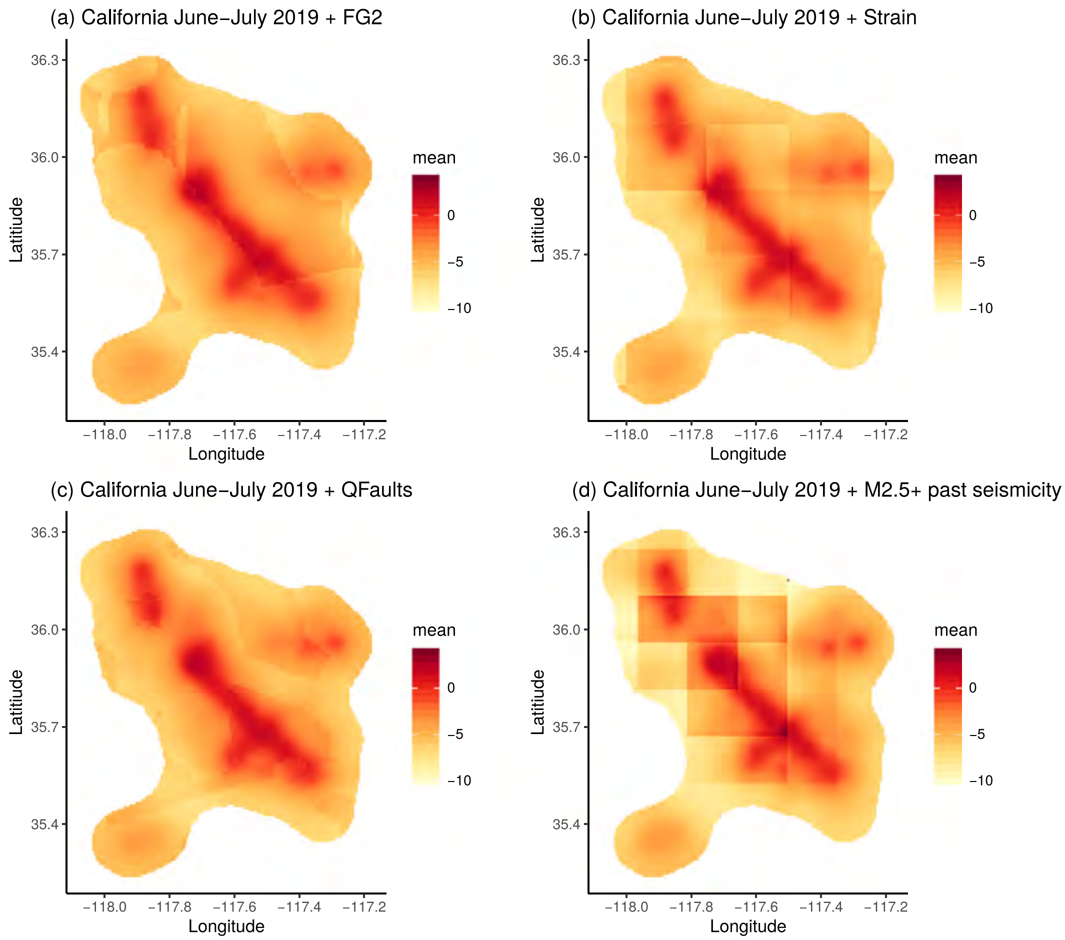


Figure 7. Predicted field intensities for the Ridgecrest events for four different models: (a)UCERF faults, (b) strain rate, (c) Quaternary fault geometry, and (d) past seismicity. The scale bar shows the log posterior mean intensity.

outperforms that of the Quaternary faults, but this appears to be a result of the buffering applied: The larger buffers of the UCERF3 model allow it to account for some of the events despite the lack of matching geometry (see supporting information figure S6). The Quaternary fault model contains many smaller faults, and it is likely that a larger spatial buffer on these faults would fill the area of the Ridgecrest events without necessarily accounting for the correct geometry. In this case neither fault map performs as well as the strain rate, which reinforces the usefulness of the strain rate in spatial models of seismicity. Choosing appropriate buffers for fault projections will remain an ongoing challenge for fault-dependent models, but by including other components in a model the effect of incomplete fault maps or poorly chosen buffers can potentially be reduced.

Table 5
Seismicity Models for July 2019 Ridgecrest Sequence, Where Δ DIC Compares DIC for the "Best" Model and δDIC Compares DIC for the Next Best Model

Model	DIC	ΔDIC	δDIC
Past seismicity + random field	1,006	0	0
Strain + random field	1,097	91	91
UCERF faults + random field	1,121	115	24
Quaternary faults + random field	1,129	123	8

6.4. Current Limitations

This paper is the first attempt at applying inlabru for modeling seismic hazard based primarily on existing functionality. There are currently several limitations that we are addressing in ongoing work. This will include extending the model to include magnitudes of events as marks jointly modeled with the random field. Event magnitudes are likely to affect the spatial correlation range of events, thus modifying the structure of our random field. They have been considered here in terms of the choice of catalog only, with the magnitudes of individual events not considered in the model. The model will then be further extended to include time dependence, which requires the model to be self-exciting, for example, applying an ETAS model to the data. Further extensions to account for uncertainty in event location and in the various input parameters will also be considered, as well as the development of a robust model assessment and comparison approach so as not to rely so heavily on the DIC and number of events alone. Marzocchi et al. (2012) suggest that when constructing hybrid models, the correlation between included forecasts should be considered when weighting model components. The components of our models are sometimes highly correlated, but the contribution from each component to the total mean intensity can be considered using the component posterior means. This allows us to identify under which circumstances the contributions of individual components are most significant to the observed intensity. Future work will make use of these component contributions in the construction of prospective forecasts of seismicity and assess the extent to which the inclusion of different (perhaps correlated) parameters affects the spatial distribution of forecast events. Nevertheless, we have presented a proof of concept for the inlabru method, demonstrating that it is a promising method that can be developed in future research. We have also used it to demonstrate several findings consistent with previous work and some that are specific to the present work.

7. Conclusions

We have demonstrated for the first time that LGCP models for earthquake data can be constructed, fitted, and tested efficiently using INLAs in a purely data-driven approach. The inlabru approach and model framework allow the quick and easy construction of seismicity models that include various different types of information, including fault polygons (with or without some attached mark of their own), derived products such as distance to the nearest fault, and spatially continuous models, such as strain rate data and past seismicity. The inlabru approach confirms results from the literature in terms of the inclusion of past seismicity and fault information in seismicity models, allows the straightforward combination of different data types, and allows the ranking of models by making use of the model DIC. Including strain rate data constrains the spatial limits of seismicity in California, while maps of distance to nearest fault prove beneficial even to models which also include the fault maps themselves by accounting for seismicity not associated with specific mapped faults. Further, the importance of information within individual components can also be tested, such as by considering changes to the model DIC when individual faults are removed. Such a framework allows the user to identify which model components truly benefit the model, the best combination of different model components, and to identify spatial areas in which their model is currently lacking by considering the random field.

Acronyms

LGCP Log-Gaussian Cox process
INLA Integrated nested Laplace approximations
GMRF Gaussian-Markov Random Field
SPDE Stochastic partial differential equations
DIC Deviance Information Criterion

Data Availability Statement

UCERF3 data are available online (https://pubs.usgs.gov/of/2013/1165/).

Acknowledgments
K. B. was funded during this work by an EPSRC PhD studentship (Grant 1519006) and during the writing of this paper by NERC-NSF grant NE/R000794/1 and by the Real-time Earthquake Risk Reduction for a Resilient Europe "RISE" project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant Agreement 821115. We thank Francesco Serafini for feedback on the synthetic models and an early version of this paper. The authors thank Shyam Nandan and an anonymous reviewer for their thoughtful and constructive comments.

References

Bach, C., & Hainzl, S. (2012). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research*, *117*, B04312. https://doi.org/10.1029/2011JB008901

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*(6), 760–766. https://doi.org/10.1111/2041-210X.13168

Bird, P., Jackson, D. D., Kagan, Y. Y., Kreemer, C., & Stein, R. S. (2015). GEAR1: A global earthquake activity rate model constructed from geodetic strain rates and smoothed seismicity. *Bulletin of the Seismological Society of America*, *105*(5), 2538–2554. https://doi.org/10.1785/0120150058
Bivand, R., Keitt, T., & Rowlingson, B. (2019). rgdal: Bindings for the "geospatial" data abstraction library. https://CRAN.R-project.org/package=rgdal
Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, second edition*. NY: Springer. https://www.asdar-book.org/
Bivand, R., & Rundel, C. (2019). rgeos: Interface to geometry engine - open source ('GEOS'). https://CRAN.R-project.org/package=rgeos
Blangiardo, M., & Cameletti, M. (2013). Spatial and spatial-temporal bayesian models with R-INLA. *Spatial and Spatiotemporal Epidemiology*, *4*, 33–49. https://doi.org/10.1016/j.sste.2012.12.001
Bommer, J. J., & Crowley, H. (2017). The purpose and definition of the minimum magnitude limit in PSHA calculations. *Seismological Research Letters*, *88*(4), 1097–1106. https://doi.org/10.1785/0220170015
Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, *17*, 129–164. https://doi.org/10.2307/2983950
Díaz-Avalos, C., Juan, P., & Serra-Saurina, L. (2016). Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes. *Forest Ecology and Management*, *381*, 360–369. https://doi.org/10.1016/j.foreco.2016.09.013
Diggle, P. J., Moraga, P., Rowlingson, B., & Taylor, B. M. (2013). Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, *28*(4), 542–563. https://doi.org/10.1214/13-STS441
Dutra Silva, L., Brito de Azevedo, E., Bento Elias, R., & Silva, L. (2017). Species distribution modeling: Comparison of fixed and mixed effects models using INLA. *ISPRS International Journal of Geo-Information*, *6*(12), 391. https://doi.org/10.3390/ijgi6120391
Field, E. H. (2007). Overview of the working group for the development of regional earthquake likelihood models (RELM). *Seismological Research Letters*, *78*(1), 7–16. https://doi.org/10.1785/gssrl.78.1.7
Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., et al. (2014). Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model. *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180. https://doi.org/10.1785/0120130164
Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., et al. (2015). Long-term time-dependent probabilities for the third uniform California earthquake rupture forecast (UCERF3). *Bulletin of the Seismological Society of America*, *105*(2), 511–543. https://doi.org/10.1785/0120140093
Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M. T., van der Elst, N., Jordan, T. H., et al. (2017). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America*, *107*(3), 1049–1081. https://doi.org/10.1785/0120160173
Gómez-Rubio, V., Cameletti, M., & Finazzi, F. (2015). Analysis of massive marked point patterns with stochastic partial differential equations. *Spatial Statistics*, *14*, 179–196. https://doi.org/10.1016/j.spasta.2015.06.003
Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the matérn correlation family. *Biometrika*, *93*(4), 989–995.
Helmstetter, A. (2003). Is earthquake triggering driven by small earthquakes? *Physical Review Letters*, *91*(5), 058501. https://doi.org/10.1103/PhysRevLett.91.058501
Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research*, *110*, B05S08. https://doi.org/10.1029/2004JB003286
Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2007). High-resolution time-independent grid-based forecast for M >= 5 earthquakes in California. *Seismological Research Letters*, *78*(1), 78–86. https://doi.org/10.1785/gssrl.78.1.78
Helmstetter, A., & Sornette, D. (2002a). Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models. *Physical Review E*, *66*(6), 061104. https://doi.org/10.1103/PhysRevE.66.061104
Helmstetter, A., & Sornette, D. (2002b). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research*, *107*(B10), 2237. https://doi.org/10.1029/2001JB001580
Helmstetter, A., & Sornette, D. (2003). Båth's law derived from the Gutenberg-Richter law and from aftershock properties. *Geophysical Research Letters*, *30*(20), 2069. https://doi.org/10.1029/2003GL018186
Hijmans, R. J. (2019). raster: Geographic data analysis and modeling. https://CRAN.R-project.org/package=raster
Huc, M., & Main, I. G. (2003). Anomalous stress diffusion in earthquake triggering: Correlation length, time dependence, and directionality. *Journal of Geophysical Research*, *108*(B7), 2324. https://doi.org/10.1029/2001JB001645
Illian, J. B., Sørbye, S. H., & Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, *6*(4), 1499–1530. https://doi.org/10.1214/11-AOAS530
Jordan, T. H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., et al. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, *54*(4), 319–391. https://doi.org/10.4401/ag-5350
Jordan, T. H., & Jones, L. M. (2010). Operational earthquake forecasting: Some thoughts on why and how. *Seismological Research Letters*, *81*(4), 571–574. https://doi.org/10.1785/gssrl.81.4.571
Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144–161.
Kreemer, C., Blewitt, G., & Klein, E. C. (2014). A geodetic plate motion and global strain rate model. *Geochemistry, Geophysics, Geosystems*, *15*, 3849–3889. https://doi.org/10.1002/2014GC005407
Lindgren, F., & Rue, H. A. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–26. https://doi.org/10.18637/jss.v063.i19
Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x
Lombardo, L., Opitz, T., & Huser, R. (2018). Point process-based modeling of multiple debris flow landslides using INLA: An application to the 2009 Messina disaster. *Stochastic Environmental Research and Risk Assessment*, *32*(7), 2179–2198. https://doi.org/10.1007/s00477-018-1518-0
Main, I. G. (1995). Earthquakes as critical phenomena: Implications for probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, *85*(5), 1299–1308.
Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis*, *67*, 68–83. https://doi.org/10.1016/j.csda.2013.04.014
Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in Italy. *Seismological Research Letters*, *85*(5), 961–969. https://doi.org/10.1785/0220130219

Marzocchi, W., Zechar, J. D., & Jordan, T. H. (2012). Bayesian forecast evaluation and ensemble earthquake forecasting. *Bulletin of the Seismological Society of America*, *102*(6), 2574–2584. https://doi.org/10.1785/0120110327

Michael, A. J., & Werner, M. J. (2018). Preface to the focus section on the collaboratory for the study of earthquake predictability (CSEP): New results and future directions. *Seismological Research Letters*, *89*(4), 1226–1228. https://doi.org/10.1785/0220180161

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, *8*(2), 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2

Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to California. *Journal of Geophysical Research: Solid Earth*, *122*, 5118–5143. https://doi.org/10.1002/2016JB013266

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. https://doi.org/10.1080/01621459.1988.10478560

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. https://doi.org/10.1023/A:1003403601725

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, *63*(3), 217–229. https://doi.org/10.5047/eps.2010.09.001

Ogata, Y., & Zhuang, J. (2006). Space time ETAS models and an improved extension. *Tectonophysics*, *413*(1–2), 13–23. https://doi.org/10.1016/J.TECTO.2005.10.016

Pebesma, E. E. J., & Bivand, R. R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13.

Rhoades, D. A., Christophersen, A., & Gerstenberger, M. C. (2015). Multiplicative earthquake likelihood models based on fault and earthquake data. *Bulletin of the Seismological Society of America*, *105*(6), 2955–2968. https://doi.org/10.1785/0120150080

Rhoades, D. A., Christophersen, A., Gerstenberger, M. C., Liukis, M., Silva, F., Marzocchi, W., et al. (2018). Highlights from the first ten years of the New Zealand earthquake forecast testing center. *Seismological Research Letters*, *89*(4), 1229–1237. https://doi.org/10.1785/0220180032

Rhoades, D. A., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., Schorlemmer, D., Werner, M. J., & Jordan, T. H. (2014). Regional earthquake likelihood models II: Information gains of multiplicative hybrids. *Bulletin of the Seismological Society of America*, *104*(6), 3072–3083. https://doi.org/10.1785/0120140035

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, *4*(1), 395–421. https://doi.org/10.1146/annurev-statistics-060116-054045

Sadykova, D., Scott, B. E., De Dominicis, M., Wakelin, S. L., Sadykov, A., & Wolf, J. (2017). Bayesian joint models with INLA exploring marine mobile predator prey and competitor species habitat overlap. *Ecology and Evolution*, *7*(14), 5212–5226. https://doi.org/10.1002/ece3.3081

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A. (2007). Earthquake likelihood model testing. *Seismological Research Letters*, *78*(1), 17–29. https://doi.org/10.1785/gssrl.78.1.17

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., et al. (2018). The collaboratory for the study of earthquake predictability: Achievements and priorities. *Seismological Research Letters*, *89*(4), 1305–1313. https://doi.org/10.1785/0220180053

Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*, 449–469. https://doi.org/10.1002/2016JB012809

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2015). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, *103*(1), 49–70. https://doi.org/10.1093/biomet/asv064

Simpson, D., Lindgren, F., & Rue, H. (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, *1*, 16–29. https://doi.org/10.1016/j.spasta.2012.02.003

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485–493. https://doi.org/10.1111/rssb.12062

Strader, A., Werner, M., Bayona, J., Maechling, P., Silva, F., Liukis, M., & Schorlemmer, D. (2018). Prospective evaluation of global earthquake forecast models: 2 yrs of observations provide preliminary support for merging smoothed seismicity with geodetic strain rates. *Seismological Research Letters*, *89*(4), 1262–1271. https://doi.org/10.1785/0220180051

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., et al. (2018). Prospective CSEP evaluation of 1 day, 3 month, and 5 yr earthquake forecasts for Italy. *Seismological Research Letters*, *89*(4), 1251–1261. https://doi.org/10.1785/0220180031

Taylor, B. M., & Diggle, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, *84*(10), 2266–2284. https://doi.org/10.1080/00949655.2013.788653

Teng, M., Nathoo, F., & Johnson, T. D. (2017). Bayesian computation for Log-Gaussian Cox processes: A comparative analysis of methods. *Journal of Statistical Computation and Simulation*, *87*(11), 2227–2252. https://doi.org/10.1080/00949655.2017.1326117

Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, *84*(6), 1–39. https://doi.org/10.18637/jss.v084.i06

Veen, A., & Schoenberg, F. P. (2008). Estimation of space time branching process models in seismology using an EM type algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624. https://doi.org/10.1198/016214508000000148

Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, *32*(2), 1–62. https://doi.org/10.2307/2984402

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, *11*(4), 503–538. https://doi.org/10.1016/0169-2070(95)00621-4

Vere-Jones, D., & Davies, R. B. (1966). A statistical survey of earthquakes in the main seismic region of New Zealand. *New Zealand Journal of Geology and Geophysics*, *9*(3), 251–284. https://doi.org/10.1080/00288306.1966.10422815

Werner, M. J., Helmstetter, A., Jackson, D. D., & Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, *101*(4), 1630–1648. https://doi.org/10.1785/0120090340

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., et al. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Annals of Applied Statistics*, *11*(4), 2270–2297. https://doi.org/10.1214/17-AOAS1078

Zechar, D. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013). Regional earthquake likelihood models I: First-order results. *Bulletin of the Seismological Society of America*, *103*(2 A), 787–798. https://doi.org/10.1785/0120120186

References From the Supporting Information

Bird, P. (2009). Long-term fault slip rates, distributed deformation rates, and forecast of seismicity in the western United States from joint fitting of community geologic, geodetic, and stress direction data sets. *Journal of Geophysical Research*, *114*, B11403. https://doi.org/10.1029/2009JB006317

Zeng, Y., & Shen, Z. (2016). A fault based model for crustal deformation, fault slip rates, and off fault strain rate in California. *Bulletin of the Seismological Society of America*, *106*(2), 766–784. https://doi.org/10.1785/0120140250

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

Pseudo-prospective testing of 5-year earthquake forecasts for California using inlabru

Kirsty Bayliss[1], Mark Naylor[1], Farnaz Kamranzad[1], and Ian Main[1]
[1]School of GeoSciences, University of Edinburgh

Correspondence: Kirsty Bayliss (Kirsty.Bayliss@ed.ac.uk)

Abstract. Probabilistic earthquake forecasts estimate the likelihood of future earthquakes within a specified time-space-magnitude window and are important because they inform planning of hazard mitigation activities on different timescales. The spatial component of such forecasts, expressed as seismicity models, generally rely upon some combination of past event locations and underlying factors which might affect spatial intensity, such as strain rate, fault location and slip rate or past seismicity. For the first time, we extend previously reported spatial seismicity models, generated using the open source inlabru package, to time-independent earthquake forecasts using California as a case study. The inlabru approach allows the rapid evaluation of point process models which integrate different spatial datasets. We explore how well various candidate forecasts perform compared to observed activity over three contiguous five year time periods using the same training window for the seismicity data. In each case we compare models constructed from both full and declustered earthquake catalogues. In doing this, we compare the use of synthetic catalogue forecasts to the more widely-used grid-based approach of previous forecast testing experiments. The simulated-catalogue approach uses the full model posteriors to create Bayesian earthquake forecasts. We show that simulated-catalogue based forecasts perform better than the grid-based equivalents due to (a) their ability to capture more uncertainty in the model components and (b) the associated relaxation of the Poisson assumption in testing. We demonstrate that the inlabru models perform well overall over various time periods, and hence that independent data such as fault slip rates can improve forecasting power on the time scales examined. Together, these findings represent a significant improvement in earthquake forecasting is possible, though this has yet to be tested and proven in true prospective mode.

1 Introduction

Probabilistic earthquake forecasts represent our best understanding of the expected occurrence of future seismicity (Jordan and Jones, 2010). Developing demonstratively robust and reliable forecasts is therefore a key goal for seismologists. A key component of such forecasts, regardless of the timescale in question, is a reliable spatial seismicity model that incorporates as much useful spatial information as possible in order to identify areas at risk. For example in probabilistic seismic hazard modelling (PSHA) a time independent spatial seismicity model is developed by combining a spatial model for the seismic sources with a frequency magnitude distribution. In light of the ever-growing abundance of earthquake data and the presence of spatial information that might help understand patterns of seismicity, Bayliss et al. (2020) developed a spatially-varying point

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

process model for spatial seismicity using Log-Gaussian Cox processes evaluated with the Bayesian integrated nested Laplace approximation method (Rue et al., 2009) implemented with the open-source R package inlabru (Bachl et al., 2019). Time-independent earthquake forecasts require not only an understanding of spatial seismicity, but also need to prove themselves to be consistent with observed event rates and earthquake magnitudes in the future.

Forecasts can only be considered meaningful if they can be shown to demonstrate a degree of proficiency at describing what future seismicity might look like. The Regional Earthquake Likelihood Model (RELM, Field, 2007) experiment and subsequent Collaboratory for the study of earthquake predictability (CSEP) experiments challenged forecasters to construct earthquake forecasts for California, Italy, New Zealand and Japan (e.g. Schorlemmer et al., 2018; Taroni et al., 2018; Rhoades et al., 2018, and other articles in this special issue) to be tested in prospective mode using a suite of pre-determined statistical tests. The testing experiments found that the best performing model for seismicity in California was the Helmstetter et al. (2007) smoothed seismicity model, whether aftershocks were included or not (Zechar et al., 2013). This model requires no mosaic of seismic source zones to be constructed, requiring only one free parameter - the spatial dimension of the smoothing kernel. In the years since this experiment originally took place, there has been considerable work both to improve the testing protocols and to develop new forecast models which may improve upon the performance of the data-driven Helmstetter et al. (2007) model, primarily by including different types of spatial information to augment what can be inferred from the seismicity alone. Multiplicative hybrid models (Marzocchi et al., 2012; Rhoades et al., 2014, 2015) have shown some promise, but these require some care in construction and further testing is needed. The performance of smoothed seismicity models has been found to be inconsistent in testing outside of California, e.g. with the Italian CSEP experiment finding smoothed past seismicity alone did not do as well as models with much longer term seismicity and fault information (Taroni et al., 2018). Thus, finding and testing new methods of allowing different data types to be easily included in developing a forecast model is an important research goal. Here we explore in particular the role of testing an ensemble of point process simulated catalogues (Savran et al., 2020) in comparison with traditional grid-based tests, where the underlying point process is locally averaged in a grid element.

In this paper we construct and test a series of time-independent forecasts for California by building on the spatial modelling approach described by Bayliss et al. (2020). As a first step in the modelling we take a pseudo-prospective approach to model design, with the forecasts being tested retrospectively on time periods subsequent to the data on which they were originally constructed, and test the models' performance against actual outcome using the pyCSEP package (Savran et al., 2021). This is not a sufficient criterion for evaluating forecast power in true prospective mode, but is a necessary step on the way, and (given similar experience of 'hindcasting' in cognate disciplines such as meteorology) can inform the development of better real-time forecasting models. The results presented here will in due course be updated and tested in true prospective mode, using a training dataset up to the present. We first test the pseudo-prospective seismicity forecasts in a manner consistent with the RELM evaluations. For this comparison we use a grid of event rates and the same training and testing time windows to provide a direct comparison to the forecasts of the smoothed seismicity models of Helmstetter et al. (2007), which use seismicity data alone as an input, and provide a suitable benchmark to our study. We then extend this approach to the updated CSEP evaluations for simulated catalogue forecasts (Savran et al., 2020) and show that the synthetic catalogue-based forecasts perform better than the grid-based equivalents, due to their ability to capture more uncertainty in the model components and the relaxation of the Poisson assumption in testing.

2 Method

We develop a series of spatial models of seismicity modelled by a time-independent Log-Gaussian Cox Process and fitted with inlabru, as described in detail in Bayliss et al. (2020), and whose workflow is summarised in Figure 1. The models take as input twenty years (1984-2004) of California earthquakes with magnitude ≥ 4.95 from the UCERF3 dataset (Field et al., 2014), with the magnitude cutoff chosen to be consistent with the RELM forecast criteria. The locations of these events are an intrinsic component of a point process model with spatially varying intensity $\lambda(\mathbf{s})$, where the intensity is described as a function of some underlying spatial covariates $x_m(\mathbf{s})$, e.g. input data from seismicity catalogues or geodetic observations of strain rate, and a Gaussian random field $\zeta(\mathbf{s})$ to account for spatial structure that is not explained by the model covariates. The spatially varying intensity then can be described with a linear predictor $\eta(\mathbf{s})$ such that

$$\lambda(\mathbf{s}) = e^{\eta(\mathbf{s})}, \tag{1}$$

and $\eta(\mathbf{s})$ can be broken down into a sum of linearly combined components

$$\eta(\mathbf{s}) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m(\mathbf{s}) + \zeta(\mathbf{s}). \tag{2}$$

The β_0 term is an intercept term, which would describe a spatially homogeneous Poisson intensity if no other components were included, and each β_m describes the weighting of individual spatial components in the model. β_0 is essentially the uniform average or base-level intensity, which allows the possibility of earthquakes happening over all of the region of interest as a null hypothesis, so 'surprises' are possible, though unlikely after adding the other terms and renormalising. The models are built on a mesh (step 2 of Figure 1) which is required to perform numerical integration in the spatial domain, with the model intensity evaluated at each mesh vertex as a function of the random field (RF, which is mapped by stochastic partial differential equations or SPDE in step 3 of Figure 1) and other components of the linear predictor function (equation 2). Fitting the model results in a posterior probability distribution for each of the model component weights, the random field and the joint posterior probability distribution for the intensity as a function of these components. The expected number of events can then be approximated by summing over the mesh and associated weights over the area of interest (Step 5 of figure 1). The performance of the models can then be evaluated by comparing the expected versus the observed number of events, and the models ranked using the resulting model deviance information criterion (DIC). DIC is commonly applied in other applications of Bayesian inference, including inlabru applications to other problems, such as spatial distributions of species in ecology. With the definition used here, DIC is lower for a model with better predictive skill.

In Bayliss et al. (2020) a range of California spatial forecast models were tested on how well the spatial model created by inlabru fitted the observed point locations, so were essentially a retrospective test of the spatial model alone in order to

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

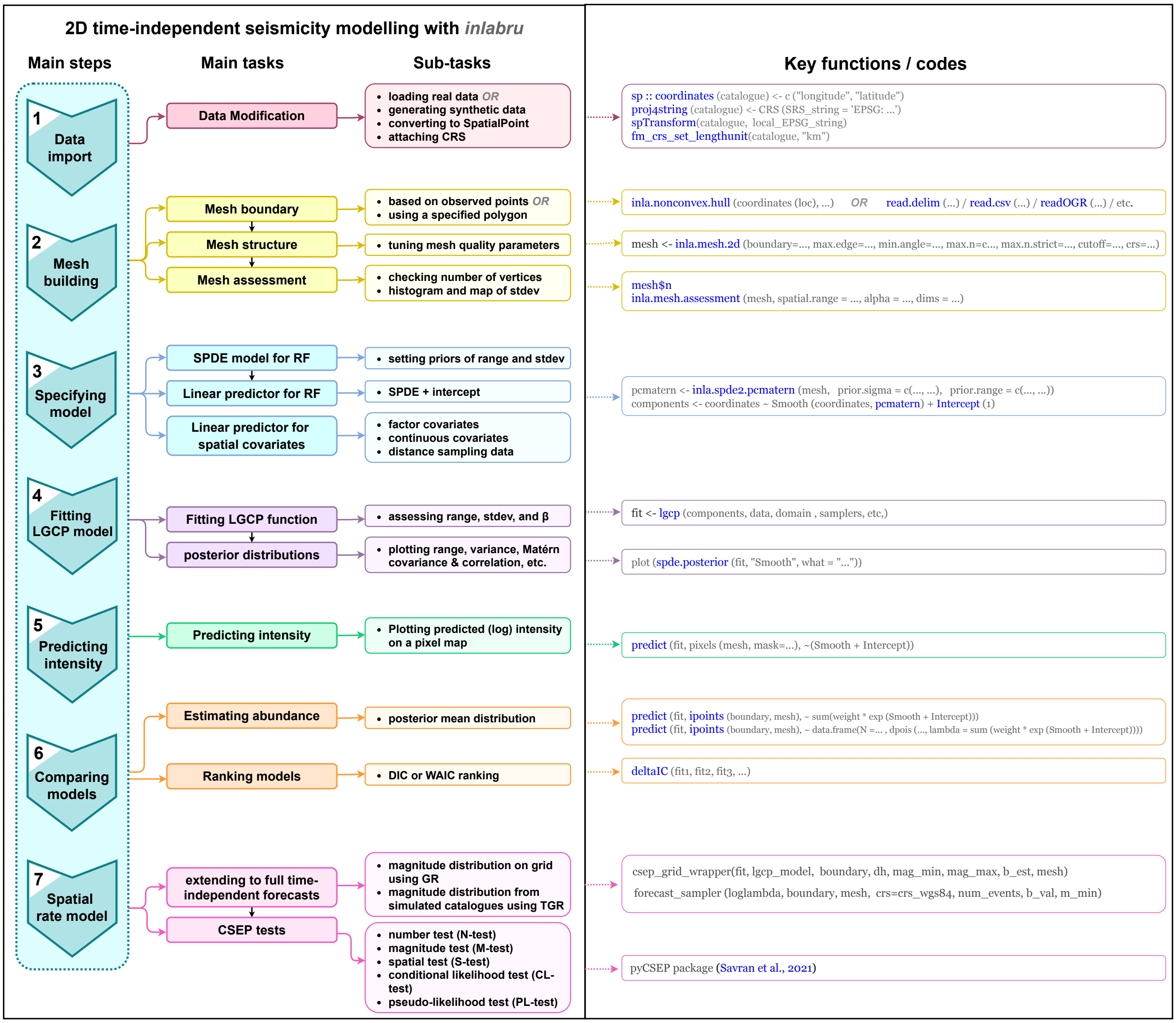


Figure 1. The workflow for generating spatial seismicity models in inlabru, with functions shown on the right.

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

understand which components were most useful in developing and improving such models. Here we test such models in pseudo-prospective mode for California, again using the approach of testing different combinations of data sets as input data. We develop a series of new spatial models to compare with the smoothed seismicity forecast of Helmstetter et al. (2007). These models contain a combination of four different covariates that were found to perform well in terms of DIC in Bayliss et al. (2020). These are shown in Figure 2 and include the GEM strain rate (Kreemer et al., 2014) (SR) map, NeoKinema model slip rates (NK) attached to mapped faults in the UCERF3 model (Field et al., 2014), a past seismicity model (MS) and a fault-distance map (FD) constructed using the UCERF3 fault geometry, with fault polygons buffered by their recorded dip. The past seismicity model used here is derived from events in the UCERF3 catalogue that occurred prior to 1984. For this data set, we fitted a model which contained only a Gaussian random field to the observed events, thus modelling the seismicity with a random field where we do not have to specify a smoothing kernel, the smoothing is an emergent property of the latent random field. This results in a smoothed seismicity map of events which occurred before our training dataset. This smoothed seismicity model also includes smaller magnitude events and those where the location or magnitude of the event is likely to be uncertain, so may account for some activity that is not observed or explicitly modelled (e.g. due to short-term clustering) at this time. Each of these components (SR, MS, NK, FD) is included as a continuous spatial covariate combined with a random field and intercept component. The M4.95+ events from 1984-2004 are used to construct the point process itself. The exact combination of components in a model is reflected in the model name as set out in Table 1. More details on each of these model components and their performance in describing locations of observed seismicity can be found in Bayliss et al. (2020). Step 7 of the workflow covers the steps described below and results presented here.

2.1 Developing full forecasts from spatial models

The inlabru models provide spatial intensity estimates which can be converted to spatial event rates by considering the timescales involved. Since the models we develop here are to be considered time-independent, we assume that the number of events expected in this time period is ‘scaleable’ in a straight-forward manner, consistent with a (temporally homogeneous) spatially-varying Poisson process. However we know that the rate of observed events is not Poissonian due to observed spatio-temporal clustering (Vere-Jones and Davies, 1966; Gardner and Knopoff, 1974) and that short time-scale spatial clustering can lead to higher rates anticipated in areas where large clusters have previously been recorded (Marzocchi et al., 2014). To test the impact of clustering on our forecasts, we include models made from both the full and declustered catalogues, assuming that the full catalogues might overestimate the spatial intensity due to observed spatio-temporal clustering and forecast higher rates in areas with recent spatial clustering. We decluster the catalogue by removing events allocated as aftershocks or foreshocks within the UCERF3 catalogue, which were determined by a (Gardner and Knopoff, 1974) clustering algorithm (UCERF3 appendix K). This results in 6 spatial models that we use from this point on, containing components as outlined in Table 1. The posterior mean of the log intensity for each of these models is shown in Figure 3. These models are constructed using an equal-area projection of California and converted to latitude and longitude only in the final step before testing. This figure represents the set of models formed by the training data set.

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

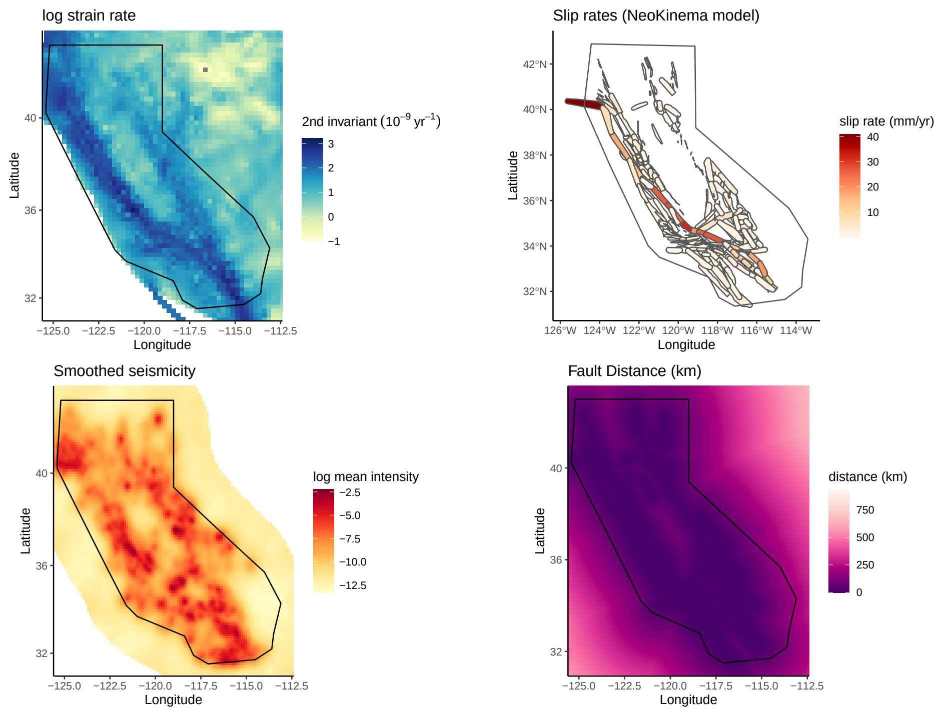


Figure 2. Input model covariates (clockwise from top left): GEM strain rate (SR), NeoKinema Slip rates from UCERF3 (NK), distance to nearest (UCERF3, dip and uniformly buffered) fault in km (FD), Smoothed seismicity from a Gaussian random field for events before 1984 (MS).

To extend this approach to a full forecast, we distribute magnitudes across the number of expected events according to a frequency-magnitude distribution. Given the small number of large events in the input training catalogue, a preference between a Tapered Gutenberg-Richter (TGR) or standard Gutenberg-Richter magnitude distribution with a rate parameter a, related to the intensity lambda, and an exponent b cannot be fully expressed. The choice of a b-value is not straightforward, as the b-value can be biased by several factors (Marzocchi et al., 2020) and is known to be affected by declustering (Mizrahi et al., 2021). In this case, we assume $b = 1$ for both clustered and declustered catalogues and for the TGR magnitude distribution we assume a corner magnitude of $M_c = 8$ for the California region proposed by (Bird and Liu, 2007) and used in the Helmstetter et al. (2007) models.

For the gridded forecasts (which assume a uniform event rate or intensity within the area of each square element), we use the posterior mean intensity as shown in Figure 3, transformed to a uniform grid of 0.1 x 0.1 latitude/longitude within the RELM region. We use latitude-longitude here as preferred by the pycsep tests. Magnitudes are then distributed across magnitude bins on a cell-by-cell basis according to the chosen magnitude-frequency distribution and the total rate expected in the cell. In this paper, we show GR magnitudes for the gridded forecasts. For the catalogue-based forecasts, we generate 10,000 samples from the full posteriors of the model components to establish 10,000 realisations of the model spatial intensity within the testing

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

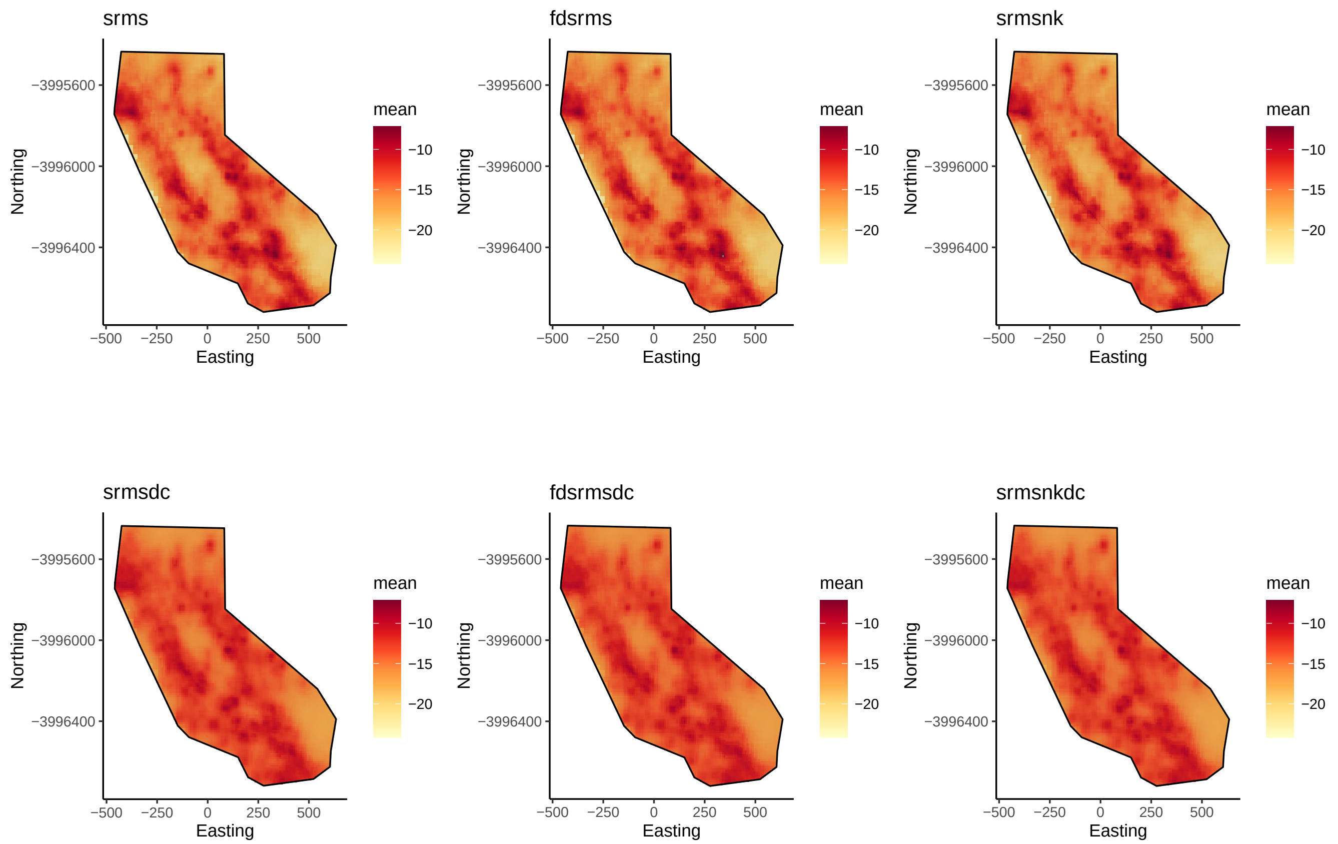


Figure 3. Posterior mean intensity for the six inlabru models created with full (top) and declustered (bottom) catalogues of events from 1985-2004.

polygon. We then sample a number of points consistent with the modelled intensity. In this case, we use the expected number of points given the mean intensity (as in step 6 in Figure 1) for one year, and randomly select an exact number of events for a simulated catalogue from a Poisson distribution about the mean rate, scaled to the number of years in the forecast. To sample events in a way that is consistent with modelled spatial rates, we sample many points and calculate the intensity value at the sampled points given the realisation of the model. We then implement a rejection sampler to retain points that have a significantly large intensity ratio compared to the largest intensity in the specific model realisation, with points retained only if the intensity ratio is greater than a uniform random variable between 0 and 1, that is points are retained with probability equal to $1 - \frac{\lambda_p}{\lambda_{max}}$. The set of retained points for each catalogue are then assigned a magnitude sampled from a TGR distribution, by methods described in Vere-Jones et al. (2001). Here we only sample magnuitudes from a TGR distribution in line with the

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

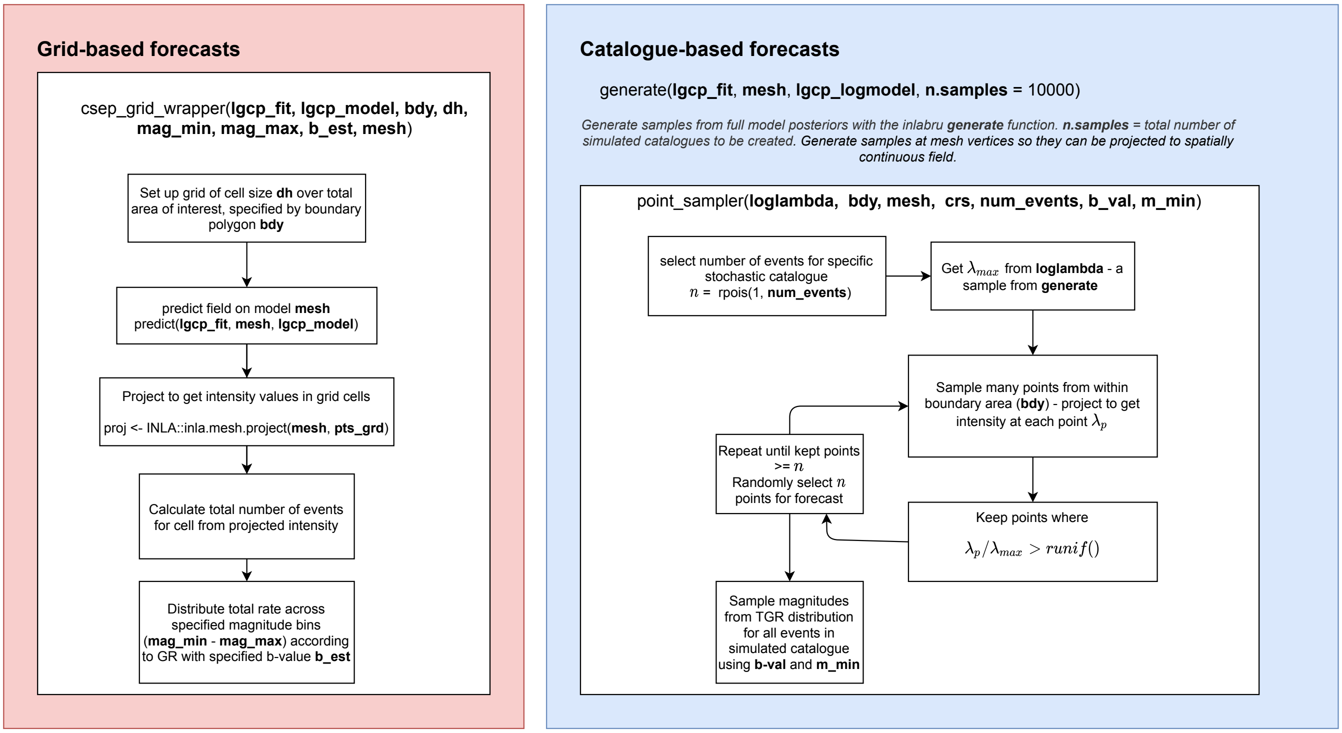


Figure 4. Schematic of the code for constructing grid-based (left) and simulated catalogue-based (right) earthquake forecasts given an inlabru LGCP intensity model. These represent step 7 of the workflow.

approach of Helmstetter et al. (2007), to allow a like for like comparison with this benchmark. A schematic diagram showing how grid and catalogue-based approaches are applied is shown in Figure 4.

2.2 CSEP tests

To test how well each forecast performs, we first test the consistency of the model forecasts, developed from data between 1984 and 2004, with observations from three subsequent and contiguous 5-year time periods, using standard CSEP tests for the number, spatial and magnitude distribution and conditional likelihood of each forecast. The original CSEP tests calculate a quantile score for the number (N), likelihood (L) (Schorlemmer et al., 2007) and spatial (S) and magnitude (M) (Zechar et al., 2010) tests, based on simulations that account for uncertainty in the forecast and a comparison of the observed and simulated likelihoods. We use 100 000 simulations of the forecasts to ensure convergence of the test results. The number test is the most straightforward, summing the rates over all forecast bins and comparing this with the total number of observed events. The quantile score is then the probability of observing at least N_{obs} events given the forecast, assuming a Poisson distribution of the number of events. Zechar et al. (2010) suggests using a modified version of the original N-test that tests the probability of a) at least N_{obs} events with score δ_1 and b) at most N_{obs} events with score δ_2 in order to test the range of events allowed by a forecast. Here we report both N-test quantile scores in line with this suggestion.

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

The likelihood test compares the performance of individual cells within the forecast. The likelihood of the observation given the model is described by a Poisson likelihood function in each cell and the total joint likelihood described by the product over all bins. The quantile score measures if the joint log-likelihood over many simulations falls within the tail of the observed likelihoods, with the score defined by the fraction of simulated joint log-likelihoods less than or equal to the observed. The conditional likelihood or CL test is a modification of the L-test developed due to the dependence of L-test results on the number of events in a forecast (Werner et al., 2010, 2011). The CL-test normalises the number of events in the simulation stage to the observed number of events in order to limit the effect of a significant mismatch in event number between forecast and observation. The magnitude and spatial tests compare the observed magnitude and spatial distributions by isolating these from the full likelihood. This is again achieved with a simulation approach and by summing and normalising over the other components. For the M-test, the sum is over the spatial bins while the S-test sums over all magnitude bins to isolate the respective components of interest. The final test statistic in both cases is again the fraction of observed log likelihoods within the range of the simulated log likelihood values. In all cases small values are considered inconsistent with the observations - we use a significance value of 0.05 for the likelihood-based tests and 0.025 for the number tests to be consistent with previous forecast testing experiments (Zechar et al., 2013).

In the new CSEP tests (Savran et al., 2020), the test distribution is determined from the simulated catalogues rather than a parametric likelihood function. For the N-test the construction of the test distribution is straightforward, being created from the number of events in each simulated catalogue and the quantile score calculated relative to this distribution. For the equivalent to the likelihood test a numerical, grid-based approximation to a point process likelihood is calculated (Savran et al., 2020). This is a more general approach than using the Poisson likelihood as in the grid-based tests, which penalises models that do not conform to a Poisson model. The distribution of pseudo-likelihood is then the collection of calculated pseudo-likelihood results for each simulated catalogue. The spatial and magnitude test distributions are derived from the pseudo-likelihood in a similar fashion to the grid-based approch, as explained in detail by Savran et al. (2020). The quantile scores are calculated similar to the original test cases, but because the simulations are based on the constructed pseudo-likelihood rather than a Poisson likelihood, the simulated-catalogue approach allows for forecasts which are overdispersed relative to a Poisson distribution. Similarly to the original tests, very small values will be considered inconsistent with the observations.

3 Full and declustered catalogue models

In constructing the three models both with and without clustering, we can examine relative contributions of the model components given differences in spatial intensity resulting from short-term spatio-temporal clustering. Table 1 shows the posterior mean component of the log intensity for each model both with and without clustering for M4.95+ seismicity, and the number of expected events per year for each model. The greatest contribution in the full-catalogue models comes from the strain rate (SR) for each model, with the past seismicity also making a significant contribution to the intensity. For the models where the catalogue has been declustered, the contribution to the posterior mean from the past seismicity is only slightly lower while the strain rate contribution is much smaller, effectively swapping the relative contributions of these components. This suggests

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

Table 1. Posterior means of model components and number of expected events for full and declustered (DC) models

	mean component contribution to log intensity				
Models	strain rate (SR)	past seismicity (MS)	slip rates (NK)	fault distance (FD)	N
SRMS	1.551	0.853	-	-	6.373
SRMSDC	0.415	0.777	-	-	3.679
SRMSNK	1.488	0.837	0.017	-	6.44
SRMSNKDC	0.425	0.779	0.001	-	3.79
FDSRMS	1.574	0.857	-	0.001	6.456
FDSRMSDC	0.491	0.784	-	0.004	3.737

that the strain rate component is more useful when considering the full earthquake catalogue than when the catalogue has been declustered. In both full- and declustered-catalogue models, the number of expected events is similar across all three models, thus we expect the models to perform similarly in the CSEP N-tests.

Figure 3 shows that the declustered-catalogue models (bottom row) appear much smoother than those constructed from the full catalogue, as they have lower intensity in areas with large seismic sequences in the training period. They also have a smaller range in intensity than the full catalogue models, with the (median) highest rates lower and the (median) lowest rates higher than the full catalogue models, meaning they cover less of the extremes at either end.

4 Model testing

We now test the models using the pyCSEP package for python (Savran et al., 2021). We begin with the standard (grid-based) CSEP test models described by Schorlemmer et al. (2007); Zechar et al. (2010) included in pyCSEP and described in section 2.2.

4.1 Grid-based forecast tests

We first compare the performance of our five-year forecasts, developed with a training window of 1984-2004, over the testing period 01/01/2006-01/01/2011 with the Helmstetter et al. (2007) forecast. In this time, the comcat catalog (https://earthquake.usgs.gov/data/comcat/). includes 32 M4.95+ events in the study region defined by the RELM polygon. All the models, regardless of their components or which catalogue is used, perform well in the magnitude tests due to the use of the GR distribution. The forecast tests are shown visually in Figure 5 and the quantile scores are reported in Table 2 for all tests and time-periods. A model is considered to pass a test if the quantile score is ≥ 0.05 for all tests except the N-test, where the significance level is set at ≥ 0.025 for both score components and the model fails if either score fails (Schorlemmer et al., 2010; Zechar et al., 2010). In Figure 4 the observed likelihood is shown as a coloured symbol (red circle for a failed test and green square for a passed one) and the forecast range is shown as a horizontal bar, for ease of comparison. In the number test (N-test), the declustered forecasts underpredict the number of expected events significantly in all cases due to the much smaller

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Table 2. Quantile scores for CSEP tests. Upper bounds for S, L and PL-tests, lower bound for N. Bold indicates consistency with observations, italics highlight declustered models.

		Gridded					Catalogue				
Time	Models	N-test (δ_1)	N-test (δ_2)	S-test	M-test	CL-test	N-test (δ_1)	N-test (δ_2)	S-test	M-test	PL-test
2006 - 2011	SRMS	**0.465**	**0.603**	0.025	**0.288**	**0.105**	**0.440**	**0.625**	**0.180**	**0.596**	**0.268**
	SRMSDC	*0.002*	*0.999*	***0.738***	***0.290***	***0.848***	*0.002*	*0.999*	***0.922***	***0.158***	*0.006*
	FDSRMS	**0.463**	**0.605**	0.032	**0.289**	**0.122**	**0.462**	**0.606**	**0.196**	**0.617**	**0.305**
	FDSRMSDC	*0.002*	*0.999*	***0.692***	***0.291***	***0.818***	*0.001*	*0.999*	***0.891***	***0.162***	*0.007*
	SRMSNK	**0.486**	**0.583**	0.028	**0.293**	**0.115**	**0.463**	**0.605**	**0.243**	**0.611**	**0.327**
	SRMSNKDC	*0.002*	*0.999*	***0.711***	***0.288***	***0.830***	*0.002*	*0.999*	***0.874***	***0.167***	*0.007*
2011 - 2016	SRMS	0.999	0	0.039	**0.158**	0.026	1	0	0.011	**1**	**0.999**
	SRMSDC	***0.963***	***0.064***	***0.766***	***0.153***	***0.485***	***0.952***	***0.081***	***0.807***	***0.963***	***0.951***
	FDSRMS	0.9999	0	0.030	**0.155**	0.021	1	0	0.011	**1**	**0.999**
	FDSRMSDC	***0.964***	***0.063***	***0.744***	***0.156***	***0.467***	***0.958***	***0.070***	***0.883***	***0.965***	***0.960***
	SRMSNK	0.999	0	**0.070**	**0.157**	**0.044**	0.999	0	0.0132	**0.999**	**0.9999**
	SRMSNKDC	***0.960***	***0.068***	***0.766***	***0.158***	***0.487***	***0.962***	***0.063***	***0.793***	***0.968***	***0.959***
2016 - 2021	SRMS	**0.792**	**0.264**	0	**0.369**	0.002	**0.772**	**0.283**	0.003	**0.564**	**0.312**
	SRMSDC	***0.029***	***0.982***	*0.008*	***0.368***	***0.068***	***0.019***	***0.989***	***0.081***	***0.141***	*0.005*
	FDSRMS	**0.791**	**0.266**	0	**0.367**	0.004	**0.795**	**0.268**	0.005	**0.602**	**0.369**
	FDSRMSDC	***0.030***	***0.981***	*0.0068*	***0.367***	***0.062***	*0.025*	*0.986*	***0.073***	***0.149***	0.007
	SRMSNK	**0.806**	**0.248**	0	**0.367**	0.006	**0.789**	**0.266**	0.006	**0.587**	**0.374**
	SRMSNKDC	***0.027***	***0.984***	0.005	***0.369***	***0.050***	***0.027***	***0.983***	***0.119***	***0.151***	0.009

number of expected events per year and the large number of events that actually occurred in the testing time period. In spatial testing (S-test), the full-catalogue models all perform poorly. In contrast, the declustered catalogue models all pass the S-test. In the conditional likelihood tests (CL-test), all of the models perform well and pass the CL-test (figure 5), with the declustered models performing better due to better spatial performance.

We then repeat the tests for two additional five year periods of California earthquakes illustrated in Figure 5. In all time windows, the M-test results remain consistent across all models. In the 2011-2016 period, there are 13 M4.95+ events within the RELM polygon, and this significant reduction in event number means that our full-catalogue models and the Helmstetter models all overestimate the actual number of events significantly, with the true number outwith the 95% confidence intervals of the models. In contrast, most of the models perform better in the S-test during this time period with the full catalogue slip-rate model and all declustered-catalogue models recording a passing quantile score (Table 2). Each of the models made with a declustered catalogue passes the CL-test.

In the 2016-2021 period (Figure 5 top) there are 30 M4.95+ events, which is within the confidence intervals shown for all tested models so all models pass the N-test for the first time. However none of the tested models pass the S-test due to the

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

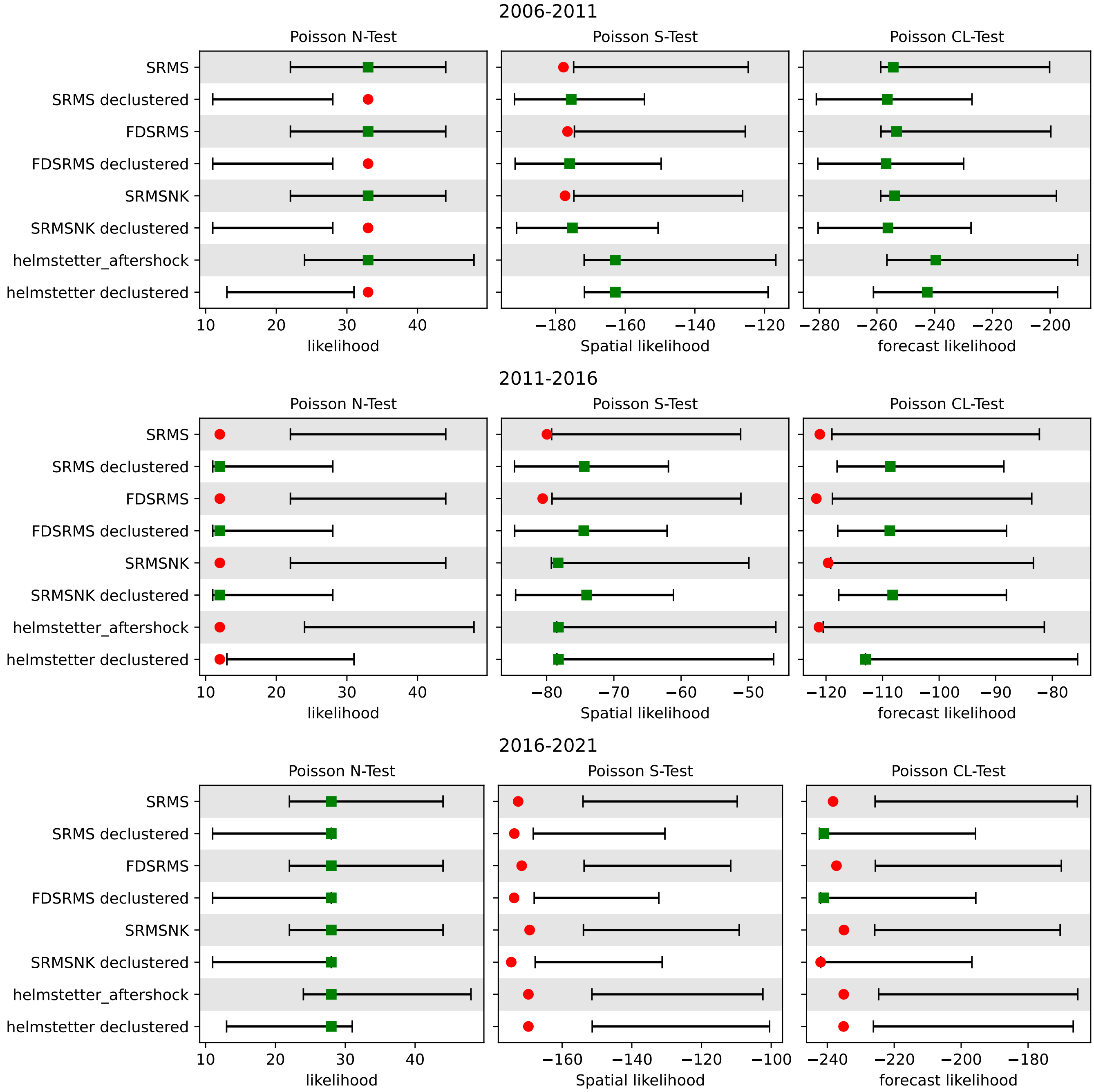


Figure 5. Grid-based forecast tests for all forecasts for three five year time periods: 2006-2011 (top), 2011-2016 (middle) and 2016-2021(bottom). The bars represent the 95% confidence interval derived from simulated likelihoods from the forecast, while the symbol represents the observed likelihood for observed events. The green square identifies that a model has passed the test and a red circle indicates inconsistency between forecast and observation. The forecasts are compared to both the full (Helmstetter aftershock) and declustered models of Helmstetter et al (2007)

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

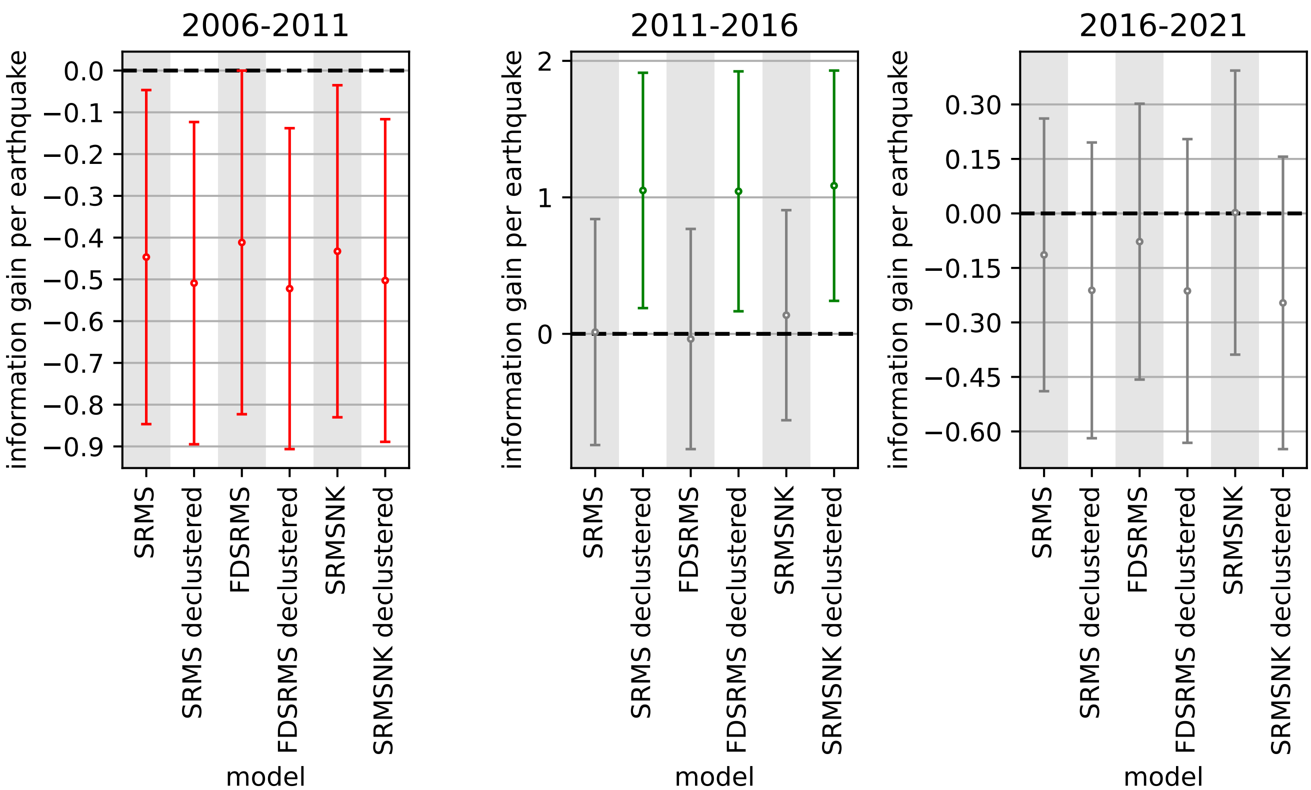


Figure 6. T-test results for the inlabru models showing information gain per earthquake relative to the full Helmstetter et al (2007) model (helmstetter aftershock in Fig. 5) for three time periods. Red indicates forecasts are worse in terms of information gain and green indicates forecasts performing better than the benchmark forecast. Grey forecasts are not significantly different in terms of information gain.

spatial distribution of the events in this time period being highly clustered in areas without exceptionally high rates, even for models developed from the full catalogue. The CL-test results for the 2016-2021 period show that none of the models perform particularly well in this time period, with two of the declustered-catalogue models passing the test only barely.

These statistical tests (N, S, M and CL) investigate the consistency of a forecast made during the training window with the observed outcome. They do not compare the performance of models directly with each other, but rather with observed events. One method of comparing forecasts is by considering their information gain relative to a fixed model with a paired T-test (Rhoades et al., 2011). Here, we implement the paired T-test for the gridded forecast to test their performance against the Helmstetter et al. (2007) aftershock forecast as a benchmark, because it performed best in comparison to other RELM models in previous CSEP testing over various timescales (Strader et al., 2017). The results of the comparison are shown in Figure 6. For the first time period (2006-2011), the models perform similarly in terms of information gain, and all of the inlabru models perform worse than the Helmstetter model. For the 2011-2016 period, the inlabru models developed from the declustered catalogues perform better in terms of information gain than those developed from the full catalogue and significantly better than

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences Discussions

the Helmstetter model. In the most recent testing period (2016-2021), the inlabru models have an information gain range that includes the Helmstetter model. Together these results imply the inlabru models provide a positive and significant information gain on a 5-10 year time period after the end of the training period for declustered-catalogue models, and not otherwise.

4.2 Simulated-catalogue forecasts

Our second stage of testing uses simulated catalogues in order to make use of the newer CSEP tests (Savran et al., 2020). We use the number, spatial and pseudolikelihood (PL) tests to evaluate these forecasts, with the PL test replacing the grid-based L-test. In our case, as described above the number of events in the simulated catalogues is inherently Poisson due to the way they are constructed, but the spatial distribution is perturbed from a homogeneous Poisson distribution due to the contributions of model covariates and the random field itself (e.g. see equation 1, where a homogenous Poisson process would include only the intercept term β_0) and the parameter values are sampled from the posterior at each simulation, so vary from simulation to simulation. Figure 7 shows the test distributions for each forecast as a letter-value plot (Hofmann et al., 2011), an extended boxplot which includes more quantiles of the distribution until the quantiles become too uncertain to discriminate. This allows us to understand more of the full distribution of model pseudo-likelihood than a standard quantile range or boxplot, while allowing easy comparisons between the results for different forecast models.

We expect the grid-based and simulated-catalogue approaches to have similar results in terms of the magnitude (M) tests due to the similarity of magnitude distributions used in construction, and all models do similarly well in this test (Table 2). Similarly, we do not expect significant differences in the number tests with this approach, since our method of determining the number of events will result in a Poisson distribution of the number of events. However, since the number of events varies in each synthetic catalogue we can look at the distribution of the number of events in the synthetic data produced by the ensemble of forecast catalogs relative to the observed number. This is shown in the left panel of Figure 7, with the observed number of events for each time period shown with a dashed line. Again, the declustered models do better in the 2011-2016 period, though it is clear the observed number of events is low even for them.

We might expect the most noticeable differences to occur in the spatial test, because it measures the spatial component consistency with observed events and because we are now using the full posterior distribution of spatial components, and therefore potentially allowing more variation in the observed spatial models. The middle panel of Figure 7 shows the spatial likelihood distribution constructed from simulated catalogues.

Similar to the grid-based examples, for the 2006-2011 period (red star indicator) the spatial performance of the SRMS and FDSRMS models is better when the full, rather than declustered catalogue, has been used in model construction.

All of the models pass the S-test when considering quantile scores in this time period. Similarly, when testing the 2011-2016 period (test statistic shown with a blue diamond), all of the models built from the declustered catalogue pass the S-test, while the full-catalogue models do more poorly. In 2016-2021 (green circle), the spatial performance of all models is again poor. The best-performing model in this time period is the FDSRMS-declustered model (Table 2), with the declustered-catalogue models generally doing better than the full-catalogue models.

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

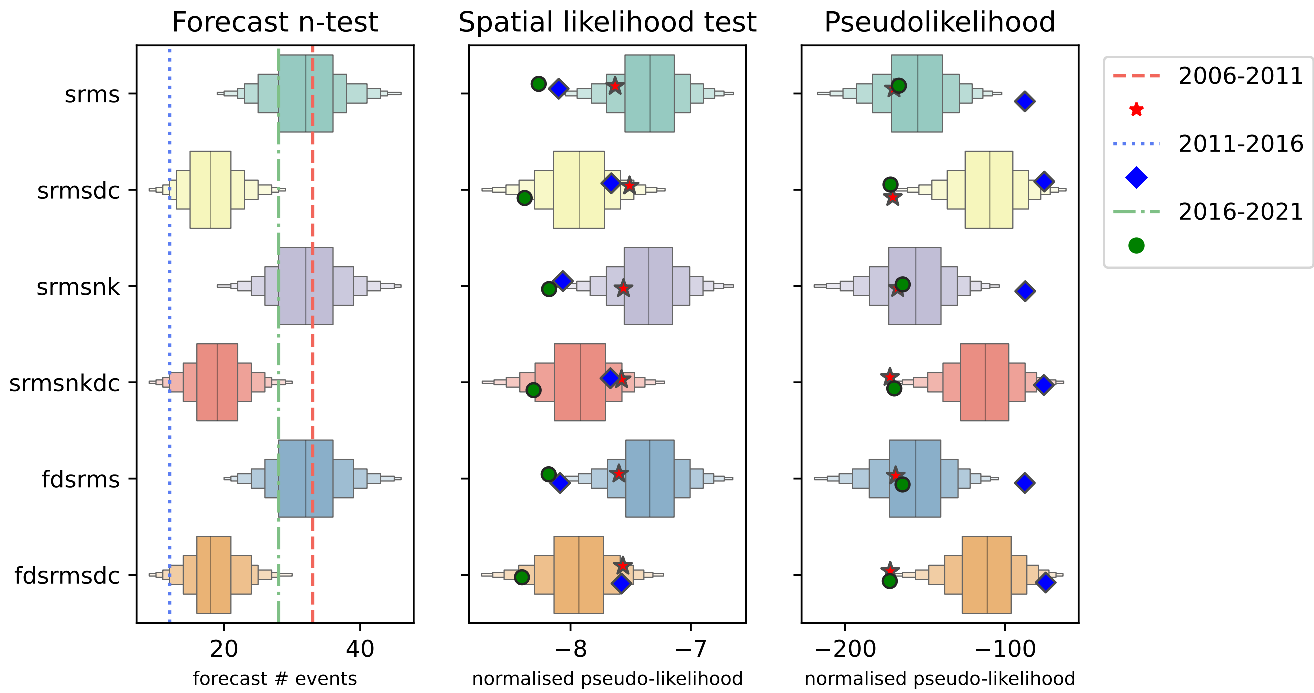


Figure 7. N-test, S-test and pseudo-likelihood results for each of the 6 inlabru models when forecasts are generated from 10 000 synthetic catalogues sampling from the full inlabru model posteriors. For the n-test, the number of observed events for the 2006-2011, 2011-2016 and 2016-2021 are shown by the red, blue and green dashed lines respectively. For the S- and Pseudo-likelihood tests, the observed test statistic for each time period is shown as a symbol (red star for 2006-2011, blue diamond for 2011-2016 and green circle for 2016-2021)

Finally, the pseudo-likelihood test (Figure 7, right) incorporates both spatial and rate components of the forecast, much like the grid-based likelihood. For the inlabru models, the preference between the models for the full and declustered catalogues changes with time period with both sets of models doing poorly in the 2016-2021 period (green circle). All of the full-catalogue models pass in 2006-2011 and surprisingly in 2011-2016 and 2016-2021, though some of the quantile scores are again quite large and in the upper tails of the likelihood distributions. Like the grid-based likelihood test, the pseudo-likelihood test penalises for the number of events in the forecast, which allows the full-catalogue models to pass the pseudo-likelihood test even when they have poor spatial performance, as in the 2011-2016 and 2016-2021 testing periods.

5 Discussion

5.1 Number of events

While the full-catalogue models performed well in the tests for the first five-year time window, the other two sets of test results were less promising. This can be largely explained by the number of events that occurred in the 10 year period from 2006-2016

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

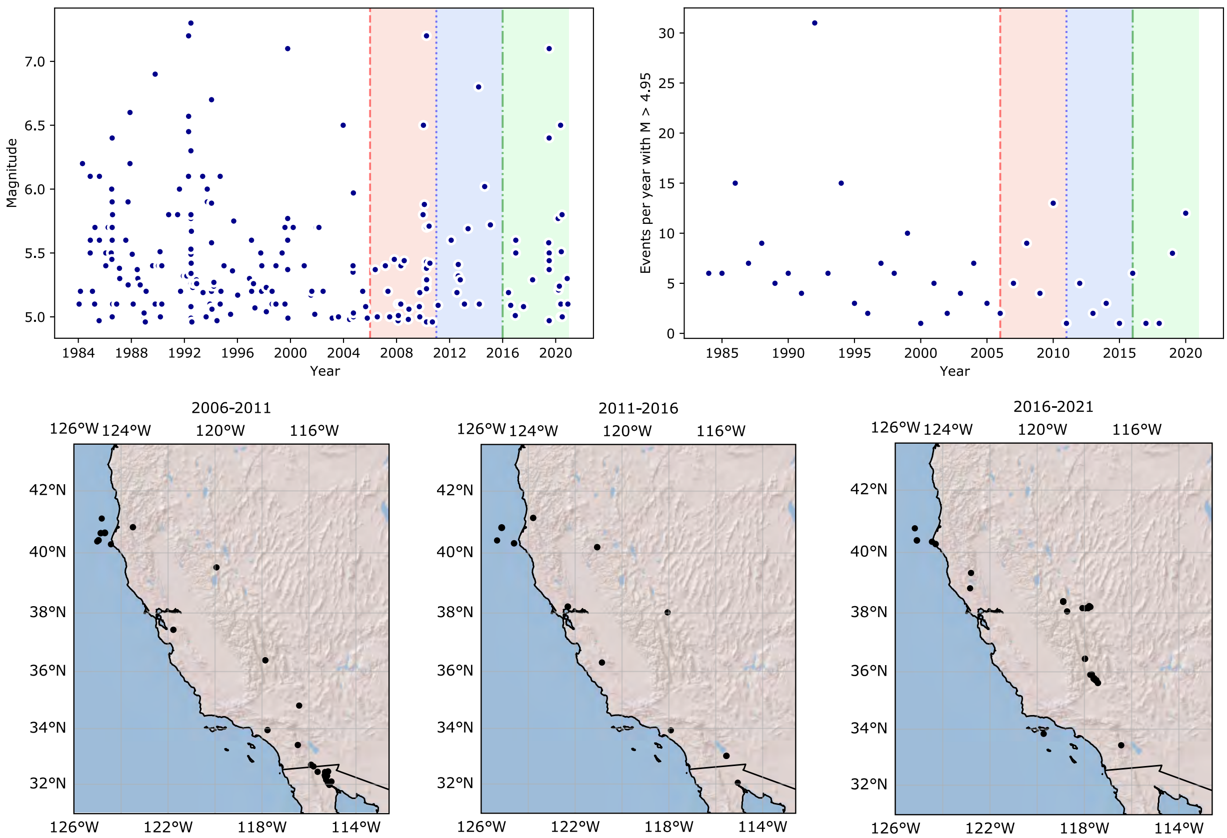


Figure 8. Top: Catalogue of events in California from 1985-2021. The period 1984-2004 is used for model construction, and the three testing periods are shown with red, blue and green backgrounds. The left panel shows the magnitude of events in time and the right the number of events in each year. Bottom: the comcat catalogues for the three five-year testing intervals.

(red and blue backgrounds in Figure 8, top right). In this time 45 events were recorded in the comcat catalog, compared to 32 events in the five years between 2006-2011. In the twenty years from 1984-2004 used in our model construction, a total of 156 events with $M > 4.95$ were recorded, which is an average of 7.8 events/year. Helmstetter et al. (2007) explicitly use the average number events per year with magnitude > 4.95 (7.38) to condition their models. It is therefore not surprising that the declustered forecasts perform oppositely, with poor performance in the 2006-2011 time period and better performances in the 2011-2016 time period when fewer events occurred. This is a common issue in CSEP testing, reported both in Italy when the five-year tests occurred in a time period with a large cluster of events in a historically low-seismicity area (Taroni et al., 2018) and in New Zealand, where the Canterbury earthquake sequence occurred in the middle of the CSEP testing period (Rhoades et al., 2018) resulting in significantly more events than expected. Strader et al. (2017) found that four of the original RELM forecasts overpredicted the number of events in the 2006-2011 time window and 11 overpredicted the number of events in the second 5-year testing window (2011-2016), including the Helmstetter model. Overall, the inlabru model N-test results were comparable to the Helmstetter model performance in the grid-based assessment and performed well at forecasting at least the minimum number of events in all but the declustered models in the first testing period (table 2).

5.2 Full- and declustered-catalogue models

We did not filter for mainshocks in the observed events, so we might expect the N-test results for the declustered models to do poorly, but they were consistent with observed behaviour in 2 of the 3 tested time periods in both the grid-based and catalogue testing. If we consider only the lower bound of the N-test, the declustered models pass the test in the full 2011-2021 time period and only perform poorly in 2006-2011, a time period which arguably contained many more than average events (Figure 8). Similarly, the full catalogue models do poorly on the upper N-test in 2011-16 but otherwise pass in time windows with higher numbers of events.

The declustered models pass spatial tests more often than the full catalogue models because they are less affected by recent clustering, and perhaps benefit from being smoother overall than the full-catalogue models (Figure 3). The superior performance of the declustered models may not have been entirely obvious had we tested only the 2006-2011 period and relied solely on the ‘pass’ criterion from the full suite of tests: only the full-catalogue synthetic catalogue forecast models get a pass in all consistency tests in this time period. This highlights a need for forecast to be assessed over different timescales in order to truly understand how well they perform, a point previously raised by Strader et al. (2017) when assessing the RELM forecasts, and more generally embedded in the evaluation of forecasting power since the early calculations of Lorenz (1963) for a simple but nonlinear model for Earth’s atmosphere in meteorological forecasting. .

We conclude that neither a full nor declustered catalogue necessarily gives a better estimate of the future number of events in any 5-year time-period, though the declustered models tend to perform better spatially, and may be more suitable for longer-term forecasting. Given different declustering methods may retain different specific events and different total numbers of events, different declustering approaches may lead to significant differences in model performances, especially in time periods with a small number of events in the full catalogue. To truly discriminate between which approach is best, a much longer testing time frame would be needed to ensure a suitably large number of events.

5.3 Spatial performance of gridded and simulated catalogue forecasts

In general, the simulated catalogue-based forecasts were more likely to pass the tests than the gridded models. This is most obvious in the first testing period, when the simulated catalogue-based models based on the full-catalogue passed all tests and those for the declustered catalogues only fail due to the smaller expected number of events. Similarly, in the most recent testing period (2016-2021) the simulated-catalogue forecasts are able to just pass the S-test where all models fail in the gridded approach.

The simulated catalogue approach allows us to consider more aspects of the uncertainty in our model. For example, we could further improve upon this by considering potential variation in the b-value in the ensemble catalogues which arises from magnitude uncertainties, an issue that may be particularly relevant when dealing with homogenised earthquake catalogues (Griffin et al., 2020) or where the b-value of the catalogue is more uncertain (Herrmann and Marzocchi, 2020).

5.4 Roadmap - where next?

The main limitation of the work presented here, and many other forecast methodologies, is how aftershock events are handled. Our choice of (a relatively high) magnitude threshold for modelling may have also benefited the full model by ignoring many small magnitude events that would be removed by a formal declustering procedure. The real solution to this is to formally model the clustering process.

The approach presented here conforms strongly with current practice. In time-independent forecasting and PSHA, catalogues are routinely declustered to be consistent with Poisson occurrence assumptions. Operational forecasting already relies heavily on models such as the epidemic type aftershock sequence model(ETAS, Ogata (1988)) to handle aftershock clustering (Marzocchi et al., 2014), but few attempts have been made to account for background spatial effects beyond a simple continuous Poisson rate. The exceptions to this are changes to the spatial components of ETAS models (Bach and Hainzl, 2012), the recent developments in spatially-varying ETAS (Nandan et al., 2017) and extensions to the ETAS model that also incorporate spatial covariates (Adelfio and Chiodi, 2020). However, the more versatile inlabru approach allows for more complex spatial models than has yet been implemented with these approaches. The inlabru approach also provides a general framework to test the importance of different covariates in the model, and a fully Bayesian method for forecast generation as we have implemented here.

One way to handle these conflicts is to model the seismicity formally as a Hawkes process, where the uncertainty in the tradeoff between the background and clustered components is explicit and can be formally accounted for. In future work we will modify the workflow of Figure 1 to test the hypothesis that this approach will improve the ability for inlabru to forecast using both time-independent and time-dependent models.

6 Conclusions

For the first time, we present time independent forecasts for California developed with inlabru. We developed three earthquake forecasts for California considering different combinations of spatial covariates and developed with both the full and declustered catalogue in each case, resulting in 6 models in total. These models each include spatial covariates that perform well in retrospective testing of spatial seismicity, which are then extended to spatio-temporal models by considering the frequency-magnitude distribution and assuming a Poisson distribution of events in time. The full-catalogue models each pass the standard CSEP tests for number, magnitude and spatial distribution, and perform favourably with the Helmstetter model tested in the original RELM experiment over the 2006-2011 period, demonstrating the suitability of inlabru models for time-independent earthquake forecasting. The declustered catalogues perform less well in this time period due to the lower expected number of events, but perform better in spatial tests and overall in the 2011-2016 time period, where the full catalogue models overestimate the number of events quite significantly. Neither the full-catalogue or declustered-catalogue models perform well in

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

the most recent testing period, with much worse spatial performance. Simulated catalogue forecasts that make use of the full posteriors of the model pass consistency tests more often than their grid-based equivalents by better accounting for uncertainty in the model itself.

Code and data availability. The code and data required to produce all of the results in this paper, including figures, can be downloaded from https://doi.org/10.5281/zenodo.5793157

Author contributions. Kirsty Bayliss developed the methodology, carried out the formal analysis and interpretation, and wrote the first draft of the paper. Farnaz Kamranzad contributed significantly to visualisation, particularly development of Figure 1. Mark Naylor and Ian Main contributed to the conceptual design, the interpretation of the results, and the writing of the paper. All authors contributed to paper review and drafting.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is funded by the Real-time Earthquake Risk Reduction for a Resilient Europe ’RISE’ project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 821115. We thank Francesco Serafini and Finn Lindgren for helpful discussions and suggestions.

https://doi.org/10.5194/nhess-2021-403
Preprint. Discussion started: 4 January 2022

References

Adelfio, G. and Chiodi, M.: Including covariates in a space-time point process with application to seismicity, Statistical Methods & Applications, https://doi.org/10.1007/s10260-020-00543-5, 2020.

Bach, C. and Hainzl, S.: Improving empirical aftershock modeling based on additional source information, Journal of Geophysical Research: Solid Earth, 117, n/a–n/a, https://doi.org/10.1029/2011JB008901, 2012.

Bachl, F. E., Lindgren, F., Borchers, D. L., and Illian, J. B.: inlabru: an R package for Bayesian spatial modelling from ecological survey data, Methods in Ecology and Evolution, 10, 760–766, https://doi.org/10.1111/2041-210X.13168, 2019.

Bayliss, K., Naylor, M., Illian, J., and Main, I. G.: Data-Driven Optimization of Seismicity Models Using Diverse Data Sets: Generation, Evaluation, and Ranking Using Inlabru, Journal of Geophysical Research: Solid Earth, 125, e2020JB020 226, https://doi.org/https://doi.org/10.1029/2020JB020226, 2020.

Bird, P. and Liu, Z.: Seismic Hazard Inferred from Tectonics: California, Seismological Research Letters, 78, 37–48, https://doi.org/10.1785/gssrl.78.1.37, 2007.

Field, E. H.: Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM), Seismological Research Letters, 78, 7–16, https://doi.org/10.1785/gssrl.78.1.7, 2007.

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon, R. J., and Zeng, Y.: Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model, Bulletin of the Seismological Society of America, 104, 1122–1180, https://doi.org/10.1785/0120130164, 2014.

Gardner, J. K. and Knopoff, L.: Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, Bulletin of the Seismological Society of America, 64, 1363–1367, https://doi.org/10.1785/0120160029, 1974.

Griffin, J. D., Allen, T. I., and Gerstenberger, M. C.: Seismic Hazard Assessment in Australia: Can Structured Expert Elicitation Achieve Consensus in the "Land of the Fair Go"?, Seismological Research Letters, 91, 859–873, https://doi.org/10.1785/0220190186, 2020.

Helmstetter, A., Kagan, Y. Y., and Jackson, D. D.: High-resolution Time-independent Grid-based Forecast for M >= 5 Earthquakes in California, Seismological Research Letters, 78, 78–86, https://doi.org/10.1785/gssrl.78.1.78, 2007.

Herrmann, M. and Marzocchi, W.: Inconsistencies and Lurking Pitfalls in the Magnitude–Frequency Distribution of High-Resolution Earthquake Catalogs, Seismological Research Letters, 92, 909–922, https://doi.org/10.1785/0220200337, 2020.

Hofmann, H., Kafadar, K., and Wickham, H.: Letter-value plots: Boxplots for large data, 2011.

Jordan, T. H. and Jones, L. M.: Operational Earthquake Forecasting: Some Thoughts on Why and How, Seismological Research Letters, 81, 571–574, https://doi.org/10.1785/gssrl.81.4.571, 2010.

Kreemer, C., Blewitt, G., and Klein, E. C.: A geodetic plate motion and Global Strain Rate Model, Geochemistry, Geophysics, Geosystems, 15, 3849–3889, https://doi.org/10.1002/2014GC005407, 2014.

Lorenz, E. N.: Deterministic Nonperiodic Flow, Journal of Atmospheric Sciences, 20, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Marzocchi, W., Zechar, J. D., and Jordan, T. H.: Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting, Bulletin of the Seismological Society of America, 102, 2574–2584, https://doi.org/10.1785/0120110327, 2012.

Marzocchi, W., Lombardi, A. M., and Casarotti, E.: The Establishment of an Operational Earthquake Forecasting System in Italy, Seismological Research Letters, 85, 961–969, https://doi.org/10.1785/0220130219, 2014.

Marzocchi, W., Spassiani, I., Stallone, A., and Taroni, M.: How to be fooled searching for significant variations of the b-value, Geophysical Journal International, https://doi.org/10.1093/gji/ggz541, 2020.

Mizrahi, L., Nandan, S., and Wiemer, S.: The Effect of Declustering on the Size Distribution of Mainshocks, Seismological Research Letters, https://doi.org/10.1785/0220200231, 2021.

Nandan, S., Ouillon, G., Wiemer, S., and Sornette, D.: Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to California, Journal of Geophysical Research: Solid Earth, 122, 5118–5143, https://doi.org/10.1002/2016JB013266, 2017.

Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes, Journal of the American Statistical Association, 83, 9–27, https://doi.org/10.1080/01621459.1988.10478560, 1988.

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., and Imoto, M.: Efficient testing of earthquake forecasting models, Acta Geophysica, 59, 728–747, https://doi.org/10.2478/s11600-011-0013-5, 2011.

Rhoades, D. A., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., Schorlemmer, D., Werner, M. J., and Jordan, T. H.: Regional Earthquake Likelihood Models II: Information Gains of Multiplicative Hybrids, Bulletin of the Seismological Society of America, 104, 3072–3083, https://doi.org/10.1785/0120140035, 2014.

Rhoades, D. A., Christophersen, A., and Gerstenberger, M. C.: Multiplicative Earthquake Likelihood Models Based on Fault and Earthquake Data, Bulletin of the Seismological Society of America, 105, 2955–2968, https://doi.org/10.1785/0120150080, 2015.

Rhoades, D. A., Christophersen, A., Gerstenberger, M. C., Liukis, M., Silva, F., Marzocchi, W., Maximilian, J., and Jordan, T. H.: Highlights from the First Ten Years of the New Zealand Earthquake Forecast Testing Center, Seismological Research Letters, 89, 1229–1237, https://doi.org/10.1785/0220180032, 2018.

Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, Journal of the Royal Statistical Society. Series B: Statistical Methodology, 71, 319–392, https://doi.org/10.1111/j.1467-9868.2008.00700.x, 2009.

Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., and Michael, A.: Pseudoprospective Evaluation of UCERF3-ETAS Forecasts during the 2019 Ridgecrest Sequence, Bulletin of the Seismological Society of America, 110, 1799–1817, https://doi.org/10.1785/0120200026, 2020.

Savran, W. H., Werner, M. J., Schorlemmer, D., and Maechling, P. J.: pyCSEP - Tools for Earthquake Forecast Developers, https://github.com/SCECcode/pycsep, 2021.

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., and Rhoades, D. A.: Earthquake Likelihood Model Testing, Seismological Research Letters, 78, 17–29, https://doi.org/10.1785/gssrl.78.1.17, 2007.

Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., and Jordan, T. H.: First Results of the Regional Earthquake Likelihood Models Experiment, Pure and Applied Geophysics, 167, 859–876, https://doi.org/10.1007/s00024-010-0081-5, 2010.

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., Mak, S., Rhoades, D. A., Gerstenberger, M. C., Hirata, N., Liukis, M., Maechling, P. J., Strader, A., Taroni, M., Wiemer, S., Zechar, J. D., and Zhuang, J.: The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities, Seismological Research Letters, 89, 1305–1313, https://doi.org/10.1785/0220180053, 2018.

Strader, A., Schneider, M., and Schorlemmer, D.: Prospective and retrospective evaluation of five-year earthquake forecast models for California, Geophysical Journal International, 211, 239–251, https://doi.org/10.1093/gji/ggx268, 2017.

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., and Euchner, F.: Prospective CSEP Evaluation of 1-Day, 3-Month, and 5-Yr Earthquake Forecasts for Italy, Seismological Research Letters, 89, 1251–1261, https://doi.org/10.1785/0220180031, 2018.

Vere-Jones, D. and Davies, R. B.: A statistical survey of earthquakes in the main seismic region of New Zealand, New Zealand Journal of Geology and Geophysics, 9, 251–284, https://doi.org/10.1080/00288306.1966.10422815, 1966.

Vere-Jones, D., Robinson, R., and Yang, W.: Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation, Geophysical Journal International, 144, 517–531, https://doi.org/10.1046/j.1365-246x.2001.01348.x, 2001.

Werner, M. J., Zechar, J. D., Marzocchi, W., and Wiemer, S.: Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, Annals of Geophysics, 53, https://doi.org/10.4401/ag-4840, 2010.

Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y.: High-Resolution Long-Term and Short-Term Earthquake Forecasts for California, Bulletin of the Seismological Society of America, 101, 1630–1648, https://doi.org/10.1785/0120090340, 2011.

Zechar, D. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., and Jordan, T. H.: Regional Earthquake Likelihood Models I: First-order results, Bulletin of the Seismological Society of America, 103, 787–798, https://doi.org/10.1785/0120120186, 2013.

Zechar, J. D., Gerstenberger, M. C., and Rhoades, D. A.: Likelihood-Based Tests for Evaluating Space–Rate–Magnitude Earthquake Forecasts, Bulletin of the Seismological Society of America, 100, 1184–1195, https://doi.org/10.1785/0120090192, 2010.

ADVANCING EARTH AND SPACE SCIENCE

JGR Solid Earth

RESEARCH ARTICLE
10.1029/2021JB023897

Key Points:

- We compile 148 afterslip studies after 53 earthquakes to probe the scaling and variability of aseismic afterslip moment
- Afterslip moment scales near-linearly with the coseismic moment but varies from <1% to >300% of coseismic moment between earthquakes
- Different modeling approaches appear to be the dominant cause of variability, but fault slip rate and rupture shape may exert some control

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. M. Churchill,
robert.churchill@bristol.ac.uk

Citation:

Churchill, R. M., Werner, M. J., Biggs, J., & Fagereng, Å. (2022). Afterslip moment scaling and variability from a global compilation of estimates. *Journal of Geophysical Research: Solid Earth*, *127*, e2021JB023897. https://doi.org/10.1029/2021JB023897

Received 23 DEC 2021
Accepted 2 APR 2022

Afterslip Moment Scaling and Variability From a Global Compilation of Estimates

R. M. Churchill[1], M. J. Werner[1], J. Biggs[1], and Å. Fagereng[2]

[1]School of Earth Sciences, University of Bristol, Bristol, UK, [2]School of Earth and Environmental Sciences, Cardiff University, Cardiff, UK

Abstract Aseismic afterslip is postseismic fault sliding that may significantly redistribute crustal stresses and drive aftershock sequences. Afterslip is typically modeled through geodetic observations of surface deformation on a case-by-case basis, thus questions of how and why the afterslip moment varies between earthquakes remain largely unaddressed. We compile 148 afterslip studies following 53 M_w6.0–9.1 earthquakes, and formally analyze a subset of 88 well-constrained kinematic models. Afterslip and coseismic moments scale near-linearly, with a median Spearman's rank correlation coefficient (CC) of 0.91 after bootstrapping (95% range: 0.89–0.93). We infer that afterslip area and average slip scale with coseismic moment as $M_o^{2/3}$ and $M_o^{1/3}$, respectively. The ratio of afterslip to coseismic moment (M_{rel}) varies from <1% to >300% (interquartile range: 9%–32%). M_{rel} weakly correlates with M_o (CC: −0.21, attributed to a publication bias), rupture aspect ratio (CC: −0.31), and fault slip rate (CC: 0.26, treated as a proxy for fault maturity), indicating that these factors affect afterslip. M_{rel} does not correlate with mainshock dip, rake, or depth. Given the power-law decay of afterslip, we expected studies that started earlier and spanned longer timescales to capture more afterslip, but M_{rel} does not correlate with observation start time or duration. Because M_{rel} estimates for a single earthquake can vary by an order of magnitude, we propose that modeling uncertainty currently presents a challenge for systematic afterslip analysis. Standardizing modeling practices may improve model comparability, and eventually allow for predictive afterslip models that account for mainshock and fault zone factors to be incorporated into aftershock hazard models.

Plain Language Summary Afterslip is the gentle slipping, or sliding, of a fault over several months or years following an earthquake. Afterslip is not an earthquake but does release energy that may trigger other earthquakes, called aftershocks. Therefore, we wish to understand why some earthquakes produce much more afterslip than others. We compile and analyze 148 afterslip studies (following 53 earthquakes) from the literature and show that the amount of afterslip is mainly determined by the magnitude of the earthquake. However, there is considerable variation beyond this dependence which might be linked to characteristics of the earthquake and fault setting. We find that more afterslip tends to occur when the earthquake rupture is less elongated in shape, or when the causative fault has a greater long-term slip rate. However, different studies following the same earthquake sometimes yielded different moment estimates that we cannot explain. We propose that the unknowns and methodological differences in afterslip modeling currently make comparing events difficult; future methods should be more standardized so that afterslip can be meaningfully considered in hazard models following an earthquake.

1. Introduction

Following an earthquake, various postseismic mechanisms act to relax and redistribute stress concentrations in the crust and upper mantle (Freed, 2005). In addition to seismic aftershocks, postseismic mechanisms include aseismic afterslip (Bürgmann et al., 1997; Marone et al., 1991; Shen et al., 1994), pore fluid flow or poroelastic rebound (Peltzer et al., 1996, 1998; Piombo et al., 2005) and deeper viscoelastic relaxation or viscous flow (Deng et al., 1998; Freed & Lin, 2001; F. F. Pollitz et al., 2001). These processes may generate geodetically observable surface deformation (e.g., with GNSS, InSAR), which can be modeled to provide insight into fault zones, crustal structure, and the earthquake cycle (e.g., Ingleby & Wright, 2017; Massonnet et al., 1994).

Aseismic afterslip may provide particularly valuable insight into fault zone rheology and earthquake cycle processes (Avouac, 2015; Bürgmann, 2018). Afterslip is transient, fault scale, aseismic shear that occurs on and close to the fault planes of the parent earthquake, as postseismic readjustment (Avouac, 2015; Harris, 2017), distinct from generally deeper and more distributed viscoelastic relaxation (K. Wang et al., 2012). Aseismic afterslip is also distinct from seismic aftershocks and is specifically a response to coseismic stress concentrations, thus is also a distinct mechanism from triggered slow slip, which is driven by stresses that have built up over longer timescales (Bürgmann, 2018). Afterslip is globally widespread and relatively easy to detect as the associated surface deformation is initially greater and more near-field than that caused by viscoelastic relaxation (Diao et al., 2014; Reilinger et al., 2000). There is also mounting evidence that afterslip may drive aftershock sequences (Bürgmann et al., 2002; Hsu et al., 2006; Huang et al., 2017; Peng & Zhao, 2009; Perfettini & Avouac, 2004), therefore it is highly desirable to better understand the phenomenon.

First order behaviors of afterslip, such as the scaling of afterslip moment with coseismic moment, are still poorly understood. Some existing studies have considered afterslip following multiple earthquakes but have been limited in scope. For example, Lange et al. (2014) compared afterslip models for three large subduction thrust events, whilst Hawthorne et al. (2016) and Alwahedi and Hawthorne (2019) analyzed afterslip following $M_w < 5$ Californian earthquakes. Wimpenny et al. (2017) and Alwahedi and Hawthorne (2019) compiled afterslip moment estimates for approximately 30 global earthquakes and showed that relative afterslip moment (M_{rel}), defined as:

$$M_{rel} = \frac{M_o^{aft}}{M_o}, \tag{1}$$

where M_o^{aft} is the afterslip moment and M_o is the coseismic moment, can vary by up to two orders of magnitude between different earthquakes. It is not clear what drives this, for example, can we explain why the similar-magnitude El Mayor Cucapah and Landers earthquakes generated M_{rel} of 74% and 2%, respectively (Fialko, 2004; Rollins et al., 2015)? A global synthesis of studies is needed to better establish the average and outlying behaviors of afterslip and provide observational constraints for physical models.

We compile 148 aseismic afterslip studies that follow 53 M_w6.0–9.1 earthquakes from 1979 to 2018. Using a refined subset of 88 better-constrained kinematic afterslip models (after 46 earthquakes), we investigate whether afterslip and coseismic moment scale in a discernible way. We explore whether observed variability in M_{rel} depends on characteristics of the mainshock (moment, rake, dip, depth, rupture aspect), measures of local deformation rate (fault slip rate, local strain rate, plate velocity), and data availability (the start date and duration of data collection). We discuss additional factors that are difficult to quantify and test statistically, including fault zone composition, earthquake history, and the influence of data availability and modeling methodology. We also investigate whether the occurrence of updip or downdip afterslip may be influenced by vertical rupture directivity, measured by a one-dimensional estimate. Determining what controls M_{rel} variation may offer new empirical constraints on afterslip, which could lead to improved predictive models of stress transfer for aftershock modeling and forecasting (Cattania et al., 2015; Mancini et al., 2020).

In Section 2, we outline the observations, kinematics, and a mechanical interpretation of afterslip and formulate hypotheses regarding the potential factors that M_{rel} might depend on, to later test. In Section 3, we explain our compilation and statistical methods and describe our database, which includes study and model metadata, and information on the mainshock and fault zone setting. This database is available online (doi:10.5281/zenodo.6414330). We present our analysis of the database in Section 4 and discuss our findings in Section 5.

2. Background

2.1. Observations and Mechanical Interpretation

The kinematics of afterslip can be well approximated by combining a constitutive framework for shear strength (τ) at an interface with elastic theory (Rubin, 2008). Rate and state dependent friction describes τ and the conditions under which materials strengthen or weaken with an imposed velocity step (Dieterich, 1979, 1987; Ruina, 1983). The Dieterich-Ruina formulation gives τ as:

$$\tau = \sigma\left(\mu_o + a\ln\frac{V}{V_o} + b\ln\frac{V_o\theta}{D_c}\right), \tag{2}$$

where σ is the effective normal stress and μ_o is the friction coefficient when slip velocity (V) equals the reference velocity (V_o). The direct effect term ($a\ln(V/V_o)$) describes an initial frictional strength increase and the evolution

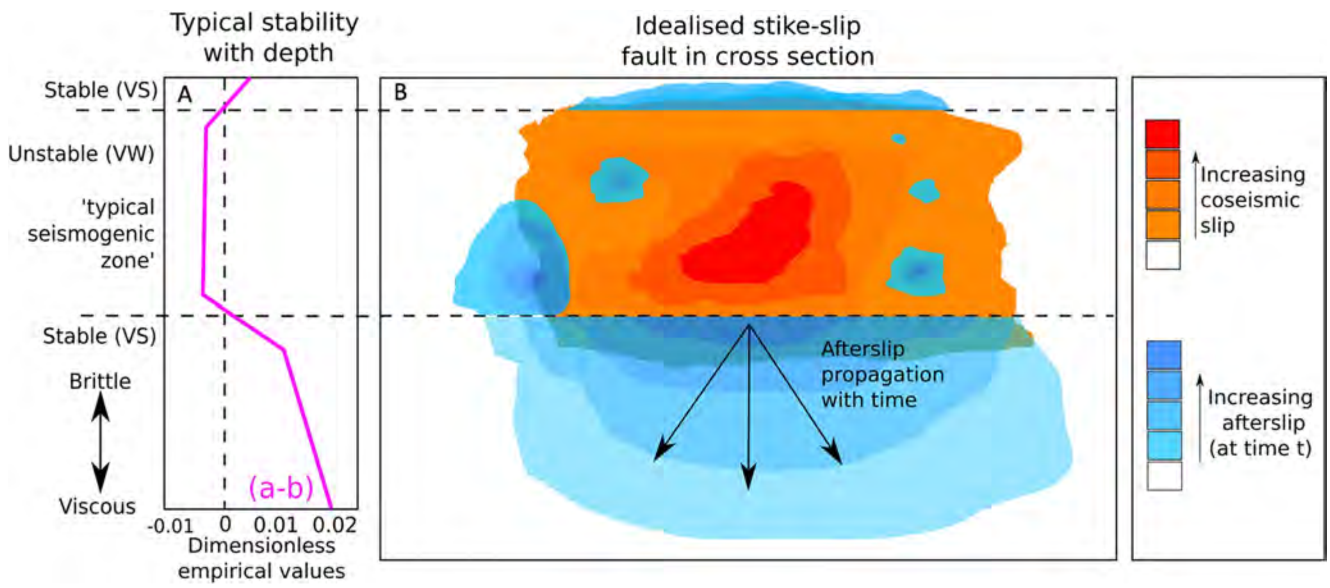


Figure 1. (a) A simplified frictional slip stability (a)–(b) profile with depth (modified from the study by Avouac [2015]; Bürgmann [2018]; Hillers et al. [2006]; Perfettini & Avouac [2007]), and (b), a schematic fault-parallel cross-section of idealized coseismic rupture and aseismic afterslip on a well-coupled strike slip fault. The approximate seismic-aseismic transition and brittle-viscous/ductile transitions are shown, but conditional stability is not shown for simplicity.

term ($b\ln(V_o\theta/D_c)$) describes a frictional strength reduction over slip distance and time, where D_c is the critical slip distance and θ is the state variable. a and b are empirical, dimensionless quantities weighing these terms. $a - b$ expresses the velocity dependence of a material under given environmental conditions, including stress, temperature, slip velocity, and effects of fluids (Blanpied et al., 1991; Marone, 1998). $a - b$ is depth-dependent (Blanpied et al., 1991), with distinct and strongly frictionally stable regions typically updip and downdip of the seismogenic zone, as shown in Figure 1a (e.g., Hillers et al., 2006; Imber et al., 2001). Rate and state dependent friction does not imply a microscale mechanism (Van den Ende et al., 2018), but one key interpretation of afterslip is a brittle creep (e.g., Marone et al., 1991; Perfettini & Avouac, 2004).

In brittle creep interpretations, afterslip occurs principally in frictionally stable fault regions, where $(a - b) > 0$ (Marone et al., 1991; Perfettini & Avouac, 2004). Here, seismic nucleation is prohibited and small increments of immediately-arrested brittle failure (Perfettini & Avouac, 2004) erode away at the stress concentrations left by an earthquake to produce aseismic, macroscale fault slip over time (Bürgmann, 2018; Harris, 1998). As the direct effect term dominates, afterslip can be approximated by a steady-state process (Marone et al., 1991; Scholz, 1998):

$$\tau = \sigma\left(\mu_o + (a-b)\ln\frac{V}{V_o}\right). \tag{3}$$

Figure 1b shows an idealized schematic of coseismic rupture and afterslip on a well coupled strike slip fault. Here, coseismic rupture is mostly confined to the well-defined frictionally unstable seismogenic zone but may propagate into adjacent stable regions through dynamic weakening (Noda & Lapusta, 2013; Shaw & Wesnousky, 2008). Afterslip then migrates away from the rupture edges within the frictionally stable regions (Bie et al., 2014; Peng & Zhao, 2009), but some occurs at traditionally seismogenic depths due to rheological heterogeneity or conditional stability. Afterslip has often been observed at traditionally seismogenic depths (e.g., Langbein et al., 2006; Reilinger et al., 2000; Riva et al., 2007), and thus is clearly not limited to distinct and strongly velocity-strengthening regions (Bürgmann et al., 2002; Helmstetter & Shaw, 2009). In the case of a poorly-coupled fault, such as the creeping section of the San Andreas (Bürgmann, 2018; Jolivet et al., 2015), isolated velocity-weakening rupture patches may exist within an overall more velocity strengthening fault. In another case, where fault material is only weakly velocity-weakening or conditionally stable, where $(a - b) \approx 0$ or <0, aseismic slip may occur if the slip velocities (Scholz, 1998) or the nucleation length scales (related to D_c; Boatwright & Cocco, 1996; Bürgmann, 2018; Rubin & Ampuero, 2005) required for seismic slip are not reached. In this case, the steady-state approximation does not hold, and triggered slow slip events may also be common (Bürgmann, 2018; Rolandone et al., 2018; Taira et al., 2014; Wallace et al., 2018).

2.2. Afterslip and the Mainshock and Fault Setting

The nature of a scaling relationship between the afterslip moment and the coseismic moment is not well constrained. Establishing what the expected or average afterslip moment for a given earthquake should be would allow for more informed investigations into behavioral variation. Moment (M_o) is given by the product of shear modulus (G), average slip $\left(\overline{D}\right)$ and slip area (A):

$$M_o = GA\overline{D}. \tag{4}$$

Coseismic ruptures are commonly assumed to scale self-similarly, whereby M_o, $\overline{D}$ and A grow in a consistent and scale-invariant way (e.g., Leonard, 2014; Wells & Coppersmith, 1994). Given this, and assuming that coseismic static stress change drives afterslip, we posit a monotonic relationship between afterslip and coseismic moments. Basic elastic theory predicts that the magnitude of stress change around a rupture and the area on the fault plane exposed to a given stress change increase with the coseismic moment (Segall, 2010). Therefore, assuming that shear modulus (G) remains approximately constant across seismogenic and afterslip zones, the average slip, area, and overall moment of afterslip should increase with the coseismic moment.

If M_{rel} is observed to vary, we can investigate factors that might drive this through testing the following hypotheses. If $a - b$ principally controls afterslip occurrence throughout the fault zone (e.g., Marone et al., 1991; Perfettini & Avouac, 2004) and is largely controlled by temperature and depth (e.g., Blanpied et al., 1991; Hillers et al., 2006; Imber et al., 2001), we hypothesize that low dip angle faults may permit more afterslip, by providing a greater area of unruptured fault in purely and conditionally frictionally stable regions. We also could expect a relationship with a rake, as this typically correlates with dip (Anderson, 1905). Again, assuming that $a - b$ is depth controlled, we hypothesize that shallower earthquakes may permit more downdip, and therefore overall, afterslip. However, this is complicated by the fact that updip afterslip can occur and that deep ruptures might be required to activate deeper frictionally stable regions in the first place. Finally, Hawthorne et al. (2016) alluded to a potential link between rupture elongation and M_{rel}, thus we investigate the influence of coseismic moment and the aspect ratio (length to downdip width) of coseismic rupture on M_{rel}. We, therefore, investigate relationships between M_{rel} and mainshock moment, fault dip, rake, depth, and rupture aspect.

Rheology may vary across different fault zones. We hypothesize that mature faults might promote more afterslip as they are suggested to contain higher proportions of velocity-strengthening materials like gouges and smoothed asperities (Choy & Kirby, 2004; Collettini et al., 2019; Ikari et al., 2011; Imber et al., 2008). Fluids or specific materials that might promote aseismic slip might also be present, such as the talc-bearing serpentinites in the creeping section of the San Andreas fault (Moore & Rymer, 2007) or well-connected phyllosilicate gouges (Niemeijer, 2018). We use measures of local deformation rate: fault slip rate (i.e., the long term rate at which a fault slips), local strain rate (i.e., how localized deformation is, a combination of fault zone width and slip rate), and plate velocity, as proxies for fault maturity and potential for abundant $(a - b) > 0$ material, as there is evidence that factors such as fault slip rate are linked to maturity (e.g., Goldberg et al., 2020; Manighetti et al., 2007). We, therefore, investigate relationships between M_{rel} and fault slip rate, local strain rate, and plate velocity.

Certain additional factors that may influence M_{rel} cannot be easily statistically tested. For example, the size and shape of different coseismic ruptures can vary at the same fault patch throughout multiple earthquake cycles (e.g., Bakun et al., 2005; Jiang & Lapusta, 2016; Shaw & Wesnousky, 2008), which may influence subsequent postseismic behaviors, as indicated in some earthquake cycle simulations (e.g., Barbot et al., 2012). This implies that any single observed earthquake and afterslip episode may not reflect the average behavior of events at that fault. Additionally, the variable presence and role of conditionally stable regions across different faults may also drive variations in M_{rel}. This includes whether these regions are locked or creep interseismically, whether they can rupture coseismically, or whether they can fail in either or both spontaneous or triggered slow slip events (e.g., Scholz, 1998; Noda & Lapusta, 2013; M. Wei et al., 2013; Bürgmann, 2018). Finally, the interseismic coupling may be linked with M_{rel} through factors such as fault maturity, rheology, fluid pressure, and structural heterogeneity (Chaussard et al., 2015; Harris, 2017; Kaneko et al., 2010). A lack of reliable interseismic coupling estimate

10.1029/2021JB023897

at many host faults makes this difficult to evaluate but is desirable for the future. These factors will be discussed alongside the implications of our results in Section 5.2.

2.3. Methods and Limitations of Observation and Modeling

Our understanding of afterslip derives principally from geodetic observations of its surface deformation (Bürgmann, 2018). The broad types of data used to analyze afterslip are ground-based surveys (e.g., creep- and strainmeters, etc.), GNSS (including GPS), InSAR, and satellite gravimetry. As these observation methods have absolute detection thresholds, we expect a bias in the literature toward readily detectable afterslip episodes. This could manifest as an apparent dependence of M_{rel} on mainshock magnitude, as low M_{rel} following large earthquakes may be detectable, whereas low M_{rel} following smaller earthquakes may not. Additionally, the deformation signals of different postseismic mechanisms may be overlain and concurrent (Barbot & Fialko, 2010), making it difficult to distinguish their individual contributions (e.g., Biggs et al., 2009; Ryder et al., 2007). Separating the contributions of afterslip and viscoelastic relaxation becomes particularly difficult above mainshock magnitudes M_w6.5-7, and the two processes can trade off strongly in models (e.g., Jacobs et al., 2002; Sun & Wang, 2015; Luo & Wang, 2021; M. Wang et al., 2021).

Afterslip moment estimates will be sensitive to the temporal window of observation. The steady-state approximation predicts afterslip velocity V at time t as:

$$V(t) = \frac{V_0}{1 + \frac{t}{C}}, \tag{5}$$

where V_0 is the initial velocity and C is a constant of decay. This approximation is well supported by observations, where the afterslip signal has a high onset amplitude (e.g., S. Wei et al., 2015; Tsang et al., 2019) and decays approximately with the inverse of time (e.g., Azúa et al., 2002; Ingleby & Wright, 2017; Marone, 1998; Perfettini & Avouac, 2004; Wennerberg & Sharp, 1997), and implies that the earliness and duration of study are crucial for capturing a representative afterslip signal.

Afterslip studies fall into three broad categories, each of which has different outputs and implications for this analysis. First, geodetic analyses are studies that typically fit decay equations to surface displacements (e.g., Savage & Svarc, 2009) or estimate the first-order spatial extent of afterslip (e.g., Ergintav et al., 2007), but do not produce a spatial distribution model of afterslip. Their conclusions regarding the spatial distribution of afterslip are generally qualitative and do not include a moment estimate. Kinematic slip modeling refers to studies that fit a spatial slip model to geodetic observations through dislocation theory (Okada, 1992; Segall, 2010). This may involve iterative forward modeling (e.g., Reilinger & Larsen, 1986) or explicit numerical inversion (e.g., L. Wang et al., 2009; Menke, 2018). Finally, dynamic slip modeling refers to studies that use a nonlinear inversion to constrain frictional parameters within frameworks such as the steady-state approximation. These can then produce a model of evolving afterslip from an initial postseismic stress field, which also satisfies geodetic observations (e.g., Johnson et al., 2009; Perfettini & Avouac, 2007). The inversion process is associated with considerable uncertainty arising from the validity of assumptions, inherent non-uniqueness, and regularization (Scales & Tenorio, 2001), discussed further in Section 5.3.

3. Data Compilation and Methods

3.1. Compilation From the Literature

We compile afterslip studies that follow M_w6.0 or greater earthquakes from 1979 onward, published until 2018 (inclusive). We omit earthquakes before this due to poor data quality, notably excluding: the 1959 Hebgen Lake (Nishimura & Thatcher, 2003), the 1978 Tabas E Golshan (Copley, 2014), and the 1940 Imperial Valley (Reilinger, 1984) earthquakes. The inclusion of a study in our compilation is irrespective of whether additional postseismic mechanisms are considered, but we note when viscoelastic relaxation and pore fluid effects are considered or modeled.

We systematically extract information about the afterslip model(s) from each study. We identify each study's preferred afterslip model, proposed by the authors as the best compromise of physical sense and data fit and record the proposed moment, any bounds on this (from error analysis or viable alternative models), and the depth

10.1029/2021JB023897

extent of 'most' afterslip. The latter is approximate and often derived from qualitative discussions or inferred from figures, as digitized afterslip models are scarcely provided. We omit one-moment estimate from a study by Paul et al. (2007) because it considers only a proportion of the total spatial extent of afterslip and would not be comparable. If multiple viable models are proposed without a strong preference, we average the proposed moments. For magnitude-moment conversions we use Hanks and Kanamori (1979), where M_o is in N m (Nm):

$$M_w = \left(\log_{10} M_o - 9.05\right)/1.5. \tag{6}$$

We assume that a significant deformation signal related to aftershocks has been removed (e.g., Hoffmann et al., 2018; Howell et al., 2017) or is negligible (e.g., Barnhart et al., 2016; Béjar-Pizarro et al., 2010). However, seismic afterslip and aftershocks are not treated as separate mechanisms in some studies. To consistently consider only aseismic afterslip, we reduce the moment estimates of Donnellan et al. (2002), Gahalaut et al. (2008), and Shrivastava et al. (2016) by 13%, 47%, and 10%, respectively, which they explicitly gave as the seismic proportions of their afterslip moment estimates.

We record data and modeling information for each study. This includes the data type(s) used, the start and end time of observation (converted to an approximate number of days since mainshock), the broad modeling type, and many individual modeling choices, where possible (see supplementary materials or database for detail). We assume a start time of 1 day when one is not explicitly given, as these are generally continuous GPS studies, and/or we assume that longer delays between the parent earthquake and data collection would be mentioned explicitly. Some studies also account for early missing afterslip by extrapolation (e.g., D'Agostino et al., 2012; Perfettini et al., 2010) or by estimating how much afterslip is contained within the coseismic model (e.g., Hutton et al., 2001), and we use these estimates.

3.2. Compilation of Mainshock Data

We compile mainshock information from global earthquake catalogs. For each earthquake, we record the moment, magnitude, longitude, latitude, depth, dip, and rake from the preferred W-phase moment tensor (M_{ww}) solution of the USGS ComCat database (U.S. Geological Survey, 2017) and from the Global Centroid Moment Tensor (GCMT) catalog (Dziewonski et al., 1981; Ekström et al., 2012). In this study, we do not need to distinguish between left and right-lateral strike slip, thus we convert the circular rake values to semi-circular values, with normal and thrust faulting as endmembers and strike slip in between. To deduce the correct fault plane from the two nodal planes of each focal mechanism, we use figures and dip and strike values given in the compiled literature. We obtain a hypocentral depth and an approximate coseismic slip depth extent, bounded by at least 1 cm of slip, from coseismic slip models in the Earthquake Source Model database: SRCMOD (Mai & Thingbaijam, 2014). We use slip models by Hayes (2017) where possible, but otherwise choose a simple, preferably single fault plane model to be as systematic as possible.

In most cases, we use the USGS ComCat preferred solution's seismic moment as the 'driving' moment of afterslip. However, for the following cases where the mainshock is ambiguous to define (i.e., mainshock sequences), we use a summed driving moment: (a) the six M_w5.2–6.3 1994 Sefidabeh earthquakes, (b) the two M_w6.5 2000 South Iceland earthquakes, (c) the M_w7.1 2005 Miyagi mainshock and its M_w6.6 aftershock, after Miura et al. (2006), (d) the entire 2009 Karonga swarm, as given by Hamiel et al. (2012), (e) the M_w8.1 and M_w8.3 2006/7 Kuril islands earthquakes, and (f) the M_w5.7 and M_w6.0 1997 Umbria-Marche earthquakes. We divide each afterslip moment estimate by the driving moment to obtain M_{rel}.

3.3. Compilation of Tectonic Data

We obtain tectonic and fault setting information for each earthquake from external, global data sets. We identify the major fault closest to each mainshock hypocenter in the Global Earthquake Model Foundation (GEM) global active faults database (Styron & Pagani, 2020) and extract the net fault slip rate (i.e., long term average value in the direction of maximum displacement) for each earthquake from the GEM data-base. We calculate the second invariant of the strain rate tensor closest to the mainshock hypocenter from Kreemer et al. (2014) for continental events. For subduction events, the projection of the hypocenter to the surface is generally far from the fault trace which caused issues in selecting a representative strain rate value systematically. Instead, we obtain a value for plate velocity from the GEM Strain Rate Model: GSRM 2.1 (Kreemer et al., 2014; UNAVCO, 2021) at the hypocentral location of each earthquake, as this is more meaningful.

3.4. Statistical Tests

We investigate variations in absolute and relative afterslip moment and test for correlations between relative afterslip moment and various factors. We compile 95 moment estimates from individual studies, but formally analyze a slightly reduced data set of 88 well-constrained kinematic slip model estimates that follow 46 earthquakes. This small reduction ensures standardization and comparability between the models we analyze.

As there are multiple moment estimates for some events, we bootstrap to fairly sample data and robustly test correlations. For each test, we create 2000 subsets of data, each with one randomly sampled estimate for every earthquake ($n = 46$, the number of earthquakes and data points of each subset). We calculate Spearman's rank correlation coefficient between each subset and the characteristic we are testing and present the median value and 95% range of the distribution. We use Spearman's rank to test for monotonic relationships (Dodge, 2008), as testing specifically for linearity (i.e., Pearson's) may miss complex, nonlinear relationships and could be disproportionately affected by outliers in our data. As the bootstrapped distributions are not necessarily Gaussian, we use the median and 95% range rather than the mean and standard deviation, which could be less representative and more sensitive to outlying values. As our correlation coefficients are based on rank rather than absolute value, we cannot provide a data-fit or measure of an error on individual coefficients, thus our statistical measures do less well at reflecting the additional uncertainty in individual moment estimates, but we discuss these uncertainties further in Section 4.3. We interpret a result as statistically interesting if the entire 95% range does not cross the zero coefficient line.

We use reduced data sets with specific criteria to further probe the relationships between M_o^{aft} and M_o, and M_{rel} and M_o. The following reduced data sets contain one estimate per earthquake and do not need bootstrapping: (a) the model with the longest duration for each earthquake ($n = 45$), (b) the largest afterslip moment estimate for each earthquake ($n = 45$), and (c) the longest duration model that also starts within 1 day of the earthquake ($n = 32$). Data set 3 is further refined as (e) removing the two outlying M_{rel} endmembers ($n = 30$), (f) including only subduction events ($n = 17$), and (g) including only earthquakes M_w7.0 or greater. The Sefidabeh study by Copley (2014) is an extreme outlier in terms of the start time (more than 2 years), that we omit in all of these reduced data sets.

4. Results

4.1. The Database

The database contains 148 studies of afterslip following 53 mainshocks (doi:10.5281/zenodo.6414330). The earthquakes span M_w 6.0–9.1 and comprise 32 thrust, 14 strike-slip, and 7 normal mechanisms (Figure 2a). Analysis of the GCMT catalog indicates that the database contains 100% of the M_w9 earthquakes that occurred during the study period, 32% of M_w8, 4% of M_w7, and less than 1% of M_w6 earthquakes. Smaller earthquakes are underrepresented in our compilation and those included may have a bias toward higher M_{rel} due to more readily detectable afterslip.

Studies vary in data practices and modeling methodologies. Overall, we categorize 18 geodetic analyses, 117 kinematic slip models, and 13 dynamic slip models. Approximately 41% of all studies considered only afterslip as a viable postseismic mechanism, 32% considered afterslip and viscoelastic relaxation, 3% considered afterslip and pore fluid factors, and 24% considered all three mechanisms. Figure 2b shows InSAR emerging and GNSS becoming dominant in the 1990s, with gravity-based methods emerging more recently and ground-based surveys scarcely used this century.

The database contains multiple afterslip studies for some earthquakes, although not every study proposes a moment estimate. There are multiple studies for 32 mainshocks, as shown in Figure 2c, and six particularly well-studied examples: M_w7.1 1999 Hector Mine (6 studies), M_w6.0 2004 Parkfield (7), M_w7.6 1999 Izmit (9), M_w9.1 2011 Tohoku (9), M_w9.1 2004 Sumatra (10) and M_w7.3 1992 Landers (11). Overall, 95 studies provide a meaningful afterslip moment estimate as geodetic analyses generally cannot estimate moment and many

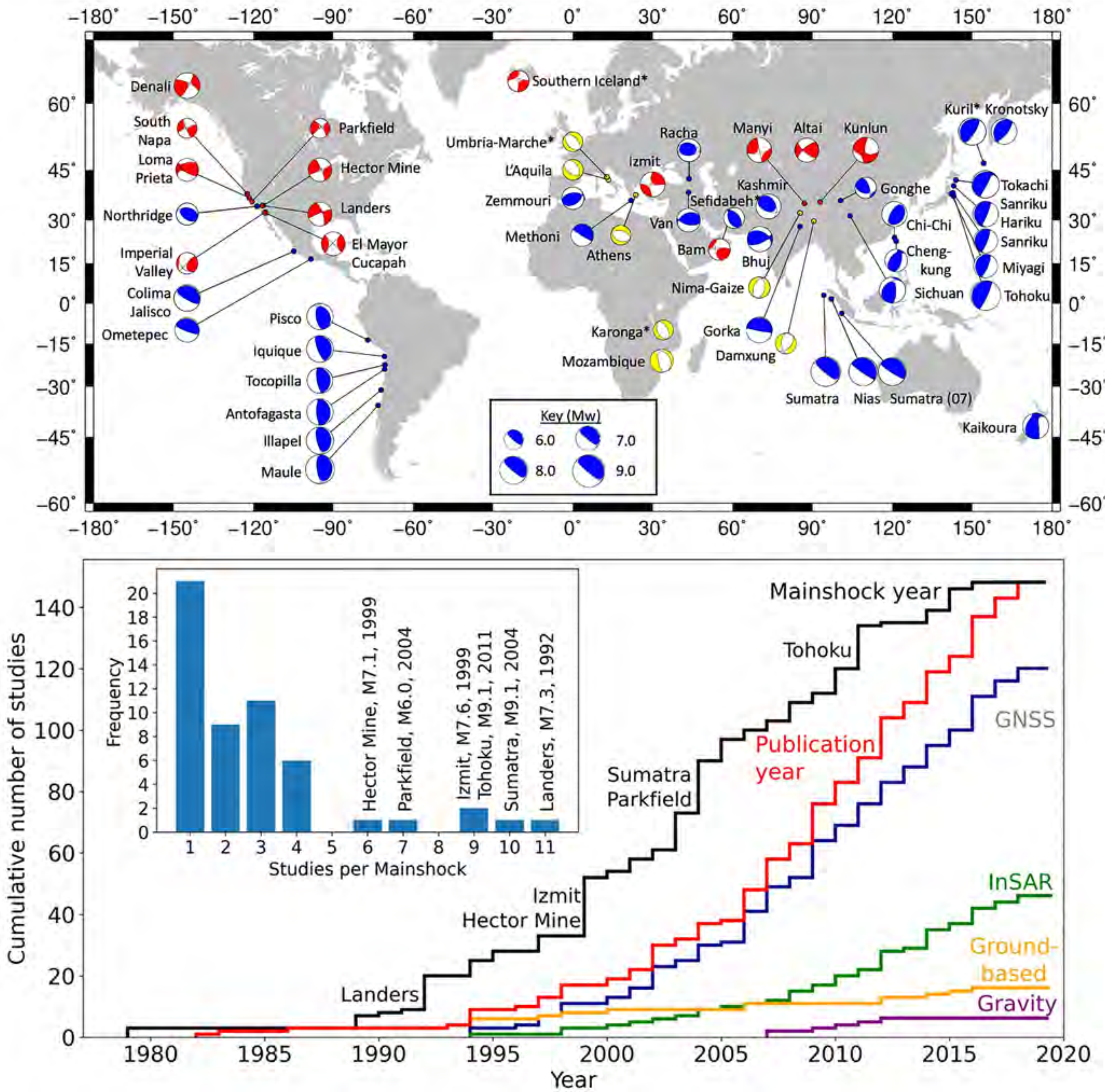


Figure 2. (a) The Global Centroid Moment Tensor focal mechanism solutions of the earthquakes in our database (red: strike slip, blue: thrust, yellow: normal, * indicates a mainshock sequence with the largest event shown). (b) The cumulative number of compiled studies is shown by year of mainshock and year of publication, and the cumulative use of data types by year of publication. (c) The frequency of studies per mainshock, with the most-represented mainshocks, annotated, corresponding to large steps in panel (b).

kinematic and dynamic slip models do not explicitly calculate or give one. Eighty eight moment estimates come from kinematic slip models, whose methodologies are better constrained and thus more comparable.

The start times and durations of all studies are summarized in Figure 3, ordered by the mainshock. If afterslip velocities decay according to Equation 5 and this is linearly proportional to moment release rate, the cumulative moment release should be proportional to the logarithm of time, thus we present logarithmic time on the *x*-axis. Most studies start within a few days of the mainshock, with approximately 1 day being the soonest and 2 years being the latest, and typically last for several months to around 2 years, with approximately 1 day being the

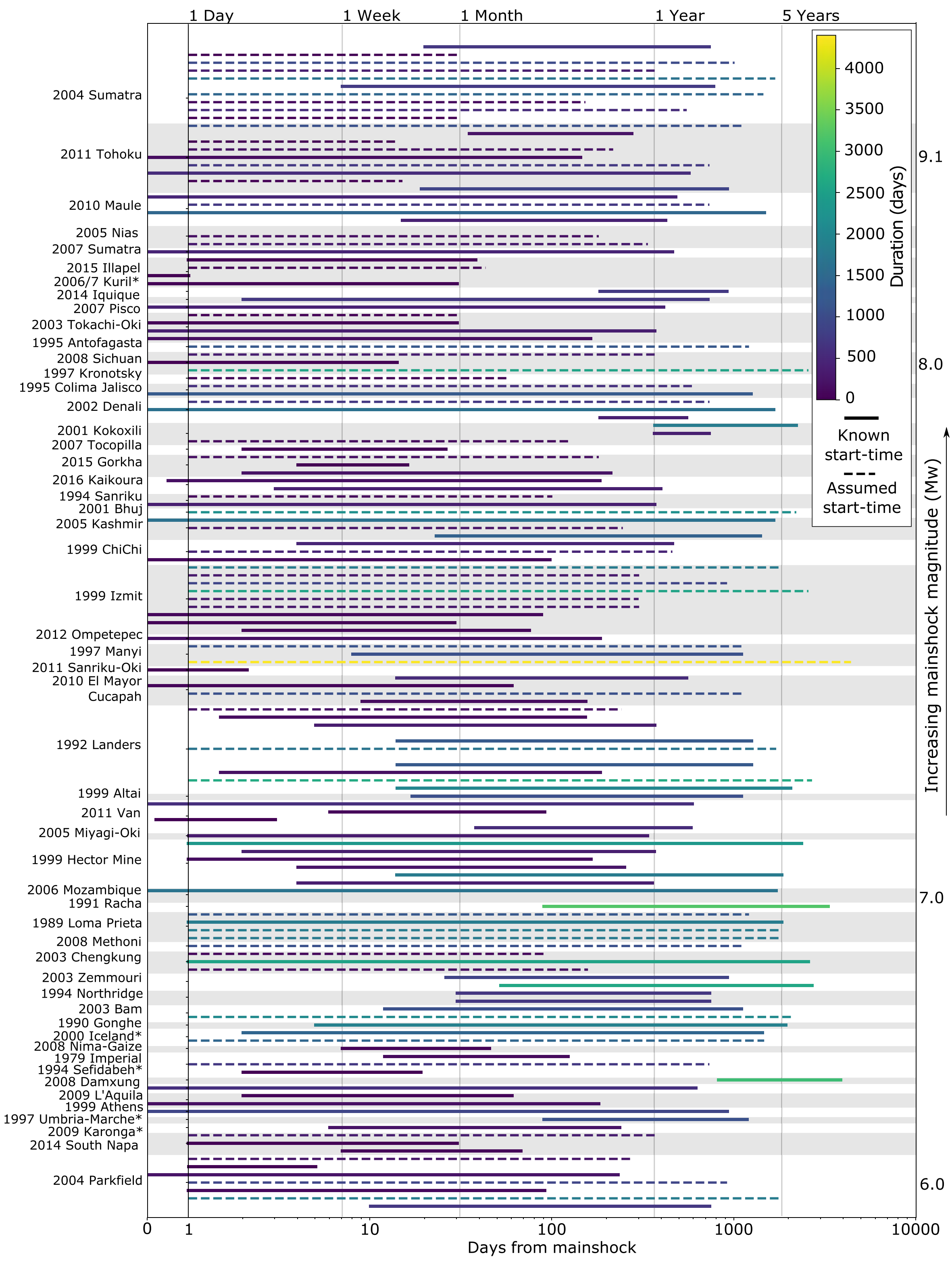


Figure 3. Temporal observation windows for all compiled studies, where available. The line length indicates the base-10 logarithmic duration and the color gives linear duration. Dashed lines indicate studies without an explicitly provided start time, which we assume is 1-day as most are continuous GPS.

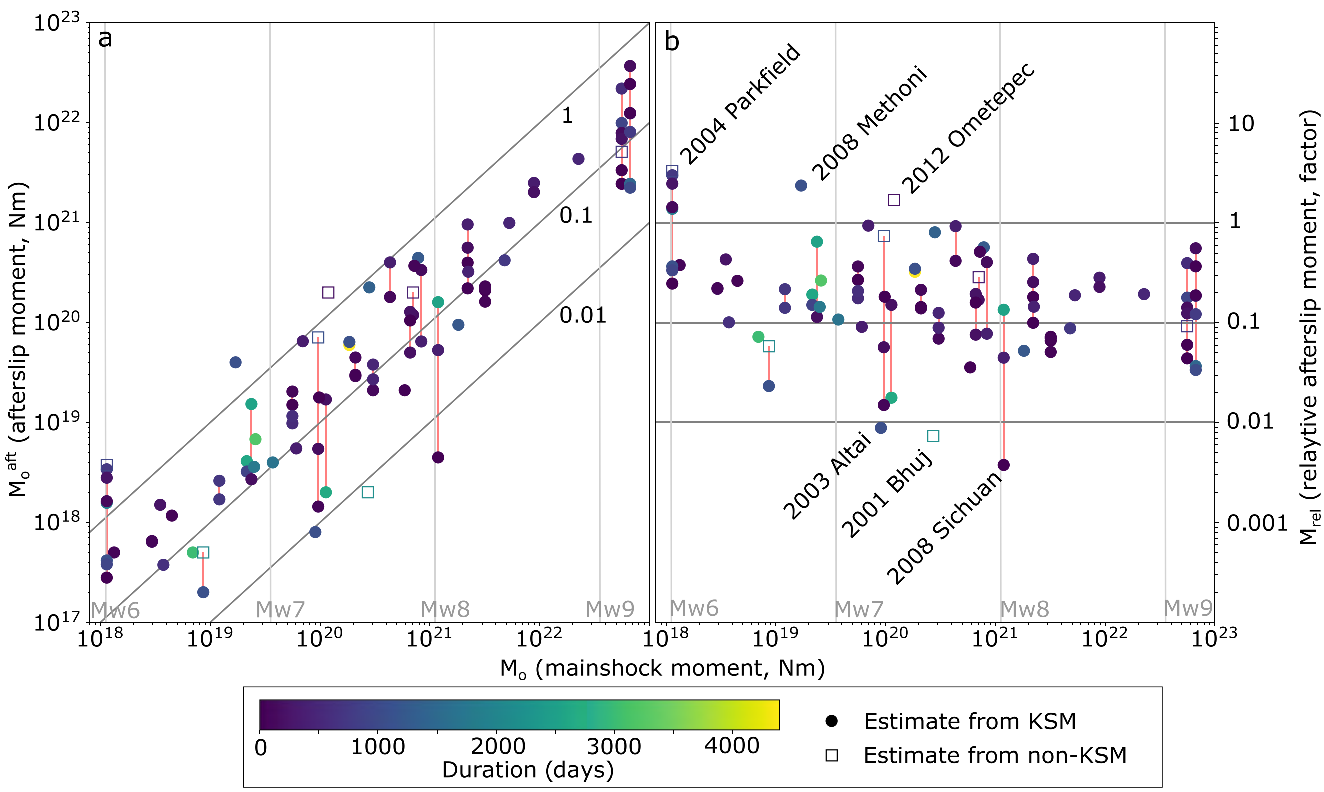


Figure 4. (a) Afterslip moment estimates against corresponding coseismic moments (M_o^{aft} vs. M_o) and (b) relative afterslip moment estimates against coseismic moment (M_{rel} vs. M_o). The color scale shows the linear temporal duration of each model. Red bars link estimates for the same earthquake from different studies. The 88 circles denote the kinematic slip model estimates (KSMs) that are analyzed further. Relative afterslip moment estimates <1% and >100% are labeled.

shortest and 12 years being the longest. We explore the relationship between M_{rel} estimates and the start time and duration of observation in Sections 4.3 and 4.4 and discuss our findings in 5.3.

4.2. Afterslip Moment Scaling and Variation

Figure 4a gives afterslip moment estimates against the corresponding coseismic moment. For the 88 kinematic slip model estimates, the median Spearman's rank correlation coefficient between M_o^{aft} and M_o is 0.91 after bootstrapping, with the 95% range between 0.89 and 0.93 (Figure 4a). This supports the hypothesis that aseismic afterslip moment scales with coseismic moment. We also note that the median Pearson's correlation coefficient between $\log\left(M_o^{aft}\right)$ and $\log(M_o)$ is 0.92 after bootstrapping, with a gradient close to one. We infer near-linear scaling of the afterslip moment with the coseismic moment for our mainshocks, which we discuss further in Section 5.1.

The 95% range of Spearman's rank correlation coefficients better reflects variation due to bootstrapping than the variations in individual afterslip moment estimates. We analyze the uncertainty in some individual estimates in Section 4.3, but further test the robustness of the M_o^{aft}/M_o correlation by examining the reduced data sets defined in Section 3.4. These correlation coefficients range from 0.85 to 0.93, shown in Figure 5a, which is close to that obtained by bootstrapping over the entire data set, further supporting a robust and strong relationship.

Relative afterslip moment (M_{rel}) varies over three orders of magnitude from <1% to >300% (Figure 4b). The median value for the 88 kinematic slip model estimates is 18% with an interquartile range of 9%–32%. Endmembers include two estimates below 1%: the M_w7.2 2003 Altai earthquake (Barbot et al., 2008) and the M_w8.0 2008 Sichuan earthquake (Shao et al., 2011), and five greater than 100%: the M_w6.0 2004 Parkfield earthquake (Bruhat et al., 2011; Freed, 2007; Johanson et al., 2006; Langbein et al., 2006) and the M_w6.8 2008 Methoni earthquake (Howell et al., 2017).

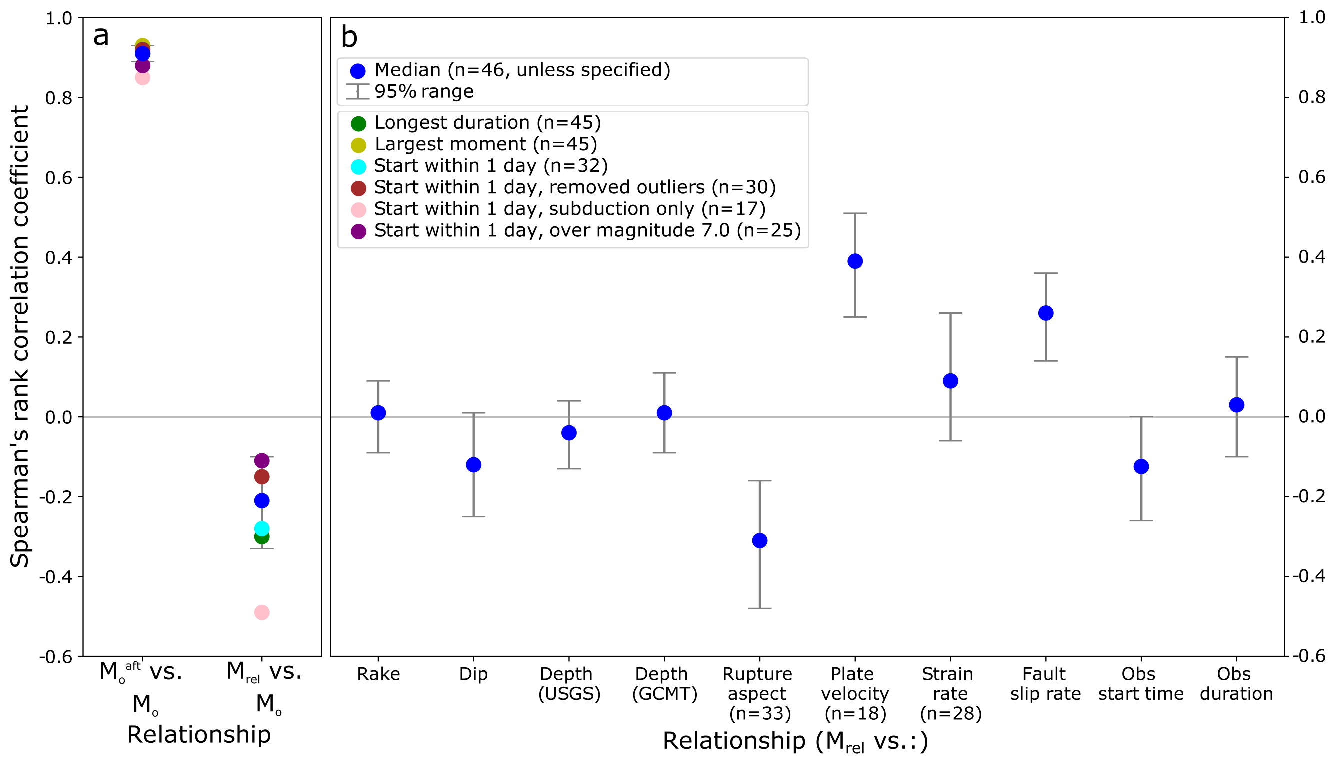


Figure 5. Median Spearman's rank correlation coefficients and 95% ranges for relationships tested. n = 46, with 88 data points bootstrapped over throughout, unless specified. (a) Absolute and relative afterslip moment against the coseismic moment. The correlation coefficients for reduced data sets are also shown, which do not require bootstrapping as there is only one data point per earthquake (n is given individually) and (b) relative afterslip moment against our tested metrics. The rake value is calculated slightly differently and is explained in Section 4.4.

M_{rel} weakly and negatively correlates with the mainshock moment. The median Spearman's rank correlation coefficient for M_{rel} and M_o is −0.21 after bootstrapping, with the 95% range from −0.32 to −0.09. This could suggest that larger earthquakes are prone to less M_{rel}, but this may be due to the publication bias, and is discussed further in Section 5.2. Correlation coefficients from the reduced data sets vary from −0.11 to −0.49 (see Figure 5a), likely because the data sets are smaller and thus less stable. The most outlying coefficient (−0.49) is from the smallest data set (n = 17), and removing a single outlying data point (for the M_w6.8 2008 Methoni earthquake) highlights this instability as the correlation coefficient falls from −0.49 to −0.38. As the overall correlation between M_{rel} and M_o is much weaker than between M_o^{aft} and M_o, this motivates the investigation of other factors to account for variability in M_{rel}.

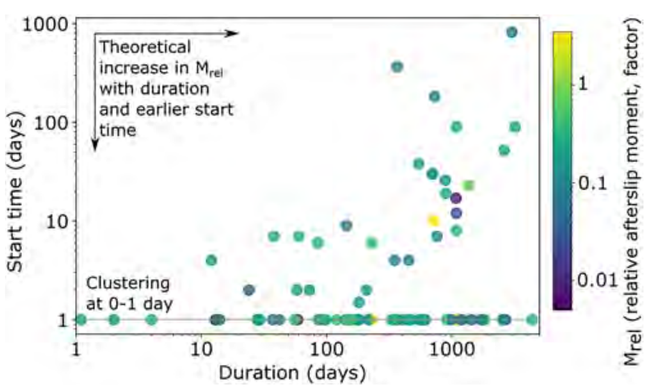


Figure 6. Observation start times and durations for all 88 kinematic afterslip models with relative afterslip moment estimates (M_{rel}), are shown in color. Models with start times given before or on day one are shown at 1-day, models without an explicitly provided start time begin on day 1.

4.3. Temporal Dependence and Uncertainty of Individual M_{rel} Estimates

Figure 6 shows the relationship between estimates of relative afterslip moment and the start time and duration of observation. The median Spearman's rank correlation coefficients after bootstrapping are −0.13 and 0.03, respectively, with 95% ranges of −0.24 to 0.00 and −0.09 to 0.16, respectively (Figure 5), indicating that across the data, there is no strong relationship between M_{rel} and observation start time and duration. This is surprising, given the theoretical temporal decay of afterslip: early and longer observation windows should result in greater afterslip moment estimates. We discuss the implications of this in Section 5.3.

Figure 7 shows that different afterslip moment estimates following the same earthquake can vary considerably. If differences in observation start time and duration cannot explain these differences, this would imply significant

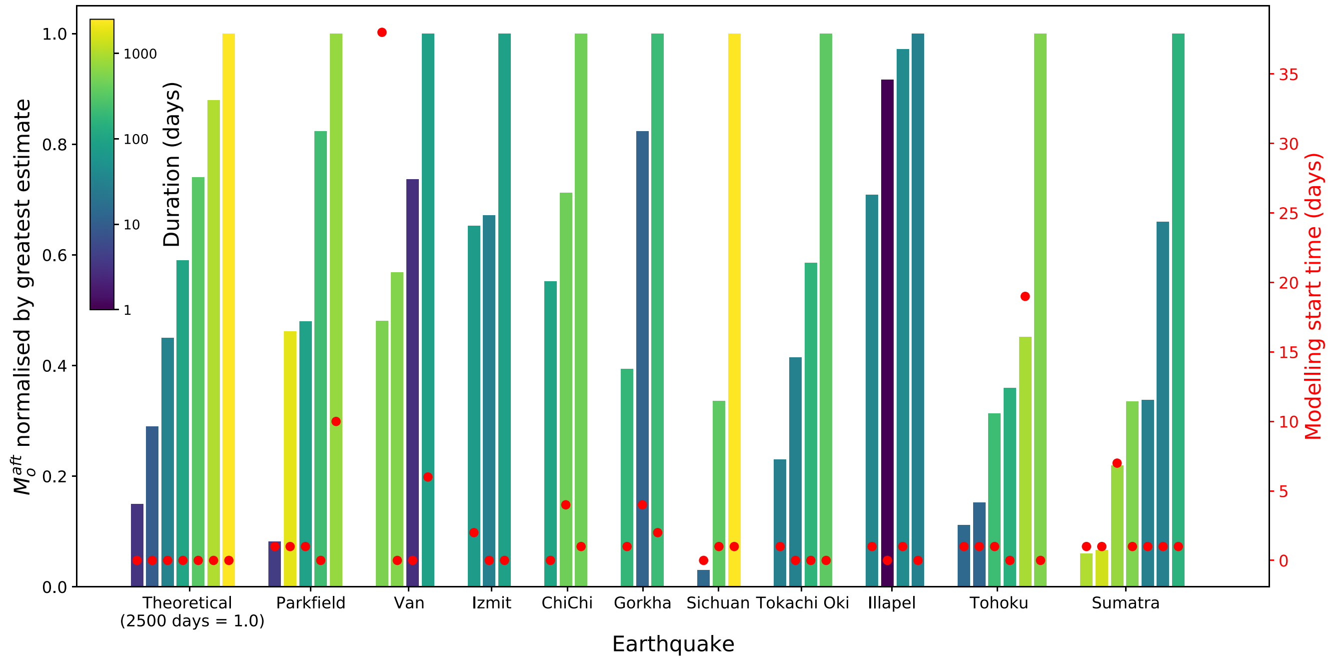


Figure 7. Afterslip moment estimates of the 10 mainshocks which have three or more kinematic slip model estimates. For each earthquake (shown across the *X*-axis), different afterslip moment estimates are shown as bars, normalised to the largest and arranged from smallest to largest (scale is given on the left *Y*-axis). The start time of the data used for each estimate is given by red circles (right *Y*-axis) and the duration is given by color. A theoretical case of how the afterslip moment should grow with time (based on an assumed steady-state, velocity-strengthening decay behavior) is also shown for comparison. Here, estimates from different durations over which data were analyzed are shown in ascending order, reaching 1.0 at 2,500 days. Moment estimates of the Sichuan earthquake, for example, appear to follow the expected trend if afterslip estimates were solely determined by duration and onset of the analyzed dataset, whereas Parkfield estimates do not follow the expected trend, suggesting other modeling sources of uncertainty.

modeling uncertainty. We analyze 10 earthquakes that have at least three afterslip moment estimates from different kinematic slip model studies. Estimates are normalized to the largest value for each earthquake to highlight the relative spread. We also present an expected, theoretical case in which afterslip is fully captured following an idealized earthquake by studies of different durations. To calculate this, we assume that afterslip velocity (Equation 5) linearly relates to the afterslip moment release rate, thus the integral with respect to time gives moment release. We assume the initial rate and constant c both equal one for simplicity and normalize to one unit of afterslip at 2,500 days to attain this idealized case.

For four earthquakes, the relative spread in individual afterslip moment estimates can be explained by differences in the temporal observation window. Afterslip moment estimates for the M_w7.6 1999 Izmit and the M_w7.6 1999 Chi-Chi earthquakes are relatively well-constrained within a factor of 2 (Figure 7), with larger estimates corresponding to increased observation duration or decreased observation start time. More varied afterslip moment estimates follow the M_w8.2 2003 Tokachi Oki earthquake (varying by up to a factor of approximately 5) but correspond with the duration of observation. Similarly, following the M_w8.0 2008 Sichuan earthquake, the smallest afterslip moment estimate is only 3% of the largest but corresponds to an observation duration of approximately two weeks compared to seven years. Both the theoretical and Sichuan case are normalized to one unit of afterslip at approximately 2,500 days, thus differences could be interpreted as (a) afterslip decaying faster after the Sichuan earthquake than the theoretical case, (b) the largest Sichuan estimate is erroneously high, (c) the two smaller Sichuan estimates are erroneously low, or (d) a combination.

For six earthquakes, the relative spread in individual afterslip moment estimates cannot be easily explained by differences in the temporal observation window. In two cases: the M_w8.0 2011 Van and the M_w8.3 2015 Illapel earthquakes, afterslip moment estimates are relatively well-constrained by a factor of approximately two, but there are four cases where the relative spread is considerably greater: the M_w6.0 2004 Parkfield, the M_w7.8 2015 Gorkha, the M_w9.1 2011 Tohoku and M_w9.1 2004 Sumatra earthquakes. For example, following the Sumatra earthquake, the two longest duration studies produced afterslip moment estimates of approximately only 6% of the largest and 10% of the second largest (both of which happened to be among the shortest duration studies). This indicates that an individual afterslip moment estimate may be more than an order of magnitude too small or too large (i.e., <10 to >1000%). The extreme variation following the Tohoku and Sumatra earthquakes is surprising as these are among the best-studied earthquakes and postseismic periods, and also suggests that uncertainty does not decrease with the coseismic moment. As six out of the 10 examples analyzed show spread in afterslip moment estimates which cannot be easily attributed to differences in observation start time or duration, we conclude that there is significant uncertainty associated with the modeling process.

This analysis indicates that the relative uncertainty in afterslip moment estimates can obscure the dependence we expect to see from either the observation start time or duration. For this reason, we do not attempt to normalize afterslip moment estimates for observational time window and instead consider individual afterslip moment estimates as given, but recognize potential for substantial uncertainty. Using the 10 analyzed earthquakes, we can assess the uncertainty of a typical afterslip moment estimate. The average mean and average variance of these 10 groups of estimates (each relative to the largest) is 0.62 and 0.1, respectively. Assuming, therefore, that a given afterslip moment estimate is 0.62 ± 0.1 of a full population of estimates, and that the best estimate solution lies somewhere in that population, the given estimate is likely within a factor of ∼two or three of the best estimate solution. However, in the most extreme case (as illustrated by the M_w9.1 2004 Sumatra earthquake) estimates could be out by an order of magnitude. The sources and implications of this uncertainty are discussed in Section 5.3.

4.4. Factors Contributing to M_{rel} Variation

We investigate potential controls on relative afterslip moment by testing the hypotheses formed in Sections 2.2 and 2.3. Figure 5b summarizes the median Spearman's correlation coefficients between M_{rel} and our testable metrics after bootstrapping. These coefficients range from near zero to |0.39|, a weak to moderate correlation. The 95% ranges vary in width and reflect the full distribution of correlation coefficients from bootstrapping to indicate a sense of the robustness of the relationship.

Figure 8 shows M_{rel} against mainshock rake, fault dip, depth, and rupture aspect ratio. The correlation coefficients between M_{rel} and the vertical component of rake and dip are 0.01 and −0.12 (Figures 8a and 8b), respectively, with both 95% ranges crossing the zero coefficient baseline, indicating no obvious control on afterslip (GCMT and USGS rake and dip values were very similar). Whilst we show the actual rake value in Figure 8a, we test adjusted (semi-circular not circular) values whereby thrust and normal mechanisms are endmembers and strike slip sits in between (i.e., right and left lateral slip are treated the same in the context of our hypothesis).

M_{rel} correlates with rupture aspect ratio but not with mainshock depth (Figures 8c and 8d). The median Spearman's rank correlation coefficients are −0.04 and 0.01 for the USGS and GCMT depths, respectively, indicating no obvious control on M_{rel}. Figure 8d shows the approximate length-to-width rupture aspect ratio against M_{rel} for the 33 earthquakes for which a coseismic slip model was available. The associated median bootstrapped Spearman's rank correlation coefficient is a moderate −0.31 and has a 95% range entirely negative. As continental and subduction earthquake populations might behave differently in terms of aspect ratio (e.g., Ampuero & Mao, 2017), we also calculate the correlation coefficients for continental (−0.34) and subduction (−0.24) populations individually, but these are quite similar to one another and the overall average.

Figure 9 shows M_{rel} against the local strain rate, plate velocity, and fault slip rate. These have correlation coefficients of 0.09, 0.39, and 0.26, respectively. The 95% ranges for the more strongly correlated plate velocity and fault slip rate relationships are also entirely above zero. The moderate relationship with plate velocity is for only 18 events on subduction interfaces, thus having less scope for interpretation as the fewer data points mean a less robust coefficient. However, the moderate relationship with fault slip rate is over the entire kinematic slip model data set of 46 earthquakes and 88 estimates, implying some robustness.

4.5. Afterslip Depth Analysis

We investigate whether the occurrence of up- or downdip afterslip may be influenced by vertical rupture directivity, using the simple, one-dimensional proxy of whether an earthquake's centroid is above or below the hypocenter. We conduct this analysis for 31 earthquakes for which we have: approximate afterslip and coseismic depth extents, hypocenter depths, and centroid depths. Figure 10 shows that in at least one study for each earthquake,

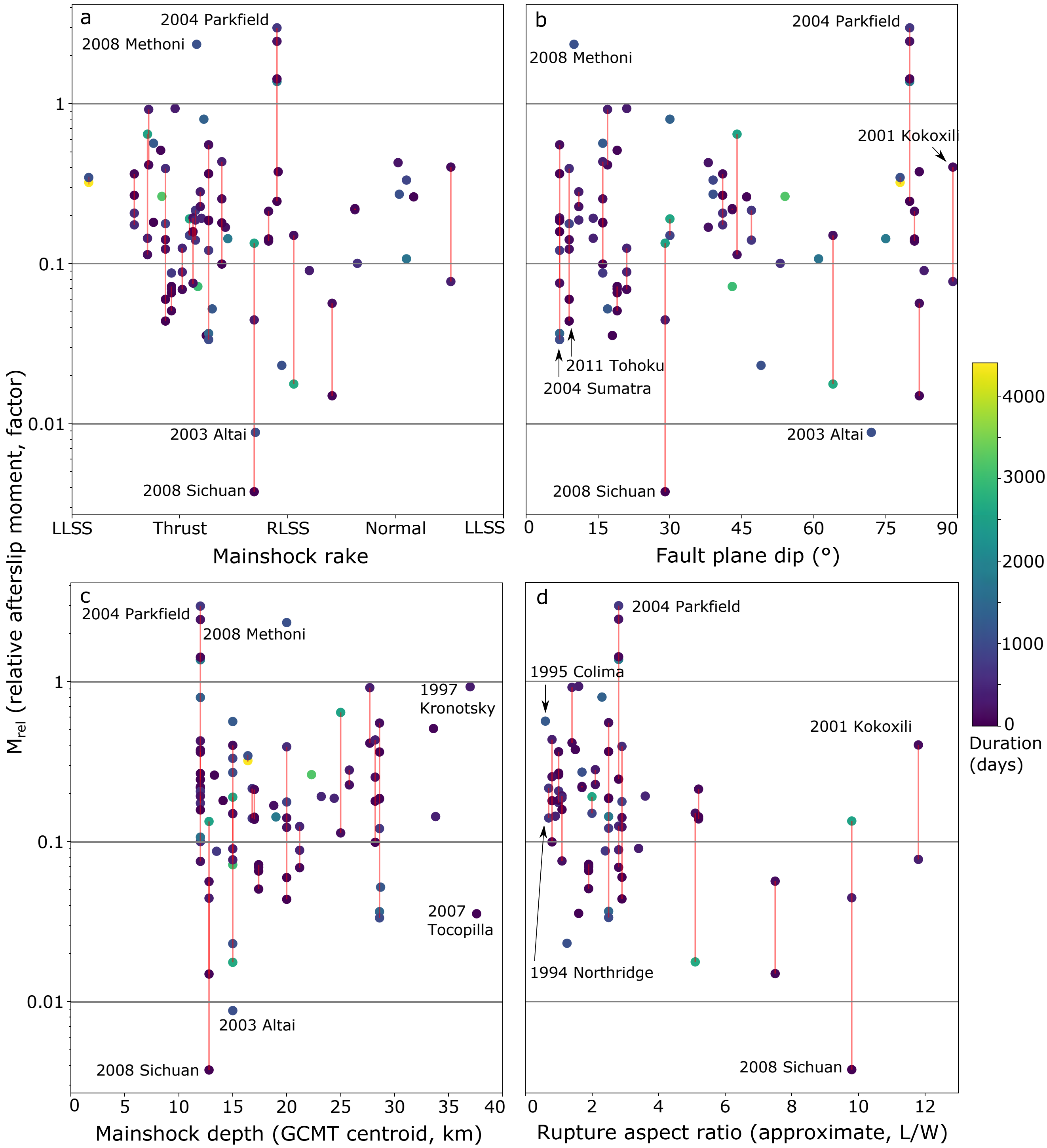


Figure 8. Relative afterslip moment (M_{rel}) against (a) mainshock fault plane rake (we test the vertical component of rake, median Spearman's rank correlation coefficient: 0.01), (b) mainshock fault dip (−0.12), (c) mainshock centroid depth (0.04), and (d) approximate rupture aspect ratio (−0.31). a and b show USGS preferred solution moment tensor values, c shows Global Centroid Moment Tensor centroid depth values, and d uses models from the Earthquake Source Model Database (SRCMOD, for the 33 available events only). Red lines connect different estimates from the same earthquake and color indicates the temporal duration of each study.

afterslip and coseismic slip depths overlap by at least one km. We cannot comment on whether specific slip patches overlap, as this could be due to afterslip and coseismic slip distributions varying laterally. However, this at least indicates that rheological heterogeneity (i.e., deviations from simple, one-dimensional slip stability

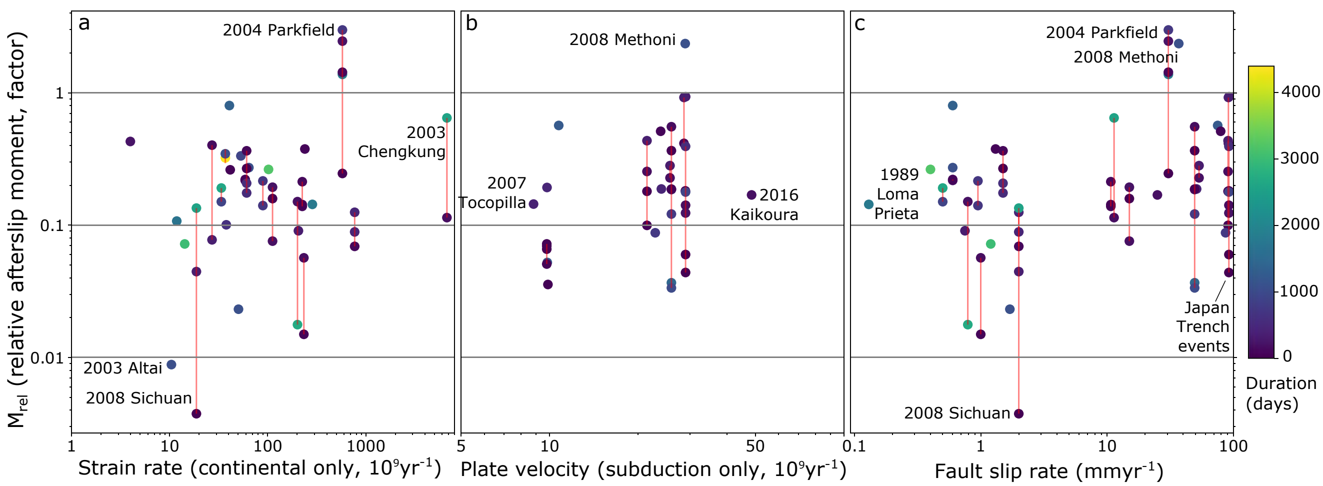


Figure 9. Relative afterslip moment (M_{rel}) against (a) local strain rate for the 28 continental-setting events (median Spearman's rank correlation coefficient: 0.09), (b) plate velocity for the 18 events broadly on a subduction interface (0.39) and (c) local fault slip rate for all 46 events (0.26). Red lines connect different estimates from the same earthquake.

models with constant depths, Figure 1) may be quite common in fault zones, especially as there is evidence of afterslip occurring throughout the entire coseismic slip depth range for approximately a third of the earthquakes in this analysis.

The relative depths of the centroid and hypocenter do not appear to influence the depth extent of afterslip. Twelve earthquakes have a hypocenter at least one km deeper than centroid, which we describe as net-updip propagating. Six of these earthquakes show some evidence of afterslip significantly above coseismic rupture depths, whilst six do not. Additionally, five earthquakes that cannot be described as net-updip propagating also show evidence of significant updip afterslip. At a threshold of five km, only five earthquakes qualify as net-updip propagating and only two of these show evidence of significant afterslip updip of coseismic rupture.

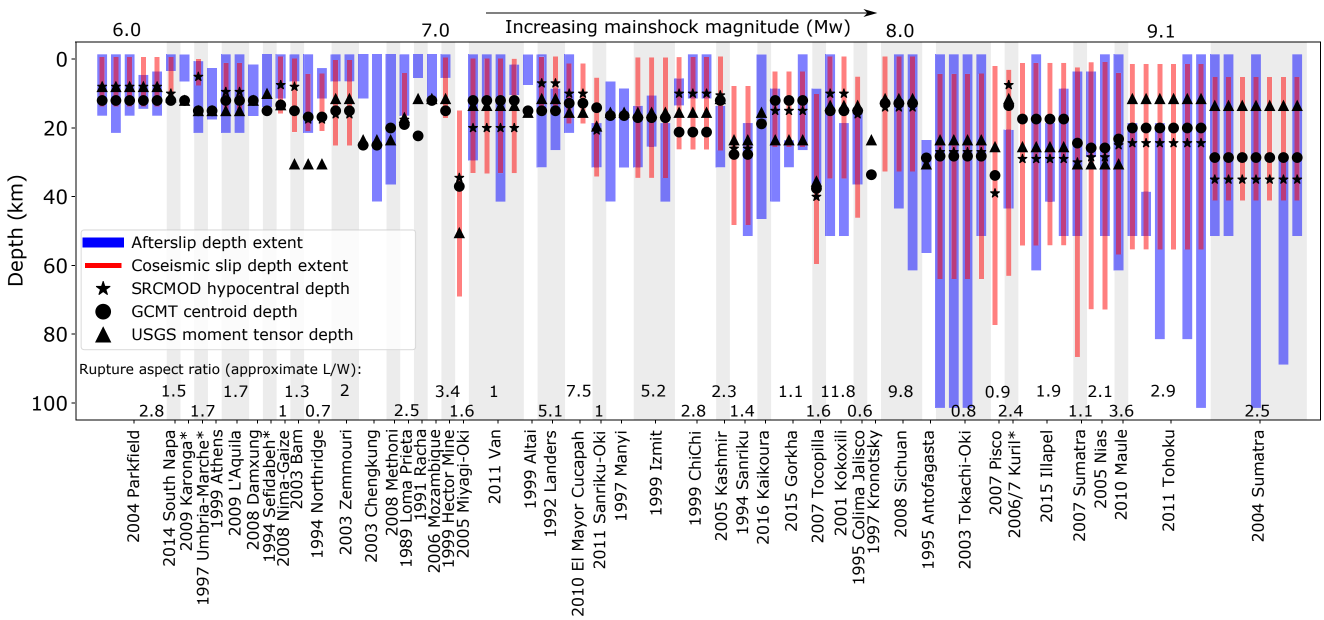


Figure 10. The approximate depth extents of aseismic afterslip for the 88 kinematic slip models studies and corresponding coseismic ruptures from SRCMOD coseismic slip models. Moment tensor depths from the USGS preferred solution, centroid depths from the Global Centroid Moment Tensor catalog, and rupture aspect ratios and hypocentral depths from SRCMOD coseismic slip models are also shown but may be erroneous in some cases (e.g., default values, not relocated). Not all afterslip depth extents, coseismic depth extents, hypocentral depths and rupture aspect ratios were available.

Fourteen earthquakes have a hypocenter at least 1 km shallower than the centroid (net-downdip propagating). Nine of these show some evidence of significant afterslip below coseismic rupture depths, but five do not. Similarly, seven earthquakes that cannot be described as net-downdip propagating also show evidence of significant downdip afterslip. Again, at a threshold of 5 km, only five earthquakes can be described as net-downdip propagating and only two of these show evidence of significant afterslip downdip of coseismic rupture. We, therefore, find no evidence that rupture directivity effects afterslip depth distribution, but our analysis is only one-dimensional; effective slip analysis would be useful for more insight here, but digitized afterslip models are scarcely provided.

5. Discussion

5.1. Afterslip and Coseismic Moment Scaling

We find that afterslip and coseismic moment scale approximately linearly, with a gradient close to one. We explore what this finding might mean for the average slip and area of afterslip, through assumptions grounded in elastic theory and self-similar rupture scaling. Whilst M_{rel} is distributed with an interquartile range of 0.09–0.32, to form a simple argument, we assume that M_{rel} can be approximated by a constant and rewrite Equation 1 as:

$$M_o^{aft} = M_{rel} M_o. \tag{7}$$

We can substitute Equation 4 into 7 and assume that the shear modulus (G) remains approximately constant across the seismogenic and afterslip zones when compared to variations in A and $\overline{D}$. The average slip $\left(\overline{D}^{aft}\right)$ and area (A^{aft}) of afterslip thus scale as:

$$A^{aft}\overline{D}^{aft} \sim A\overline{D} \sim M_o. \tag{8}$$

We can consider the scaling of A^{aft} and $\overline{D}^{aft}$ separately, by first considering an 'activated area' around a rupture that is primed for afterslip. For a simple circular rupture, stress change decays as the inverse of the distance cubed from the dislocation (Segall, 2010). Assuming that afterslip is entirely driven by coseismic static stress change, the distance (d^{aft}) to the minimum activating shear stress bounding the activated area (A^{aft}) scales with M_o as:

$$d^{aft} \sim M_o^{1/3}, \tag{9}$$

(Marsan, 2005). Squaring the equation gives an area A^{aft} that scales as:

$$A^{aft} \sim M_o^{2/3}, \tag{10}$$

with proof given in the supplementary materials. Given Equation 8, $\overline{D}^{aft}$ must therefore scale as:

$$\overline{D}^{aft} \sim M_o^{1/3}. \tag{11}$$

Our empirical finding (Equation 7) also allows us to substitute M_o^{aft} into Equations 10 and 11 in the place of M_o (although the constants of proportionality change):

$$A^{aft} \sim M_o^{aft\,2/3}, \tag{12}$$

$$\overline{D}^{aft} \sim M_o^{aft\,1/3}. \tag{13}$$

Equations 12 and 13 refer to afterslip area, slip, and moment, but are essentially equivalent to well established coseismic scaling relations (e.g., Allen & Hayes, 2017; Blaser et al., 2010; Hanks & Bakun, 2002; Leonard, 2010, 2014; Murotani et al., 2013; Skarlatoudis et al., 2016; Somerville et al., 1999; Strasser et al., 2010; Wells & Coppersmith, 1994). Interestingly, Michel et al. (2019) also proposed that the area of slow slip events follow a relationship equivalent to Equation 12, implying that the area of afterslip and generic slow slip events may scale similarly.

We propose that the afterslip area and average slip approximately obey the scaling relations given by Equations 10 and 11. Afterslip moment grows by a combination of coseismic area and average slip, thus with the overall coseismic moment. We believe these scaling relations provide a good first order approximation of afterslip behavior, around which secondary factors can cause variation. Our results indicate that characteristics such as rupture aspect ratio and fault slip rate (a potential proxy for fault zone maturity and composition) may influence M_{rel} and therefore cause systematic deviations from this scaling.

5.2. Mainshock and Fault Setting Factors

In this section, we discuss the potential influences of mainshock characteristics, fault setting, and the broader earthquake cycle on M_{rel}. We expect that afterslip occurrence is driven principally by how strongly and how much of the fault zone interface is velocity-strengthening ($a - b < 0$), and that $a - b$ is largely controlled by depth.

We find no evidence to support strong relationships between M_{rel} and mainshock rake, dip, or depth. This indicates that globally, neither mechanism nor depth overwhelmingly affects the afterslip moment and that our original hypothesis (that shallow ruptures and low fault dip angles may allow more afterslip) is not supported. However, our metrics may be overly simplistic or too insensitive to serve as good proxies for mainshock geometry, as they do not consider rupture shape, fault roughness, or kinks. It is also possible that any relationship is obscured by data or modeling uncertainties, which we discuss in Section 5.3.

The rupture aspect ratio may be a second-order control on afterslip. The correlation coefficient between M_{rel} and rupture aspect ratio is moderate −0.31, with an entirely negative 95% range. Rupture aspect ratio depends on characteristics such as nucleation area, local seismogenic thickness, and whether an earthquake is sufficiently large to interact with the edges of the seismogenic zone (Ampuero & Mao, 2017; Weng & Yang, 2017), thus is inherently linked to coseismic moment. This is seen in Figure 10, in which the largest rupture aspect ratios belong to larger continental earthquakes, which generally saturate the seismogenic zone around M_w6-7 (Hawthorne et al., 2016) and then elongate with increasing magnitude. Subduction interface events generally occur on much wider and lower dip angle faults (Anderson, 1905), thus form a separate population of rupture aspect ratios in Figure 10, but the overall relationship is still seen. The correlation coefficients between M_{rel} and rupture aspect ratio for continental- and subduction-only event populations are −0.34 and −0.24, respectively, similar to the overall value.

A relationship between rupture aspect ratio and M_{rel} may have more than one explanation. Hawthorne et al. (2016) suggested that larger, elongated ruptures may have a reduced capacity for relative afterslip compared to smaller, less elongated earthquakes, because of the relative size of the region surrounding the coseismic rupture that can undergo afterslip. Smaller and less elongate earthquakes may also generate more of their afterslip closer to the rheologically controlled seismic-aseismic transition, which has greater scope to vary from location to location, than larger, more elongate ruptures which generate more of their afterslip closer to the temperature-controlled brittle-ductile transition. However, this argument assumes that the seismic-aseismic transition is consistently above the brittle-ductile transition, which may not hold everywhere. A greater scope for relative afterslip variability in smaller earthquakes, combined with a publication bias whereby smaller earthquakes with larger M_{rel} are preferentially studied, provides one explanation for the relationship we observe but implies that it is (at least in part) due to the publication bias. The dependence of shear stress change on rupture stress drop (Segall, 2010) may provide an alternative, physical argument. For the same coseismic moment, a larger area and presumably more elongated rupture will have a lower stress drop, and thus a smaller average stress concentration at its edges than a less elongated, more compact earthquake. Assuming that afterslip occurs generally downdip, this could imply that less elongate ruptures are able to generate more (downdip) afterslip than more elongate ruptures. Whilst rupture aspect ratio is not independent of the coseismic moment, M_{rel} is more strongly correlated to rupture aspect ratio than it is to M_o, suggesting that rupture aspect ratio may provide some independent control on afterslip, although the specific reasoning is unclear.

M_{rel} correlates moderately with plate velocity and fault slip rate. Plate velocity, local strain rate, and fault slip rate are measures of deformation rate that we treat as proxies for fault maturity and high proportions of frictionally stable fault zone materials such as gouges and smoothed asperities (Choy & Kirby, 2004; Ikari et al., 2011). The moderate correlation between M_{rel} and plate velocity (0.39) is based on only 18 subduction interface earthquakes and is thus not particularly robust. The correlation between M_{rel} and strain rate for the remaining 28 continental events is a weak 0.09. The most significant finding is the moderate correlation between M_{rel} and fault slip rate (0.26) over the entire data set. Some geological evidence supports fault slip rate as a proxy for fault maturity (e.g., Goldberg et al., 2020; Manighetti et al., 2007), whilst reported slip rates may inadvertently be a good proxy of fault maturity, as measurements at immature faults may be systematically underestimated because the strain is less localized (Dolan & Haravitch, 2014). Regardless, the reported fault slip rate may be a reasonable first order proxy for maturity, and a weak to moderate indicator of M_{rel}.

Endmember case examples can link high fault slip rates, specific geological characteristics, and high M_{rel}. The highest M_{rel} estimates belong to the M_w6.0 2004 Parkfield and M_w6.8 2008 Methoni earthquakes, which have high fault slip rates, and high strain rates and plate velocities, respectively (see Figure 9). Near Parkfield, the fast creeping section of the San Andreas fault (Jolivet et al., 2015) contains several meters of highly velocity-strengthening material gouge and talc-bearing serpentinites (Johnson et al., 2006; Moore & Rymer, 2007; Savage & Langbein, 2008) which may explain the high M_{rel} and, perhaps, the relatively shallow afterslip observed (Bruhat et al., 2011; Johanson et al., 2006). In addition to this, Johanson et al. (2006) also posited that two M_w5 aftershocks may have served to unpin an additional, adjacent fault section and trigger enhanced afterslip which explains the high M_{rel}. High slip and strain rates might not be sufficient for abundant afterslip, however. Whilst our lowest M_{rel} earthquakes have relatively low fault slip rates (and strain rates and plate velocities), the M_w6.8 2003 Chengkung and the Japan Trench earthquakes have the highest strain and slip rates, respectively, but more moderate M_{rel}. Better estimates of lithology, rheology, and structure that can be used to describe the $a - b$ profile at a fault would be helpful to further assess this dependence.

So far, we have only considered contemporary factors, but M_{rel} might vary over multiple earthquake cycles at a given fault. Simulations of different ruptures on the same fault patch have shown penetration to variable depths (Jiang & Lapusta, 2016; Shaw & Wesnousky, 2008), which could theoretically affect the fault area left primed for afterslip in future earthquakes, assuming that frictional stability is principally controlled by depth. Postseismic behaviors have even been shown to vary at the same fault patch in some of these simulations (e.g., Barbot et al., 2012). Furthermore, as stress conditions evolve with tectonic loading, exactly when an earthquake occurs could affect its afterslip. For example, regions adjacent to an 'early' earthquake might require less afterslip to catch up with the surrounding interseismic creep, than for a 'late' earthquake. Studies of several quasi-periodic earthquake cycles at Parkfield have suggested this, indicating that the 1966 earthquake possibly produced more afterslip than the 1934 earthquake, which was 'early' (Segall & Du, 1993; Segall & Harris, 1987). However, data for these earthquakes and afterslip events are quite poor and the entire concept of quasi-periodic seismic cycles is debated (Kagan et al., 2012). Interseismic coupling may be an important factor in determining M_{rel}. More velocity-strengthening fault surfaces surrounding rupture are likely to allow both more interseismic creep and afterslip and be less conducive to larger seismic ruptures, thus interseismic coupling could potentially be an indicator of afterslip potential, but requires reliable estimates at every fault.

In summary, M_{rel} does not appear to be overwhelmingly affected by earthquake mechanism, fault dip, or depth, but may be favored by higher fault slip rates and lower rupture aspect ratios. The uncertainty in M_{rel} estimates for the same event discussed in Section 4.3 highlights that any of these relationships may be obscured by data and modeling uncertainty. We discuss this further below, but perhaps stronger or additional relationships could be established by observing more earthquakes in the same locations over time and attempting to model the afterslip in a systematic way.

5.3. Data and Modeling Factors

We have identified significant uncertainty in afterslip moment estimates that must be due to data and modeling factors. In this section, we explore factors within the modeling methodology that might have led to (a) the lack of strong relationships between M_{rel} and the start time and duration of observation across global and individual earthquake scales and (b) the observed variability in afterslip moment estimates. The identification of this uncertainty should lead to a more informed analysis of afterslip models and perhaps an effort to standardize afterslip modeling methodology to improve model comparability and help us to better understand aseismic afterslip.

Afterslip moment should tend toward an asymptotic limit with earlier and longer observation windows, but we did not see strong evidence for this globally. The theoretical importance of observational duration is highlighted in Figure 7 by the synthetic afterslip decay case and can also be seen within some individual examples (e.g., estimates for the M_w8.0 2008 Sichuan earthquake). The importance of an early start time is highlighted clearly in studies such as Jiang et al. (2021), who proposed that M_{rel} may have reached 34% within 24 hr of the 2004 M_w6.0 Parkfield earthquake. There are several potential explanations for the lack of these relationships in the data. First, other potential dependencies such as rupture aspect ratio and fault slip rate may contribute to obscuring temporal

10.1029/2021JB023897

relationships in analysis across different regions. Secondly, as afterslip can decay at different rates across different regions (Ingleby & Wright, 2017), 3 months of observation following one earthquake might capture a greater fraction of its afterslip than in 1 year following another; thus global-scale correlations between M_{rel} and observation duration may be obscured. However, if either (or both) of these arguments are true, we would still expect to see correlations within different estimates following the same earthquake and within similar regions. In Section 4.3, we show that this is often not the case and that modeling must therefore be a significant source of afterslip estimate uncertainty, ranging from a typical factor of ∼two or three to over an order of magnitude.

A number of modeling choices may contribute to more variable afterslip moment estimates. A total of 47 of the 88 kinematic afterslip models we analyze do not properly account for or reasonably consider additional postseismic mechanisms (i.e., they did not model viscoelastic relaxation or pore fluid effects, or indicate why this is not required), which could lead to erroneous afterslip moment estimates (e.g., McCormack et al., 2020; Sun & Wang, 2015). The implications of not considering viscoelastic relaxation could be especially significant. For example, following the M_w8.0 2008 Sichuan earthquake, M. Wang et al. (2021) suggested that an afterslip-only model produced an afterslip moment estimate several times that of a model that included viscoelasticity. Conversely, they also suggested that not considering afterslip in viscoelastic relaxation models can lead to incorrect inferred effective viscosities. Additional examples where the trade-off of afterslip and viscoelasticity in models may be significant include following the M_w7.8 2015 Gorkha earthquake (e.g., B. Zhao et al., 2017), the M_w7.9 2001 Kokoxili earthquake (e.g., D. Zhao et al., 2021) and the great M_w9.1 2011 Tohoku (e.g., Sun & Wang, 2015) and M_w9.1 2004 Sumatra (e.g., F. Pollitz et al., 2008) subduction thrust earthquakes. This may explain why uncertainty does not decrease with coseismic moment. When considering both mechanisms, separating their respective contributions is also a difficult problem, particularly in the lower crust (Jacobs et al., 2002; Luo & Wang, 2021). Modeling additional mechanisms also requires more complex rheological model spaces, thus additional free parameters (e.g., Bruhat et al., 2011; Muto et al., 2016; B. Zhao et al., 2017). The validity of different rheological spaces is an ongoing debate and an obvious source of uncertainty. Bedford et al. (2016) argue that the homogeneous, elastic half-space is established, acceptable and useful for modeling afterslip, whilst others (e.g., Hearn & Burgmann, 2005; Sun & Wang, 2015) propose that layered elastic and viscoelastic half-spaces are more valid and can recover more afterslip. Finally, the failure to remove the deformation signal due to aftershocks could lead to overestimates of the afterslip moment and distorted spatial models (Lange et al., 2014). Aftershocks are commonly ignored in afterslip studies due to a comparatively small cumulative moment (e.g., Diao et al., 2018). However, if particularly large aftershocks are not explicitly accounted for, this could amount to significant errors in afterslip moment estimates: we adjusted one estimate by Gahalaut et al. (2008) by 47%, but only because they explicitly stated this. We encourage researchers to reserve the term afterslip for a specific phenomenon outlined in Section 2, rather than generic postseimsic deformation.

Uncertainty surrounding different methodological practices remains a significant barrier to comparing afterslip models. More general sources include the non-uniqueness and regularization inherent to the inversion process (Menke, 2018; Scales & Tenorio, 2001), approximations of topography and fault geometry, and data sensitivities, resolution, and distribution (Marchandon et al., 2021). For example, InSAR may often miss early afterslip or struggle to detect far-field deformation resulting from deep afterslip (Marchandon et al., 2021; Wimpenny et al., 2017). Many of the modeling choices outlined in this section are compiled and summarized in our database for further investigation. A push toward the standardization of kinematic afterslip modeling methods would help improve the comparability of afterslip models and allow better deductions of afterslip behaviors, fault zone structure, and the relationship between afterslip and aftershock sequences. Many specific best modeling practices are still unclear and require further research before implementation, such as how appropriate different rheological models spaces are for modeling postseismic mechanisms. However, we recommend transparency and explicit quantification of parameters and uncertainties, the provision of digital afterslip models (if possible) for further analyses, and a push toward standardized data quality and temporal observation windows (i.e., an effort to start observation periods as early as possible and ensure a long duration), while recognizing that this is not always possible.

6. Conclusion

We compile a database of 148 afterslip studies after 53 earthquakes, containing detailed information on mainshock characteristics, modeling methods, and outputs (doi:10.5281/zenodo.6414330). By analyzing a subset of 88 well-constrained kinematic slip models, we find that: (a) coseismic moment is the principal control on the ensuing afterslip moment, which scales near-linearly with a median value of 18% of the coseismic moment, (b) relative afterslip moment (M_{rel}) varies from less than 1% to over 300% of the coseismic moment, with an inter-quartile range of 9%–32%, (c) global variation in M_{rel} cannot be accounted for by variation in factors such as fault dip, rake, and depth, (d) global variation in M_{rel} may be related to rupture aspect ratio and fault slip rate (which might be indicative of fault maturity), (e) there is an unexpected lack of strong, correlation between M_{rel} and the start time and duration of observation window on global scales, which could be obscured by other relationships or because afterslip decays sufficiently differently in different regions. However, as differences in start time and duration of observation window cannot always account for different M_{rel} estimates by different studies following the same earthquake, we infer that: (f) there is significant, up to order-of-magnitude uncertainty in afterslip moment estimates related to the modeling process, which currently provides a barrier to systematic comparison. Our database and analysis help expose the current uncertainty in afterslip moment estimates and hopefully encourage the community to consider standardizing processes to provide increased ability to compare studies. Such comparisons can better constrain variability in afterslip behaviors, and deduce their controls. Understanding the controls on afterslip moment may allow the eventual incorporation of afterslip as a source of postseismic stress transfer in aftershock sequence hazard models.

Data Availability Statement

Data used in this study are accessible through U.S. Geological Survey (2017), Dziewonski et al. (1981); Ekström et al. (2012), Mai and Thingbaijam (2014), Styron and Pagani (2020), Kreemer et al. (2014), UNAVCO (2021), as indicated in text, and through the database (doi:10.5281/zenodo.6414330).

Acknowledgments

R. M. Churchill is supported by a NERC GW4+ Doctoral Training Partnership studentship from the Natural Environment Research Council [NE/L002434/1]. M. J. Werner received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 821115, Real- Time Earthquake Risk Reduction for a Resilient Europe (RISE, http://www.rise-eu.org/home/). J. Biggs and M. J. Werner. are funded by the NERC Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics (COMET, http://comet.nerc.ac.uk), a partnership between UK Universities and the British Geological Survey. J. Biggs is also funded by a Leverhulme Prize (PLP-2018-362). A. Fagereng received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715836 "MICA"). We are grateful for constructive comments from Roland Bürgmann and an anonymous reviewer that helped improve this paper.

References

Allen, T. I., & Hayes, G. P. (2017). Alternative rupture-scaling relationships for subduction interface and other offshore environments. *Bulletin of the Seismological Society of America*, *107*(3), 1240–1253. https://doi.org/10.1785/0120160255

Alwahedi, M. A., & Hawthorne, J. C. (2019). Intermediate-magnitude postseismic slip follows intermediate-magnitude (m 4 to 5) earthquakes in California. *Geophysical Research Letters*, *46*(7), 3676–3687. https://doi.org/10.1029/2018gl081001

Ampuero, J. P., & Mao, X. (2017). Upper limit on damage zone thickness controlled by seismogenic depth. *Fault zone dynamic processes: Evolution of fault properties during seismic rupture*, (Vol. 227, pp. 243–253). https://doi.org/10.1002/9781119156895.ch13

Anderson, E. M. (1905). The dynamics of faulting. *Transactions of the Edinburgh Geological Society*, *8*(3), 387–402. https://doi.org/10.1144/transed.8.3.387

Avouac, J.-P. (2015). From geodetic imaging of seismic and aseismic fault slip to dynamic modeling of the seismic cycle. *Annual Review of Earth and Planetary Sciences*, *43*, 233–271. https://doi.org/10.1146/annurev-earth-060614-105302

Azúa, B. M., DeMets, C., & Masterlark, T. (2002). Strong interseismic coupling, fault afterslip, and viscoelastic flow before and after the Oct. 9, 1995 Colima-Jalisco earthquake: Continuous gps measurements from Colima, Mexico. *Geophysical Research Letters*, *29*(8), 122-1–122-4. https://doi.org/10.1029/2002gl014702

Bakun, W., Aagaard, B., Dost, B., Ellsworth, W., Hardebeck, J., Harris, R., et al. (2005). Implications for prediction and hazard assessment from the 2004 parkfield earthquake. *Nature*, *437*(7061), 969–974. https://doi.org/10.1038/nature04067

Barbot, S., & Fialko, Y. (2010). A unified continuum representation of post-seismic relaxation mechanisms: Semi-analytic models of afterslip, poroelastic rebound and viscoelastic flow. *Geophysical Journal International*, *182*(3), 1124–1140. https://doi.org/10.1111/j.1365-246x.2010.04678.x

Barbot, S., Hamiel, Y., & Fialko, Y. (2008). Space geodetic investigation of the coseismic and postseismic deformation due to the 2003 mw7. 2 altai earthquake: Implications for the local lithospheric rheology. *Journal of Geophysical Research*, *113*(B3), B03403. https://doi.org/10.1029/2007jb005063

Barbot, S., Lapusta, N., & Avouac, J.-P. (2012). Under the hood of the earthquake machine: Toward predictive modeling of the seismic cycle. *Science*, *336*(6082), 707–710. https://doi.org/10.1126/science.1218796

Barnhart, W. D., Murray, J. R., Briggs, R. W., Gomez, F., Miles, C. P., Svarc, J., et al. (2016). Coseismic slip and early afterslip of the 2015 illapel, Chile, earthquake: Implications for frictional heterogeneity and coastal uplift. *Journal of Geophysical Research: Solid Earth*, *121*(8), 6172–6191. https://doi.org/10.1002/2016jb013124

Bedford, J., Moreno, M., Li, S., Oncken, O., Baez, J. C., Bevis, M., et al. (2016). Separating rapid relocking, afterslip, and viscoelastic relaxation: An application of the postseismic straightening method to the maule 2010 cgps. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7618–7638. https://doi.org/10.1002/2016jb013093

Béjar-Pizarro, M., Carrizo, D., Socquet, A., Armijo, R., Barrientos, S., Bondoux, F., et al. (2010). Asperities and barriers on the seismogenic zone in north Chile: State-of-the-art after the 2007 m w 7.7 tocopilla earthquake inferred by gps and insar data. *Geophysical Journal International*, *183*(1), 390–406. https://doi.org/10.1111/j.1365-246x.2010.04748.x

Bie, L., Ryder, I., Nippress, S. E., & Bürgmann, R. (2014). Coseismic and post-seismic activity associated with the 2008 m w 6.3 damxung earthquake, tibet, constrained by insar. *Geophysical Journal International*, *196*(2), 788–803. https://doi.org/10.1093/gji/ggt444

Biggs, J., Bürgmann, R., Freymueller, J. T., Lu, Z., Parsons, B., Ryder, I., et al. (2009). The postseismic response to the 2002 m 7.9 denali fault earthquake: Constraints from insar 2003–2005. *Geophysical Journal International*, *176*(2), 353–367. https://doi.org/10.1111/j.1365-246x.2008.03932.x

Blanpied, M., Lockner, D., & Byerlee, J. (1991). Fault stability inferred from granite sliding experiments at hydrothermal conditions. *Geophysical Research Letters*, *18*(4), 609–612. https://doi.org/10.1029/91gl00469

Blaser, L., Krüger, F., Ohrnberger, M., & Scherbaum, F. (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, *100*(6), 2914–2926. https://doi.org/10.1785/0120100111

Boatwright, J., & Cocco, M. (1996). Frictional constraints on crustal faulting. *Journal of Geophysical Research*, *101*(B6), 13895–13909. https://doi.org/10.1029/96jb00405

Bruhat, L., Barbot, S., & Avouac, J.-P. (2011). Evidence for postseismic deformation of the lower crust following the 2004 mw6. 0 parkfield earthquake. *Journal of Geophysical Research*, *116*(B8), B08401. https://doi.org/10.1029/2010jb008073

Bürgmann, R. (2018). The geophysics, geology and mechanics of slow fault slip. *Earth and Planetary Science Letters*, *495*, 112–134. https://doi.org/10.1016/j.epsl.2018.04.062

Bürgmann, R., Ergintav, S., Segall, P., Hearn, E., McClusky, S., Reilinger, R. E., et al. (2002). Time-dependent distributed afterslip on and deep below the izmit earthquake rupture. *Bulletin of the Seismological Society of America*, *92*(1), 126–137. https://doi.org/10.1785/0120000833

Bürgmann, R., Segall, P., Lisowski, M., & Svarc, J. (1997). Postseismic strain following the 1989 loma prieta earthquake from gps and leveling measurements. *Journal of Geophysical Research*, *102*(B3), 4933–4955. https://doi.org/10.1029/96jb03171

Cattania, C., Hainzl, S., Wang, L., Enescu, B., & Roth, F. (2015). Aftershock triggering by postseismic stresses: A study based on Coulomb rate-and-state models. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2388–2407. https://doi.org/10.1002/2014jb011500

Chaussard, E., Bürgmann, R., Fattahi, H., Johnson, C., Nadeau, R., Taira, T., & Johanson, I. (2015). Interseismic coupling and refined earthquake potential on the hayward-calaveras fault zone. *Journal of Geophysical Research: Solid Earth*, *120*(12), 8570–8590. https://doi.org/10.1002/2015jb012230

Choy, G. L., & Kirby, S. H. (2004). Apparent stress, fault maturity and seismic hazard for normal-fault earthquakes at subduction zones. *Geophysical Journal International*, *159*(3), 991–1012. https://doi.org/10.1111/j.1365-246x.2004.02449.x

Collettini, C., Tesei, T., Scuderi, M. M., Carpenter, B. M., & Viti, C. (2019). Beyond byerlee friction, weak faults and implications for slip behavior. *Earth and Planetary Science Letters*, *519*, 245–263. https://doi.org/10.1016/j.epsl.2019.05.011

Copley, A. (2014). Postseismic afterslip 30 years after the 1978 tabas-e-golshan (Iran) earthquake: Observations and implications for the geological evolution of thrust belts. *Geophysical Journal International*, *197*(2), 665–679. https://doi.org/10.1093/gji/ggu023

D'Agostino, N., Cheloni, D., Fornaro, G., Giuliani, R., & Reale, D. (2012). Space-time distribution of afterslip following the 2009 l'aquila earthquake. *Journal of Geophysical Research*, *117*(B2), B02402.

Deng, J., Gurnis, M., Kanamori, H., & Hauksson, E. (1998). Viscoelastic flow in the lower crust after the 1992 landers, California, earthquake. *Science*, *282*(5394), 1689–1692. https://doi.org/10.1126/science.282.5394.1689

Diao, F., Wang, R., Wang, Y., Xiong, X., & Walter, T. R. (2018). Fault behavior and lower crustal rheology inferred from the first seven years of postseismic gps data after the 2008 wenchuan earthquake. *Earth and Planetary Science Letters*, *495*, 202–212. https://doi.org/10.1016/j.epsl.2018.05.020

Diao, F., Xiong, X., Wang, R., Zheng, Y., Walter, T. R., Weng, H., & Li, J. (2014). Overlapping post-seismic deformation processes: Afterslip and viscoelastic relaxation following the 2011 m w 9.0 Tohoku (Japan) earthquake. *Geophysical Journal International*, *196*(1), 218–229. https://doi.org/10.1093/gji/ggt376

Dieterich, J. H. (1979). Modeling of rock friction: 1. Experimental results and constitutive equations. *Journal of Geophysical Research*, *84*(B5), 2161–2168. https://doi.org/10.1029/jb084ib05p02161

Dieterich, J. H. (1987). Nucleation and triggering of earthquake slip: Effect of periodic stresses. *Tectonophysics*, *144*(1–3), 127–139. https://doi.org/10.1016/0040-1951(87)90012-6

Dodge, Y. (2008). *Spearman rank correlation coefficient. The Concise Encyclopedia of Statistics* (pp. 502–505). Springer.

Dolan, J. F., & Haravitch, B. D. (2014). How well do surface slip measurements track slip at depth in large strike-slip earthquakes? The importance of fault structural maturity in controlling on-fault slip versus off-fault surface deformation. *Earth and Planetary Science Letters*, *388*, 38–47. https://doi.org/10.1016/j.epsl.2013.11.043

Donnellan, A., Parker, J. W., & Peltzer, G. (2002). Combined gps and insar models of postseismic deformation from the northridge earthquake. *Pure and Applied Geophysics*, *159*(10), 2261–2270. https://doi.org/10.1007/s00024-002-8734-7

Dziewonski, A., Chou, T.-A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, *86*(B4), 2825–2852. https://doi.org/10.1029/jb086ib04p02825

Ekström, G., Nettles, M., & Dziewoński, A. (2012). The global cmt project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, *200*, 1–9. https://doi.org/10.1016/j.pepi.2012.04.002

Ergintav, S., Doğan, U., Gerstenecker, C., Cakmak, R., Belgen, A., Demirel, H., et al. (2007). A snapshot (2003–2005) of the 3d postseismic deformation for the 1999, mw= 7.4 izmit earthquake in the Marmara region, Turkey, by first results of joint gravity and gps monitoring. *Journal of Geodynamics*, *44*(1–2), 1–18. https://doi.org/10.1016/j.jog.2006.12.005

Fialko, Y. (2004). Evidence of fluid-filled upper crust from observations of postseismic deformation due to the 1992 mw7. 3 landers earthquake. *Journal of Geophysical Research*, *109*(B8), B08401. https://doi.org/10.1029/2004jb002985

Freed, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, *33*, 335–367. https://doi.org/10.1146/annurev.earth.33.092203.122505

Freed, A. M. (2007). Afterslip (and only afterslip) following the 2004 parkfield, California, earthquake. *Geophysical Research Letters*, *34*(6). https://doi.org/10.1029/2006gl029155

Freed, A. M., & Lin, J. (2001). Delayed triggering of the 1999 hector mine earthquake by viscoelastic stress transfer. *Nature*, *411*(6834), 180–183. https://doi.org/10.1038/35075548

Gahalaut, V., Jade, S., Catherine, J., Gireesh, R., Ananda, M., Kumar, P., et al. (2008). Gps measurements of postseismic deformation in the Andaman-Nicobar region following the giant 2004 Sumatra-Andaman earthquake. *Journal of Geophysical Research*, *113*(B8), B08401. https://doi.org/10.1029/2007jb005511

Goldberg, D. E., Melgar, D., Sahakian, V., Thomas, A., Xu, X., Crowell, B., & Geng, J. (2020). Complex rupture of an immature fault zone: A simultaneous kinematic model of the 2019 ridgecrest, ca earthquakes. *Geophysical Research Letters*, *47*(3), e2019GL086382. https://doi.org/10.1029/2019gl086382

Hamiel, Y., Baer, G., Kalindekafe, L., Dombola, K., & Chindandali, P. (2012). Seismic and aseismic slip evolution and deformation associated with the 2009–2010 northern Malawi earthquake swarm, east african rift. *Geophysical Journal International*, *191*(3), 898–908.

Hanks, T. C., & Bakun, W. H. (2002). A bilinear source-scaling model for m-log a observations of continental earthquakes. *Bulletin of the Seismological Society of America*, *92*(5), 1841–1846. https://doi.org/10.1785/0120010148

Hanks, T. C., & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research*, *84*(B5), 2348–2350. https://doi.org/10.1029/jb084ib05p02348

Harris, R. A. (1998). Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *Journal of Geophysical Research*, *103*(B10), 24347–24358. https://doi.org/10.1029/98jb01576

Harris, R. A. (2017). Large earthquakes and creeping faults. *Reviews of Geophysics*, *55*(1), 169–198. https://doi.org/10.1002/2016rg000539

Hawthorne, J., Simons, M., & Ampuero, J.-P. (2016). Estimates of aseismic slip associated with small earthquakes near san juan bautista, ca. *Journal of Geophysical Research: Solid Earth*, *121*(11), 8254–8275. https://doi.org/10.1002/2016jb013120

Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, *468*, 94–100. https://doi.org/10.1016/j.epsl.2017.04.003

Hearn, E., & Burgmann, R. (2005). The effect of elastic layering on inversions of gps data for coseismic slip and resulting stress changes: Strike-slip earthquakes. *Bulletin of the Seismological Society of America*, *95*(5), 1637–1653. https://doi.org/10.1785/0120040158

Helmstetter, A., & Shaw, B. E. (2009). Afterslip and aftershocks in the rate-and-state friction law. *Journal of Geophysical Research*, *114*(B1), B01308. https://doi.org/10.1029/2007jb005077

Hillers, G., Ben-Zion, Y., & Mai, P. (2006). Seismicity on a fault controlled by rate-and state-dependent friction with spatial variations of the critical slip distance. *Journal of Geophysical Research*, *111*(B1), B01403. https://doi.org/10.1029/2005jb003859

Hoffmann, F., Metzger, S., Moreno, M., Deng, Z., Sippl, C., Ortega-Culaciati, F., & Oncken, O. (2018). Characterizing afterslip and ground displacement rate increase following the 2014 iquique-pisagua mw 8.1 earthquake, northern Chile. *Journal of Geophysical Research: Solid Earth*, *123*(5), 4171–4192. https://doi.org/10.1002/2017jb014970

Howell, A., Palamartchouk, K., Papanikolaou, X., Paradissis, D., Raptakis, C., Copley, A., et al. (2017). The 2008 methoni earthquake sequence: The relationship between the earthquake cycle on the subduction interface and coastal uplift in sw Greece. *Geophysical Journal International*, *208*(3), 1592–1610. https://doi.org/10.1093/gji/ggw462

Hsu, Y.-J., Simons, M., Avouac, J.-P., Galetzka, J., Sieh, K., Chlieh, M., et al. (2006). Frictional afterslip following the 2005 nias-simeulue earthquake, sumatra. *Science*, *312*(5782), 1921–1926. https://doi.org/10.1126/science.1126960

Huang, H., Xu, W., Meng, L., Bürgmann, R., & Baez, J. C. (2017). Early aftershocks and afterslip surrounding the 2015 mw 8.4 illapel rupture. *Earth and Planetary Science Letters*, *457*, 282–291. https://doi.org/10.1016/j.epsl.2016.09.055

Hutton, W., DeMets, C., Sanchez, O., Suarez, G., & Stock, J. (2001). Slip kinematics and dynamics during and after the 1995 october 9 mw= 8.0 colima–jalisco earthquake, Mexico, from gps geodetic constraints. *Geophysical Journal International*, *146*(3), 637–658. https://doi.org/10.1046/j.1365-246x.2001.00472.x

Ikari, M. J., Marone, C., & Saffer, D. M. (2011). On the relation between fault strength and frictional stability. *Geology*, *39*(1), 83–86. https://doi.org/10.1130/g31416.1

Imber, J., Holdsworth, R., Butler, C., & Strachan, R. (2001). A reappraisal of the sibson-scholz fault zone model: The nature of the frictional to viscous ("brittle-ductile") transition along a long-lived, crustal-scale fault, outer hebrides, scotland. *Tectonics*, *20*(5), 601–624. https://doi.org/10.1029/2000tc001250

Imber, J., Holdsworth, R., Smith, S., Jefferies, S., & Collettini, C. (2008). Frictional-viscous flow, seismicity and the geology of weak faults: A review and future directions. *Geological Society, London, Special Publications*, *299*(1), 151–173. https://doi.org/10.1144/sp299.10

Ingleby, T., & Wright, T. (2017). Omori-like decay of postseismic velocities following continental earthquakes. *Geophysical Research Letters*, *44*(7), 3119–3130. https://doi.org/10.1002/2017gl072865

Jacobs, A., Sandwell, D., Fialko, Y., & Sichoix, L. (2002). The 1999 (m w 7.1) hector mine, California, earthquake: Near-field postseismic deformation from ers interferometry. *Bulletin of the Seismological Society of America*, *92*(4), 1433–1442. https://doi.org/10.1785/0120000908

Jiang, J., Bock, Y., & Klein, E. (2021). Coevolving early afterslip and aftershock signatures of a san andreas fault rupture. *Science Advances*, *7*(15), eabc1606. https://doi.org/10.1126/sciadv.abc1606

Jiang, J., & Lapusta, N. (2016). Deeper penetration of large earthquakes on seismically quiescent faults. *Science*, *352*(6291), 1293–1297. https://doi.org/10.1126/science.aaf1496

Johanson, I. A., Fielding, E. J., Rolandone, F., & Burgmann, R. (2006). Coseismic and postseismic slip of the 2004 parkfield earthquake from space-geodetic data. *Bulletin of the Seismological Society of America*, *96*(4B), S269–S282. https://doi.org/10.1785/0120050818

Johnson, K. M., Bürgmann, R., & Freymueller, J. T. (2009). Coupled afterslip and viscoelastic flow following the 2002 denali fault, Alaska earthquake. *Geophysical Journal International*, *176*(3), 670–682. https://doi.org/10.1111/j.1365-246x.2008.04029.x

Johnson, K. M., Bürgmann, R., & Larson, K. (2006). Frictional properties on the san andreas fault near parkfield, California, inferred from models of afterslip following the 2004 earthquake. *Bulletin of the Seismological Society of America*, *96*(4B), S321–S338. https://doi.org/10.1785/0120050808

Jolivet, R., Simons, M., Agram, P., Duputel, Z., & Shen, Z.-K. (2015). Aseismic slip and seismogenic coupling along the central San Andreas fault. *Geophysical Research Letters*, *42*(2), 297–306. https://doi.org/10.1002/2014gl062222

Kagan, Y., Jackson, D. D., & Geller, R. J. (2012). *Characteristic earthquake model, 1884–2011, rip*. arXiv preprint arXiv:1207.4836.

Kaneko, Y., Avouac, J.-P., & Lapusta, N. (2010). Towards inferring earthquake patterns from geodetic observations of interseismic coupling. *Nature Geoscience*, *3*(5), 363–369. https://doi.org/10.1038/ngeo843

Kreemer, C., Blewitt, G., & Klein, E. C. (2014). A geodetic plate motion and global strain rate model. *Geochemistry, Geophysics, Geosystems*, *15*(10), 3849–3889. https://doi.org/10.1002/2014gc005407

Langbein, J., Murray, J., & Snyder, H. (2006). Coseismic and initial postseismic deformation from the 2004 parkfield, California, earthquake, observed by global positioning system, electronic distance meter, creepmeters, and borehole strainmeters. *Bulletin of the Seismological Society of America*, *96*(4B), S304–S320. https://doi.org/10.1785/0120050823

Lange, D., Bedford, J., Moreno, M., Tilmann, F., Baez, J., Bevis, M., & Kruger, F. (2014). Comparison of postseismic afterslip models with aftershock seismicity for three subduction-zone earthquakes: Nias 2005, maule 2010 and tohoku 2011. *Geophysical Journal International*, *199*(2), 784–799. https://doi.org/10.1093/gji/ggu292

Leonard, M. (2010). Earthquake fault scaling: Self-consistent relating of rupture length, width, average displacement, and moment release. *Bulletin of the Seismological Society of America*, *100*(5A), 1971–1988. https://doi.org/10.1785/0120090189

Leonard, M. (2014). Self-consistent earthquake fault-scaling relations: Update and extension to stable continental strike-slip faults. *Bulletin of the Seismological Society of America*, *104*(6), 2953–2965. https://doi.org/10.1785/0120140087

Luo, H., & Wang, K. (2021). Postseismic geodetic signature of cold forearc mantle in subduction zones. *Nature Geoscience*, *14*(2), 104–109. https://doi.org/10.1038/s41561-020-00679-9

Mai, P. M., & Thingbaijam, K. (2014). Srcmod: An online database of finite-fault rupture models. *Seismological Research Letters*, *85*(6), 1348–1357. https://doi.org/10.1785/0220140077

Mancini, S., Segou, M., Werner, M. J., & Parsons, T. (2020). The predictive skills of elastic Coulomb rate-and-state aftershock forecasts during the 2019 ridgecrest, California, earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1736–1751. https://doi.org/10.1785/0120200028

Manighetti, I., Campillo, M., Bouley, S., & Cotton, F. (2007). Earthquake scaling, fault segmentation, and structural maturity. *Earth and Planetary Science Letters*, *253*(3–4), 429–438. https://doi.org/10.1016/j.epsl.2006.11.004

Marchandon, M., Hollingsworth, J., & Radiguet, M. (2021). Origin of the shallow slip deficit on a strike slip fault: Influence of elastic structure, topography, data coverage, and noise. *Earth and Planetary Science Letters*, *554*, 116696. https://doi.org/10.1016/j.epsl.2020.116696
Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, *26*(1), 643–696. https://doi.org/10.1146/annurev.earth.26.1.643
Marone, C., Scholtz, C., & Bilham, R. (1991). On the mechanics of earthquake afterslip. *Journal of Geophysical Research*, *96*(B5), 8441–8452. https://doi.org/10.1029/91jb00275
Marsan, D. (2005). The role of small earthquakes in redistributing crustal elastic stress. *Geophysical Journal International*, *163*(1), 141–151. https://doi.org/10.1111/j.1365-246x.2005.02700.x
Massonnet, D., Feigl, K., Rossi, M., & Adragna, F. (1994). Radar interferometric mapping of deformation in the year after the landers earthquake. *Nature*, *369*(6477), 227–230. https://doi.org/10.1038/369227a0
McCormack, K., Hesse, M. A., Dixon, T., & Malservisi, R. (2020). Modeling the contribution of poroelastic deformation to postseismic geodetic signals. *Geophysical Research Letters*, *47*(8), e2020GL086945. https://doi.org/10.1029/2020gl086945
Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory*. Academic press.
Michel, S., Gualandi, A., & Avouac, J.-P. (2019). Similar scaling laws for earthquakes and cascadia slow-slip events. *Nature*, *574*(7779), 522–526. https://doi.org/10.1038/s41586-019-1673-6
Miura, S., Iinuma, T., Yui, S., Uchida, N., Sato, T., Tachibana, K., & Hasegawa, A. (2006). Co-and post-seismic slip associated with the 2005 miyagi-oki earthquake (m7. 2) as inferred from gps data. *Earth Planets and Space*, *58*(12), 1567–1572. https://doi.org/10.1186/bf03352662
Moore, D. E., & Rymer, M. J. (2007). Talc-bearing serpentinite and the creeping section of the san andreas fault. *Nature*, *448*(7155), 795–797. https://doi.org/10.1038/nature06064
Murotani, S., Satake, K., & Fujii, Y. (2013). Scaling relations of seismic moment, rupture area, average slip, and asperity size for m˜ 9 subduction-zone earthquakes. *Geophysical Research Letters*, *40*(19), 5070–5074. https://doi.org/10.1002/grl.50976
Muto, J., Shibazaki, B., Iinuma, T., Ito, Y., Ohta, Y., Miura, S., & Nakai, Y. (2016). Heterogeneous rheology controlled postseismic deformation of the 2011 tohoku-oki earthquake. *Geophysical Research Letters*, *43*(10), 4971–4978. https://doi.org/10.1002/2016gl068113
Niemeijer, A. (2018). Velocity-dependent slip weakening by the combined operation of pressure solution and foliation development. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-22889-3
Nishimura, T., & Thatcher, W. (2003). Rheology of the lithosphere inferred from postseismic uplift following the 1959 hebgen lake earthquake. *Journal of Geophysical Research*, *108*(B8), 2389. https://doi.org/10.1029/2002jb002191
Noda, H., & Lapusta, N. (2013). Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*, *493*(7433), 518–521. https://doi.org/10.1038/nature11703
Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, *82*(2), 1018–1040. https://doi.org/10.1785/bssa0820021018
Paul, J., Lowry, A., Bilham, R., Sen, S., & Smalley, R., Jr. (2007). Postseismic deformation of the andaman islands following the 26 december, 2004 great sumatra–andaman earthquake. *Geophysical Research Letters*, *34*(19), L19309. https://doi.org/10.1029/2007gl031024
Peltzer, G., Rosen, P., Rogez, F., & Hudnut, K. (1996). Postseismic rebound in fault step-overs caused by pore fluid flow. *Science*, *273*(5279), 1202–1204. https://doi.org/10.1126/science.273.5279.1202
Peltzer, G., Rosen, P., Rogez, F., & Hudnut, K. (1998). Poroelastic rebound along the landers 1992 earthquake surface rupture. *Journal of Geophysical Research*, *103*(B12), 30131–30145. https://doi.org/10.1029/98jb02302
Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 parkfield earthquake. *Nature Geoscience*, *2*(12), 877. https://doi.org/10.1038/ngeo697
Perfettini, H., & Avouac, J.-P. (2004). Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the chi-chi earthquake, taiwan. *Journal of Geophysical Research*, *109*,B02304. https://doi.org/10.1029/2003jb002488
Perfettini, H., & Avouac, J.-P. (2007). Modeling afterslip and aftershocks following the 1992 landers earthquake. *Journal of Geophysical Research*, *112*(B7), B07409. https://doi.org/10.1029/2006jb004399
Perfettini, H., Avouac, J.-P., Tavera, H., Kositsky, A., Nocquet, J.-M., Bondoux, F., et al. (2010). Seismic and aseismic slip on the central Peru megathrust. *Nature*, *465*(7294), 78. https://doi.org/10.1038/nature09062
Piombo, A., Martinelli, G., & Dragoni, M. (2005). Post-seismic fluid flow and Coulomb stress changes in a poroelastic medium. *Geophysical Journal International*, *162*(2), 507–515. https://doi.org/10.1111/j.1365-246x.2005.02673.x
Pollitz, F., Banerjee, P., Grijalva, K., Nagarajan, B., & Bürgmann, R. (2008). Effect of 3-d viscoelastic structure on post-seismic relaxation from the 2004 m = 9.2 sumatra earthquake. *Geophysical Journal International*, *173*(1), 189–204. https://doi.org/10.1111/j.1365-246x.2007.03666.x
Pollitz, F. F., Wicks, C., & Thatcher, W. (2001). Mantle flow beneath a continental strike-slip fault: Postseismic deformation after the 1999 hector mine earthquake. *Science*, *293*(5536), 1814–1818. https://doi.org/10.1126/science.1061361
Reilinger, R. (1984). Coseismic and postseismic vertical movements associated with the 1940 m7. 1 imperial valley, California, earthquake. *Journal of Geophysical Research*, *89*(B6), 4531–4537. https://doi.org/10.1029/jb089ib06p04531
Reilinger, R., Ergintav, S., Bürgmann, R., McClusky, S., Lenk, O., Barka, A., et al. (2000). Coseismic and postseismic fault slip for the 17 august 1999, m = 7.5, izmit, Turkey earthquake. *Science*, *289*(5484), 1519–1524. https://doi.org/10.1126/science.289.5484.1519
Reilinger, R., & Larsen, S. (1986). Vertical crustal deformation associated with the 1979 m = 6.6 imperial valley, California earthquake: Implications for fault behavior. *Journal of Geophysical Research*, *91*(B14), 14044–14056. https://doi.org/10.1029/jb091ib14p14044
Riva, R., Borghi, A., Aoudia, A., Barzaghi, R., Sabadini, R., & Panza, G. (2007). Viscoelastic relaxation and long-lasting after-slip following the 1997 umbria-marche (central Italy) earthquakes. *Geophysical Journal International*, *169*(2), 534–546. https://doi.org/10.1111/j.1365-246x.2007.03315.x
Rolandone, F., Nocquet, J.-M., Mothes, P. A., Jarrin, P., Vallée, M., Cubas, N., et al. (2018). Areas prone to slow slip events impede earthquake rupture propagation and promote afterslip. *Science Advances*, *4*(1), eaao6596. https://doi.org/10.1126/sciadv.aao6596
Rollins, C., Barbot, S., & Avouac, J.-P. (2015). Postseismic deformation following the 2010 m = 7.2 el mayor-cucapah earthquake: Observations, kinematic inversions, and dynamic models. *Pure and Applied Geophysics*, *172*(5), 1305–1358. https://doi.org/10.1007/s00024-014-1005-6
Rubin, A. M. (2008). Episodic slow slip events and rate-and-state friction. *Journal of Geophysical Research*, *113*(B11), B11414. https://doi.org/10.1029/2008jb005642
Rubin, A. M., & Ampuero, J.-P. (2005). Earthquake nucleation on (aging) rate and state faults. *Journal of Geophysical Research*, *110*(B11), B11312. https://doi.org/10.1029/2005jb003686
Ruina, A. (1983). Slip instability and state variable friction laws. *Journal of Geophysical Research*, *88*(B12), 10359–10370. https://doi.org/10.1029/jb088ib12p10359

Ryder, I., Parsons, B., Wright, T. J., & Funning, G. J. (2007). Post-seismic motion following the 1997 manyi (tibet) earthquake: Insar observations and modelling. *Geophysical Journal International*, *169*(3), 1009–1027. https://doi.org/10.1111/j.1365-246x.2006.03312.x

Savage, J., & Langbein, J. (2008). Postearthquake relaxation after the 2004 m6 parkfield, California, earthquake and rate-and-state friction. *Journal of Geophysical Research*, *113*(B10), B10407. https://doi.org/10.1029/2008jb005723

Savage, J., & Svarc, J. (2009). Postseismic relaxation following the 1992 m7. 3 landers and 1999 m7. 1 hector mine earthquakes, southern California. *Journal of Geophysical Research*, *114*(B1), B01401. https://doi.org/10.1029/2008jb005938

Scales, J. A., & Tenorio, L. (2001). Prior information and uncertainty in inverse problems. *Geophysics*, *66*(2), 389–397. https://doi.org/10.1190/1.1444930

Scholz, C. H. (1998). Earthquakes and friction laws. *Nature*, *391*(6662), 37–42. https://doi.org/10.1038/34097

Segall, P. (2010). *Earthquake and volcano deformation*. Princeton University Press.

Segall, P., & Du, Y. (1993). How similar were the 1934 and 1966 parkfield earthquakes? *Journal of Geophysical Research*, *98*(B3), 4527–4538. https://doi.org/10.1029/92jb02408

Segall, P., & Harris, R. (1987). Earthquake deformation cycle on the san andreas fault near parkfield, California. *Journal of Geophysical Research*, *92*(B10), 10511–10525. https://doi.org/10.1029/jb092ib10p10511

Shao, Z., Wang, R., Wu, Y., & Zhang, L. (2011). Rapid afterslip and short-term viscoelastic relaxation following the 2008 m w 7.9 wenchuan earthquake. *Earthquake Science*, *24*(2), 163–175. https://doi.org/10.1007/s11589-010-0781-z

Shaw, B. E., & Wesnousky, S. G. (2008). Slip-length scaling in large earthquakes: The role of deep-penetrating slip below the seismogenic layer. *Bulletin of the Seismological Society of America*, *98*(4), 1633–1641. https://doi.org/10.1785/0120070191

Shen, Z.-K., Jackson, D. D., Feng, Y., Cline, M., Kim, M., Fang, P., & Bock, Y. (1994). Postseismic deformation following the landers earthquake, California, 28 june 1992. *Bulletin of the Seismological Society of America*, *84*(3), 780–791. https://doi.org/10.1785/bssa0840030780

Shrivastava, M. N., González, G., Moreno, M., Chlieh, M., Salazar, P., Reddy, C., et al. (2016). Coseismic slip and afterslip of the 2015 mw 8.3 illapel (Chile) earthquake determined from continuous gps data. *Geophysical Research Letters*, *43*(20), 10–710. https://doi.org/10.1002/2016gl070684

Skarlatoudis, A., Somerville, P., & Thio, H. (2016). Source-scaling relations of interface subduction earthquakes for strong ground motion and tsunami simulation. *Bulletin of the Seismological Society of America*, *106*(4), 1652–1662. https://doi.org/10.1785/0120150320

Somerville, P., Irikura, K., Graves, R., Sawada, S., Wald, D., Abrahamson, N., et al. (1999). Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters*, *70*(1), 59–80. https://doi.org/10.1785/gssrl.70.1.59

Strasser, F. O., Arango, M., & Bommer, J. J. (2010). Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude. *Seismological Research Letters*, *81*(6), 941–950. https://doi.org/10.1785/gssrl.81.6.941

Styron, R., & Pagani, M. (2020). The gem global active faults database. *Earthquake Spectra*, *36*(1_suppl), 160–180. https://doi.org/10.1177/8755293020944182

Sun, T., & Wang, K. (2015). Viscoelastic relaxation following subduction earthquakes and its effects on afterslip determination. *Journal of Geophysical Research: Solid Earth*, *120*(2), 1329–1344. https://doi.org/10.1002/2014jb011707

Taira, T., Bürgmann, R., Nadeau, R. M., & Dreger, D. S. (2014). Variability of fault slip behavior along the san andreas fault in the san juan bautista region. *Journal of Geophysical Research: Solid Earth*, *119*(12), 8827–8844. https://doi.org/10.1002/2014jb011427

Tsang, L. L., Vergnolle, M., Twardzik, C., Sladen, A., Nocquet, J.-M., Rolandone, F., et al. (2019). Imaging rapid early afterslip of the 2016 pedernales earthquake, Ecuador. *Earth and Planetary Science Letters*, *524*, 115724. https://doi.org/10.1016/j.epsl.2019.115724

UNAVCO. (2021). *Plate motion calculator*. Retrieved from https://www.unavco.org/software/geodetic-utilities/plate-motion-calculator/plate-motion-calculator

U.S. Geological Survey. (2017). Earthquake Hazards Program, 2017, Advanced National Seismic System (ANSS) Comprehensive Catalog of Earthquake Events and Products: Various. https://doi.org/10.5066/F7MS3QZH

Van den Ende, M., Chen, J., Ampuero, J.-P., & Niemeijer, A. (2018). A comparison between rate-and-state friction and microphysical models, based on numerical simulations of fault slip. *Tectonophysics*, *733*, 273–295. https://doi.org/10.1016/j.tecto.2017.11.040

Wallace, L. M., Hreinsdóttir, S., Ellis, S., Hamling, I., D'Anastasio, E., & Denys, P. (2018). Triggered slow slip and afterslip on the southern hikurangi subduction zone following the kaikōura earthquake. *Geophysical Research Letters*, *45*(10), 4710–4718. https://doi.org/10.1002/2018gl077385

Wang, K., Hu, Y., & He, J. (2012). Deformation cycles of subduction earthquakes in a viscoelastic Earth. *Nature*, *484*(7394), 327–332. https://doi.org/10.1038/nature11032

Wang, L., Wang, R., Roth, F., Enescu, B., Hainzl, S., & Ergintav, S. (2009). Afterslip and viscoelastic relaxation following the 1999 m 7.4 i̇zmit earthquake from gps measurements. *Geophysical Journal International*, *178*(3), 1220–1237. https://doi.org/10.1111/j.1365-246x.2009.04228.x

Wang, M., Shen, Z.-K., Wang, Y.-Z., Bürgmann, R., Wang, F., Zhang, P.-Z., et al. (2021). Postseismic deformation of the 2008 wenchuan earthquake illuminates lithospheric rheological structure and dynamics of eastern Tibet. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB022399. https://doi.org/10.1029/2021jb022399

Wei, M., Kaneko, Y., Liu, Y., & McGuire, J. J. (2013). Episodic fault creep events in California controlled by shallow frictional heterogeneity. *Nature Geoscience*, *6*(7), 566–570. https://doi.org/10.1038/ngeo1835

Wei, S., Barbot, S., Graves, R., Lienkaemper, J. J., Wang, T., Hudnut, K., et al. (2015). The 2014 mw 6.1 south napa earthquake: A unilateral rupture with shallow asperity and rapid afterslip. *Seismological Research Letters*, *86*(2A), 344–354. https://doi.org/10.1785/0220140249

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, *84*(4), 974–1002.

Weng, H., & Yang, H. (2017). Seismogenic width controls aspect ratios of earthquake ruptures. *Geophysical Research Letters*, *44*(6), 2725–2732. https://doi.org/10.1002/2016gl072168

Wennerberg, L., & Sharp, R. V. (1997). Bulk-friction modeling of afterslip and the modified omori law. *Tectonophysics*, *277*(1–3), 109–136. https://doi.org/10.1016/s0040-1951(97)00081-4

Wimpenny, S., Copley, A., & Ingleby, T. (2017). Fault mechanics and post-seismic deformation at bam, se Iran. *Geophysical Journal International*, *209*(2), 1018–1035. https://doi.org/10.1093/gji/ggx065

Zhao, B., Bürgmann, R., Wang, D., Tan, K., Du, R., & Zhang, R. (2017). Dominant controls of downdip afterslip and viscous relaxation on the postseismic displacements following the mw7. 9 gorkha, Nepal, earthquake. *Journal of Geophysical Research: Solid Earth*, *122*(10), 8376–8401. https://doi.org/10.1002/2017jb014366

Zhao, D., Qu, C., Bürgmann, R., Gong, W., & Shan, X. (2021). Relaxation of Tibetan lower crust and afterslip driven by the 2001 mw7. 8 kokoxili, China, earthquake constrained by a decade of geodetic measurements. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB021314. https://doi.org/10.1029/2020jb021314

Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-022-02221-2

ORIGINAL PAPER

Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence

Christian Grimm[1] · Sebastian Hainzl[2] · Martin Käser[3,4] · Helmut Küchenhoff[1]

Accepted: 21 March 2022

Abstract

Strong earthquakes cause aftershock sequences that are clustered in time according to a power decay law, and in space along their extended rupture, shaping a typically elongate pattern of aftershock locations. A widely used approach to model earthquake clustering, the *Epidemic Type Aftershock Sequence* (ETAS) model, shows three major biases. First, the conventional ETAS approach assumes isotropic spatial triggering, which stands in conflict with observations and geophysical arguments for strong earthquakes. Second, the spatial kernel has unlimited extent, allowing smaller events to exert disproportionate trigger potential over an unrealistically large area. Third, the ETAS model assumes complete event records and neglects inevitable short-term aftershock incompleteness as a consequence of overlapping coda waves. These three aspects can substantially bias the parameter estimation and lead to underestimated cluster sizes. In this article, we combine the approach of Grimm et al. (Bulletin of the Seismological Society of America, 2021), who introduced a generalized anisotropic and locally restricted spatial kernel, with the *ETAS-Incomplete* (ETASI) time model of Hainzl (Bulletin of the Seismological Society of America, 2021), to define an ETASI space-time model with flexible spatial kernel that solves the abovementioned shortcomings. We apply different model versions to a triad of forecasting experiments of the 2019 Ridgecrest sequence, and evaluate the prediction quality with respect to cluster size, largest aftershock magnitude and spatial distribution. The new model provides the potential of more realistic simulations of on-going aftershock activity, e.g. allowing better predictions of the probability and location of a strong, damaging aftershock, which might be beneficial for short term risk assessment and disaster response.

Keywords ETAS · Short-term incompleteness · Anisotropic spatial kernel · Ridgecrest

1 Introduction

Strong earthquakes are usually observed to cause a pronounced spatio-temporal pattern of aftershocks. More precisely, according to the Omori-Utsu Law (Utsu et al. 1995), the temporal aftershock rate is subject to a power law decrease with time $t - t_{main}$ after the main triggering event, that is,

$$g(t - t_{main}) = (t - t_{main} + c)^{-p} \quad (1)$$

✉ Christian Grimm
christian.grimm@stat.uni-muenchen.de

Sebastian Hainzl
hainzl@gfz-potsdam.de

Martin Käser
martin.kaeser@geophysik.uni-muenchen.de

Helmut Küchenhoff
Kuechenhoff@stat.uni-muenchen.de

[1] Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstraße 33, 80539 Munich, Germany

[2] GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany

[3] Department of Earth and Environmental Sciences, Geophysics, Ludwig-Maximilians-University Munich, Theresienstraße 41, 80333 Munich, Germany

[4] Present Address: Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany

Published online: 16 April 2022

with the delay parameter $c > 0$ (usually a few minutes to hours) and the exponent p (usually in the range between $0.8 - 1.2$). It means that the temporal pattern of aftershocks is dominated by events occurring within short time after the mainshock. Figure 1a demonstrates this temporal behavior for the Ridgecrest sequence in California, which produced an M6.4 foreshock on July 4, 2019, followed by an M7.1 mainshock within 34 hours on July 6, 2019.

The observed spatial patterns of aftershock sequences stem from their tendency to occur on or close to the mainshock rupture plane (Marsan and Lengliné 2008). The larger the length-to-width ratio of this plane gets, the more elongate the typical aftershock region becomes. In addition, a higher dip angle reduces the width of the 3D-to-2D projection of the rupture plain to the earth's surface and therefore results in a scatter of two-dimensional aftershock epicenters that can be increasingly well approximated by a line segment.

The prevailing continental tectonic regime in southern California with typically steep, strike-slip faulting favors such elongated aftershock patterns in this region. With the exception of the M6.7 1994 Northridge earthquake, all of the most prominent mainshock-aftershock sequences of the last 40 years (M6.6 1987 Superstition Hill, M7.3 1992 Landers, M7.1 1999 Hector Mine, M7.2 2010 Baja California, M7.1 2019 Ridgecrest) demonstrate distinct linearly elongate scattering of aftershock locations (Hainzl 2021).

In this context, the Ridgecrest sequence is a special case as the M6.4 foreshock simultaneously ruptured two almost orthogonal faults, leading to a double pattern of separate linearly elongate aftershock clouds (Marsan and Ross 2021). Fig. 1b shows that the triggering M6.4 event (yellow pentagram) is located close to the intersection of the two ruptured faults. In contrast, the M7.1 mainshock (yellow hexagram) ruptured only one fault which appears to be the extension of one of the faults activated by the foreshock.

Analyzing and forecasting clustered seismicity is an established discipline in seismological research. Its goal is to understand the evolution of large aftershock sequences and to predict their size, largest aftershock magnitude, spatial distribution etc. A prominent approach to model clustered seismicity is the so-called *Epidemic Type Aftershock Sequence (ETAS)* model, which describes earthquake records as a superposition of independent background seismicity and triggered earthquake sequences (Ogata 1988, 1998). The earthquake triggering component is designed in terms of a branching process and characterized by the triad of (1) trigger-magnitude dependent aftershock productivity, (2) a temporal distribution of aftershock times typically derived from the Omori Law (see Eq. 1), and (3) an usually isotropic spatial distribution of aftershock locations (e.g. Zhuang et al. 2002; Jalilian 2019). Particularly, the aftershock productivity (i.e. expected number of offsprings) for a trigger event with magnitude m is

$$k_{A,\alpha}(m) = A\ \exp(\alpha\,(m - M_c)), \tag{2}$$

where parameters $A > 0$ and $\alpha > 0$ control the exponential growth of the trigger potential and M_c is the cut-off magnitude of the analyzed earthquake catalog.

Despite generally producing successful and insightful estimation and forecast results, ETAS models are known to be limited by a number of potential biases. In this article, we present an approach that combines solutions for three main short-comings of the conventional ETAS model, (1) the isotropic spatial aftershock distribution, (2) the infinite extent of the spatial kernel and (3) the short-term incompleteness of earthquake records after strong triggering events.

1.1 Bias 1: isotropic spatial distribution

The common assumption in ETAS models is that spatial aftershock locations are distributed isotropically around the triggering event. It is named as a shortcoming in many publications because it stands in conflict with the above-mentioned observation that aftershocks tend to occur close to the (elongate) rupture plane of the triggering event (Ogata 1998, 2011; Ogata and Zhuang 2006; Hainzl et al. 2008, 2013; Seif et al. 2017; Zhang et al. 2018). The assumption of isotropy is reasonably valid for weak earthquakes with small rupture extensions, but becomes problematic for larger magnitudes, e.g. see the spatial pattern of the Ridgecrest sequence in Fig. 1b. It has been shown that inadequate spatial models can lead to an underestimation of the productivity parameter α (Eq. 2) because the numerous small events take over the role of mimicking the "true" anisotropic distribution (Hainzl et al. 2008, 2013; Grimm et al. 2021).

1.2 Bias 2: infinite spatial extent

Under the common definition of an inifinite spatial kernel, aftershock triggering is disproportionately associated with the more numerous small events, that can more flexibly mimic anisotropic event alignments than the few strong mainshocks. This can lead to unrealistically far trigger impact of small magnitudes and to a substantial underestimation of the direct aftershock productivity of strong events, resulting in a smoothing of temporal event distributions (Grimm et al. 2021).

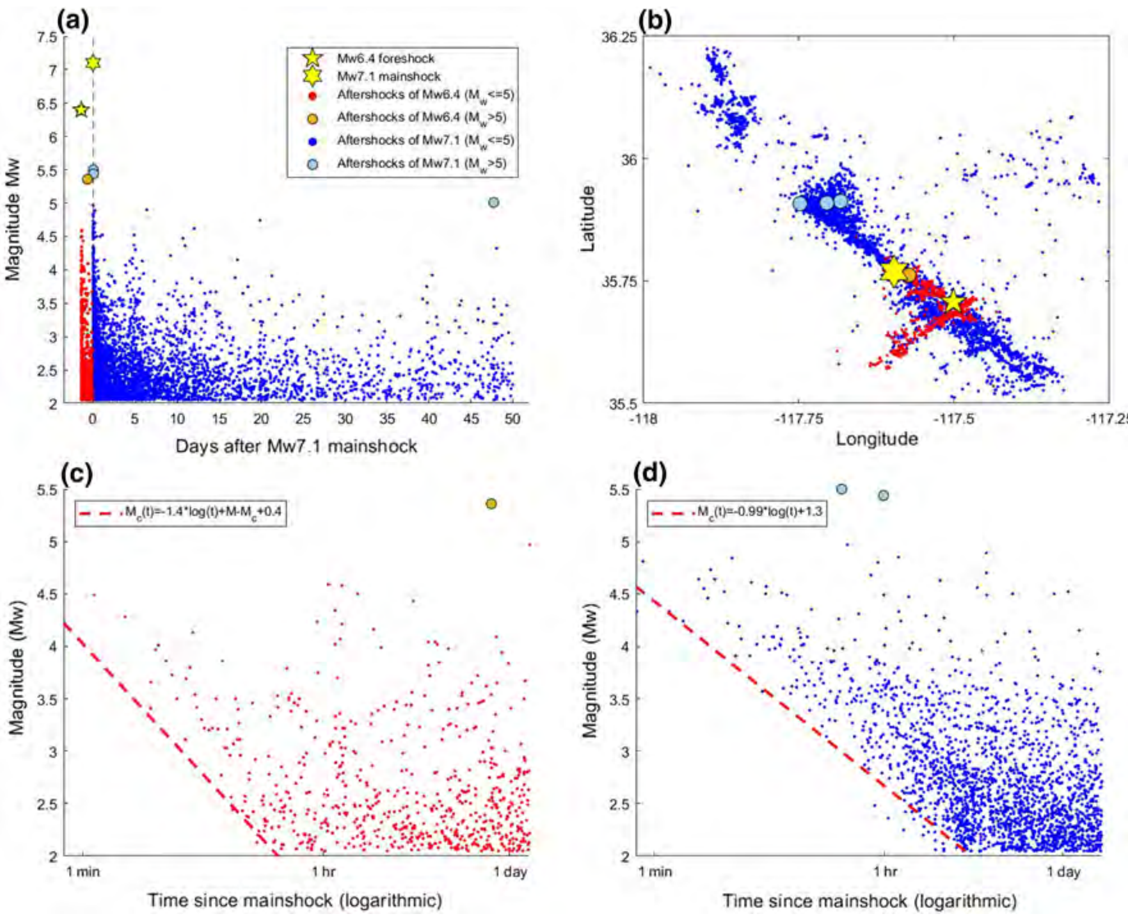


Fig. 1 **a** Magnitudes versus event times of Ridgecrest Mw6.4 (red dots) and Mw7.1 (blue dots) aftershock sequences. Event times are denoted in days before/after Mw7.1 mainshock, the dashed black line represents the time origin (M7.1 event time). Light blue and light red dots mark aftershocks with magnitudes larger than 5. Yellow pentagram symbolizes the Mw6.4 foreshock, and yellow hexagram marks the Mw7.1 mainshock. **b** Aftershock locations of the Ridgecrest Mw6.4 and Mw7.1 sequences. Legend as in **a**. **c** Magnitudes versus logarithmic event times of Ridgecrest Mw6.4 sequence. The dashed red line represents a manually fitted estimate of the empirical completeness function $M_c(t)$. **d** Magnitudes versus logarithmic event times of Ridgecrest Mw7.1 sequence. The dashed red line represents a manually fitted estimate of the empirical completeness function $M_c(t)$

1.3 Bias 3: short-term aftershock incompleteness (STAI)

Strong earthquakes typically cause incomplete aftershock records immediately after their occurrence, mainly due to an overlap of coda waves (Hainzl 2016a; de Arcangelis et al. 2018). Figure 1c and (d) confirms this phenomenon for the aftershock sequences of the M6.4 and M7.1 Ridgecrest events, respectively. Apparently, records of smaller sized aftershocks are missing in the first minutes to hours, somewhat foiling the power law decay of event rates expected from the Omori-Utsu law (Eq. 1). The short-term incomplete event records can therefore hide to a large extent both the "true" Omori Law decay (Eq. 1) and the "true" aftershock productivity of the trigger event (Eq. 2) and lead to an overestimation of Omori parameter c and an underestimation of productivity parameter α (Hainzl 2021, 2016b; Page et al. 2016; Seif et al. 2017).

Data-driven uncertainties of event locations and cut-off magnitude as well as the assumption of a *time-invariant seismic background* may lead to further inaccuracies in the parameter estimation (Harte 2013, 2016; Seif et al. 2017). However, they can be neglected in our study because they are either expected to be small in southern California datasets (e.g. location and magnitude uncertainty) or do not apply in an isolated sequence analysis (background miss-specification).

1.4 Scope of this article

In this article, we combine an ETAS approach accounting for short-term incomplete event records with the application of a generalized, anisotropic spatial model that restricts the spatial kernel to the local surrounding of the trigger source. We demonstrate the functionality and superiority of our approaches over the conventional, isotropic ETAS model by means of forecasting experiments for the Ridgecrest sequence.

We utilize the generalized anisotropic and locally restricted spatial kernel suggested by Grimm et al. (2021), which assumes uniform trigger density along an estimated rupture line segment, with power-law decay to the sides and at the end points of the rupture. Zhang et al. (2018) pursued an even more detailed approach, which assumed constant trigger rate in the entire rupture plane, with power-law decay outside of it. Different versions of elliptic Gaussian distributions were introduced and discussed by Ogata (1998, 2011) and Ogata and Zhuang (2006). The latter approaches successfully modeled spatial aftershock patterns, however, they require a new set of parameters and are therefore not flexibly combinable with the conventional, isotropic functionality. In contrast, the kernel of Grimm et al. (2021) represents a generalization of the isotropic function and therefore allows simultaneous anisotropic modeling of some events (e.g. above a certain magnitude threshold) and isotropic modeling of the rest. In order to address the abovementioned particularity of the M6.4 Ridgecrest foreshock, rupturing two almost orthogonal faults, we further generalize the approach by allowing a spatial kernel composed by a weighted superposition of two distinct rupture line segments.

Additionally, we accounts for STAI by applying an ETAS model version that incorporates rate-dependent incompleteness of event records. Recognizing alternative approaches that will be briefly described in the *Methods* section, we choose for the *ETAS-Incomplete (ETASI)* model as recently suggested by Hainzl (2021). For simplicity and to sharpen its focus on the incompleteness detection, Hainzl (2021) neglected the space dimension in his model. As this article combines the ETASI time model of Hainzl (2021) with an adequate, anisotropic spatial kernel it can be seen as the space-including extension of the latter. The focus of this study, however, is on the benefit of modeling the spatial aftershock distribution by a generalized anisotropic spatial kernel, rather than the benefit of the ETASI model.

This article is structured as follows. In the *Methods* section, we introduce the conventional ETAS model and its ETASI extension and define the anisotropic, locally restricted spatial kernel. This section includes a description of the estimation procedures for strikes and rupture positions and the spatial integral over anisotropic kernels. Next, the *Application* section explains the three forecasting experiments, introducing the data and time-space windows for the parameter estimation and forward simulations. Finally, we interpret and discuss our forecasting results and draw our conclusions. Source codes for model estimation and simulation are freely available in a Github repository (see Data and resources).

2 Methods

The ETAS model, first introduced by Ogata (1988, 1998), is a branching-tree type model which describes clustered earthquake occurrences by consecutive triggering evolving over multiple parent-child generations (i.e. allowing secondary aftershocks). The triggered seismicity is overlaying a time-invariant background process.

In this section, we will first introduce the conventional, isotropic ETAS model approach. Next, we will extend the model to obtain a time-space version of the ETASI model suggested by Hainzl (2021), which involves STAI into parameter estimation. Mostly, notations are consistent with Hainzl (2021). We will then define the anisotropic generalization of the spatial kernel, which is compatible with both the ETAS and ETASI model, and introduce the local restriction of the kernel. Finally, we explain the fitting algorithm for the strike angle and rupture position of anisotropic trigger events and the methods for spatial integral estimation.

2.1 ETAS-model

In the conventional ETAS model approach, the occurrence rate of an earthquake with *magnitude* m, occurring at *time* t and at *location* (x, y) is modeled by an inhomogeneous Poisson process with a time-space-magnitude dependent intensity function

$$\lambda(t, x, y, m) = f_0(m)\, R_0(t, x, y)$$

where

$$f_0(m) = \beta\, e^{-\beta(m-M_c)} \tag{3}$$

is the "true" probability density function (pdf) of the frequency-magnitude distribution (FMD) with Gutenberg-Richter parameter $b = \beta/ln(10)$ (Gutenberg and Richter 1944), and

$$R_0(t, x, y) = \mu\, u(x, y) + \sum_{i:t_i<t} k_{A,\alpha}(m_i)\, g_{c,p}(t-t_i)\, h_{D,\gamma,q}(r_i(x, y), m_i, l_i) \tag{4}$$

is the "true" occurrence rate of events with magnitude $m \geq M_c$, at time t and at location (x, y). The "true" event rate is modeled by a superposition of the time-invariant *seismic background rate* $\mu\, u(x, y)$ with parameter $\mu > 0$ and a sum of the trigger rate contributions of all events i that occurred prior to current time t. $k_{A,\alpha}(m_i)$ and $g_{c,p}(t-t_i)$ denote the aftershock productivity and Omori-Utsu Law decay functions as defined in Eqs. (1) and (2), respectively. $h_{D,\gamma,q}(r_i(x, y), m_i, l_i)$ models distribution of aftershock locations triggered by event i, with parameters D, γ and

q. The precise inputs and shape of the spatial kernel are discussed later.

The term "true" means that the (physical) relationships are expected to be observed with perfect earthquake records. The main assumption of the conventional ETAS model is that STAI does not significantly distort the "true" magnitude distribution and the "true" event rates.

2.2 ETASI model

2.2.1 Rate-dependent iIncompleteness

The concept of rate-dependent earthquake record incompleteness assumes that the "true" relationships underlying $f_0(m)$ and $R_0(t,x,y)$ are not accurately identifiable in available earthquake catalogs because especially events with small magnitudes are detected with lower probability in periods of high seismic activity. In these periods, the detection ability is limited typically due to overlapping seismic waves (Hainzl 2016a, 2021).

Fitting the "true" relationships to incomplete data records may therefore lead to significantly biased parameter estimates (Hainzl 2016a, b; Page et al. 2016; Seif et al. 2017; Hainzl 2021).

In recent years, there has been growing research interest in how to account for short-term incomplete datasets. For instance, Zhuang et al. (2017) developed a replenishment algorithm to fill up likely incomplete time intervals by simulated events, in order to obtain artificially complete pseudo-records. Other authors, particularly mentionable Omi et al. (2013, 2014), Lippiello et al. (2016), de Arcangelis et al. (2018), Mizrahi et al. (2021) and Hainzl (2021), tried to incorporate STAI directly into the ETAS model fit. A rather simple workaround approach is to remove likely incomplete time periods from the fitted time interval using empirical completeness functions, such as performed in Hainzl et al. (2013) and Grimm et al. (2021). A comprehensive discussion and comparison of various ETASI models is not in the scope of this article. The choice for the ETASI model proposed by Hainzl (2021) was made for rather practical reasons, mainly because of its compatibility with existing code.

2.2.2 Model formulation

The working assumption of the ETASI model described here is that an earthquake with magnitude m and occurring at time t can only be detected by the operating seismic network if no event of equal or larger magnitude occurred within the blind time $[t-T_b, t]$, where T_b is typically in the range of some seconds to few minutes (Hainzl 2021). Similar assumptions have formerly been formulated by Lippiello et al. (2016), de Arcangelis et al. (2018) and Hainzl (2016a).

Let $N_0(t)$ be the expected number of events occurring within the entire spatial window S during blind time $[t-T_b, t]$,

$$N_0(t) = \int_{t-T_b}^{t} \iint_S R_0(t,x,y) dx\, dy\, dt \approx T_b \iint_S R_0(t,x,y)\, dx\, dy,$$

where the approximation holds under the assumption that event rates are approximately constant during the blind time (Hainzl 2021). According to the "true" FMD (Eq. 3), each of the $N_0(t)$ events has a probability of $e^{-\beta(m-M_c)}$ to exceed magnitude m. We define the detection probability $p_d(m,t)$ of an earthquake at time t with magnitude m as the probability that no equal or larger event occurred during blind time T_b, i.e.

$$p_d(m,t) = e^{-N_0(t)\, e^{-\beta(m-M_c)}}.$$

Following the derivations in Hainzl (2016b, 2021), we obtain the "apparent", incompleteness-biased FMD

$$f(m,t) := f_0(m)\, N_0(t)\, \frac{p_d(m,t)}{1-e^{-N_0(t)}}$$

and the "apparent" event rate

$$R(t,x,y) := \frac{R_0(t,x,y)}{N_0(t)} \left(1 - e^{-N_0(t)}\right).$$

The term "apparent" signalizes that the functions f and R do not represent the "true", but the observable relationships that are possibly distorted by rate-dependent record incompleteness. In periods of high seismic activity, the "apparent" FMD exhibits a larger relative frequency of strong events (because they are more likely to be detected) and an event rate lowered by detection capacity. We obtain the ETASI intensity function

$$\begin{aligned}\lambda(t,x,y,m) &= f(m,t)\, R(t,x,y)\\ &= f_0(m)\, R_0(t,x,y)\, p_d(m,t)\end{aligned}$$

The two underlying, simplifying assumptions in the ETASI model are that (1) the blind time T_b is magnitude-independent, which Hainzl (2021) justifies by typically shorter source durations than travel times of coda waves, and (2) that the seismic network is equally occupied for blind time T_b by any event in the entire investigated spatial window. The second assumption is reasonable for a small spatial window, e.g. when analyzing an isolated sequence. When fitting the ETASI model over a larger region, it needs to be checked that relevant clusters do not evolve at the same time but at distinct locations as they would be assumed to simultaneously occupy the entire seismic network. A reasonable approach to prevent undesired biases is to choose a larger cut-off magnitude.

2.2.3 Log-likelihood optimization

The parameter vector $\theta = \{\mu, A, \alpha, c, p, D, \gamma, q, \beta, T_b\}$ of the ETASI model is estimated by maximizing its log-likelihood function $LL = LL_1 - LL_2$ with

$$LL_1 = \sum_{events\ j} ln\big(f_0(m_j)\, R_0(t_j, x_j, y_j)\, p_d(m_j, t)\,\big),$$

$$\begin{aligned} LL_2 &= \int_{M_c}^{\infty} \int_{T_1}^{T_2} \iint_S \lambda(t, x, y, m)\, dx\, dy\, dt\, dm \\ &\approx \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-T_b \iint_S R_0(t,x,y)\, dx\, dy} dt \end{aligned} \tag{5}$$

where the sum in LL_1 goes over all *target events* in the magnitude-time-space window $[M_c, \infty) \times [T_1, T_2] \times S$ and LL_2 integrates over this model space. In our study we optimized the parameter vector θ using the gradient-based Davidson-Fletcher-Powell algorithm (Ogata 1998; Zhuang et al. 2002; Jalilian 2019).

2.3 Generalized anisotropic spatial kernel

2.3.1 Conventional isotropic kernel

The *spatial kernel* $h_{D,\gamma,q}(r_i, m_i, l_i)$ in Eq. (4) models the 2D-distribution of aftershocks locations. In conventional ETAS model approaches, the triggering event is assumed to be a point source, distributing its offsprings isotropically around its epicenter. A classical definition of an isotropic kernel (see Ogata 1998; Grimm et al. 2021; Jalilian 2019) is

$$h^{iso}_{D,\gamma,q}(r_i(x,y), m_i) := \frac{q-1}{D\ \exp(\gamma(m_i - M_c))} \left(1 + \frac{\pi\, r_i(x,y)^2}{D\ \exp(\gamma(m_i - M_c))}\right)^{-q}$$

where $r_i(x, y)$ denotes the point-to-point distance between a potential aftershock location (x, y) and the coordinates (x_i, y_i) of the triggering event i, and m_i is the magnitude of the event i. The kernel is constrained by the parameters D and γ that control the magnitude-dependent width of the kernel, and parameter q that describes the exponential decay of the function with growing spatial distance.

2.3.2 Anisotropic generalization

Here we use the anisotropic generalization of the spatial kernel that was first introduced by Grimm et al. (2021),

$$h_{D,\gamma,q}(r_i(x,y), m_i, l_i) := \frac{q-1}{D\ \exp(\gamma(m_i - M_c))} \left(1 + \frac{2\, l_i\, r_i(x,y)\ +\ \pi\, r_i(x,y)^2}{D\ \exp(\gamma(m_i - M_c))}\right)^{-q}.$$

In this spatial model, the distance term $r_i(x, y)$ denotes the point-to-line distance between the potential aftershock location (x, y) and the estimated rupture segment of triggering event i with length l_i. That is, the kernel assigns constant density along the rupture line segment, with a power-law decay to the sides. Note that

$$h_{D,\gamma,q}(r_i(x,y), m_i, 0) = h^{iso}_{D,\gamma,q}(r_i(x,y), m_i),$$

i.e. the anisotropic kernel is a generalization and collapses to the isotropic model if the triggering location is assumed to be a point source with rupture extension $l_i = 0$. Therefore, the generalized spatial model can be used for mixing approaches of both kernels, e.g. applying anisotropy to events i with magnitudes $m_i \geq M_{aniso}$:

$$l_i = \begin{cases} 0, & \text{for } m_i < M_{aniso}, \quad \text{(isotropic trigger)} \\ 10^{-2.57+0.62 m_i}, & \text{for } m_i \geq M_{aniso}, \quad \text{(anisotropic trigger)} \end{cases} \tag{6}$$

The scaling relationship for anisotropic events is taken from the estimate of subsurface rupture lengths for strike-slip faulting events, provided in Wells and Coppersmith (1994). Alternative relationships can be applied, too, but are not expected to impact results.

2.3.3 Local spatial restriction

Both the conventional isotropic and the generalized anisotropic kernels are defined in terms of a probability density function (pdf) over infinite space. Realistically, however, small earthquakes should exert only a locally restricted trigger influence. Grimm et al. (2021) showed that an infinite spatial extent may lead to an underestimation of the aftershock productivity parameter α because it overestimates the triggering power of smaller events at the cost of the larger events. A manual analysis of the spatial aftershock patterns of the six Californian mainshocks named in the introduction shows that the cloud of potential aftershocks typically lies within one estimated rupture length (by Wells and Coppersmith 1994) from the epicenter. In case of an anisotropic shape of the kernel, the area of half a rupture length around the extended rupture segment seems sufficient. According to this observation, we choose a spatial restriction R_i for event i according to

$$R_i := \begin{cases} 10^{-2.57+0.62 m_i}, & \text{for } m_i < M_{aniso}, \quad \text{(isotropic trigger)} \\ 0.5 \cdot 10^{-2.57+0.62 m_i}, & \text{for } m_i \geq M_{aniso}, \quad \text{(anisotropic trigger)} \end{cases} \tag{7}$$

where again we use the strike-slip faulting subsurface rupture length scaling from Wells and Coppersmith (1994).

In other words, the spatial kernel for event i is only defined in the restricted area

$$S_i(R_i) := \{(x, y) \in \mathbb{R}^2 | r_i(x, y) \leq R_i\}$$

and set to 0 outside of it. Note that the restricted area $S_i(R_i)$ can assume isotropic and anisotropic shapes, depending on the point-to-point or point-to-line definition of the distance term $r_i(x, y)$. In order to retain the property of a pdf, we need to rescale the kernel within the restricted area by its analytical integral

$$\begin{aligned} H_{D,\gamma,q}(R_i, m_i, l_i) : &= \iint_{S_i(R_i)} h_{D,\gamma,q}(r_i(x, y), m_i, l_i)\, dx\, dy \\ &= 1 - \left(1 + \frac{2\, l_i\, R_i \,+\, \pi\, R_i^2}{D\, \exp(\gamma(m_i - M_c))}\right)^{1-q}. \end{aligned} \tag{8}$$

The integral term holds true for both isotropic ($l_i = 0$) and anisotropic triggers ($l_i > 0$). We obtain the generalized, restricted and anisotropic spatial kernel

$$h^{restr}_{D,\gamma,q}(r_i(x, y), m_i, l_i) = \begin{cases} \dfrac{h^{restr}_{D,\gamma,q}(r_i(x, y), m_i, l_i)}{H_{D,\gamma,q}(R_i, m_i, l_i)}, & \text{if } r_i(x, y) \leq R_i, \\ 0, & \text{if } r_i(x, y) > R_i. \end{cases} \tag{9}$$

2.4 Estimation of strike and epicenter location

The restricted, anisotropic spatial kernel in Eq. (9) requires a strike angle to define the orientation of the extended rupture for anisotropic triggers with $l_i > 0$. In retrospect, the strike angle could be taken from one of the numerous publications about the Ridgecrest sequence or from focal mechanism datasets such as the Global Moment Tensor Catalog, the ISC-GEM Global Instrumental Earthquake Catalog or from local datasets of the Southern California Earthquake Data Center (SCEDC). In order to perform a realistic forecasting test case, however, we should build upon instantaneous strike estimates such as from local agencies (e.g. the United States Geological Survey), which are typically available within several minutes to hours.

Here, we used the strike estimation algorithm proposed by Grimm et al. (2021), that optimally fits the rupture segment through the cloud of early aftershock locations by maximizing the contributed spatial rate under initial spatial parameter guesses. To be more precise, we ran through possible strikes in 1° steps (i.e. $\{1°, \ldots, 180°\}$ where we can neglect all strikes larger than 180° because we do not account for dip direction in our model). We also moved the rupture along each strike angle in order to test different positions of the rupture segment relative to the fix epicenter. Here, we go through possible relative positions in 0.01-steps (i.e. $\{0, 0.01, 0.02, \ldots, 0.99, 1\}$), where 0 and 1 means that one of the ends of the rupture segment coincides with the epicenter, and 0.5 denotes the situation where the rupture embeds the epicenter directly in its center. Among all combinations, we searched the orientation and rupture position that maximizes the forward trigger contribution of the anisotropic event i to subsequent events j within a fixed duration $\Delta t = 1$ *hour*, i.e. with $t_i < t_j < t_i + \Delta t$. The forward trigger contribution of event i is computed as

$$\sum_{t: t_i < t_j < t_i + \Delta t} h^{rest}_{D,\gamma,q}(r_i(x_j, y_j), m_i, l_i). \tag{10}$$

In order to avoid that the rupture orientation and position is dominated by single events that occurred very close to the segment candidate, we applied a minimum distance of 0.2 kilometers.

Here, we use the initial spatial parameters $D = 0.0025$, $\gamma = 1.78$ and $q = 1.71$ derived from the results of an isotropic ETAS model for a long-term California dataset, locally restricted to $R = 2.5$ rupture lengths, by Grimm et al. (2021). Tests have shown that modified initial parameters changed the level of the sum of forward rate contributions, but led to similar strike and epicenter location estimates. We also tested multiple durations Δt up to 30 hours and found that the estimation procedure provided very similar estimates for strike and rupture position. It shows that the rupture orientation and position can be appropriately identified soon after the trigger event occurred.

In the *Application* section we present the strike and rupture position estimation for the example of the M6.4 and M7.1 Ridgecrest events.

2.5 Estimation of spatial integral

The computation of the log-likelihood function in Eq. (5) requires the estimation of the spatial integral of R_0 and therefore $h^{restr}_{D,\gamma,q}$.

In the isotropic case, the integral can be estimated semi-analytically by the radial partitioning method suggested by Ogata (1998) and applied in the *R* package *ETAS* (Jalilian 2019). It uses the property, that the isotropic spatial kernel can be integrated analytically over circular areas $S_i(R)$, according to Eq. (8). As Fig. 2a illustrates, the polygon defining the spatial window S is partitioned into a fine grid, with two neighboring boundary grid points having approximately equal distances $\tilde{R}$ to the point source coordinate of event i. The integral of $h^{restr}_{D,\gamma,q}$ over each of these thin segments of a circle can then be approximated by the analytical full integral, weighted by the ratio of the circle segment $\phi/360°$, where ϕ is the angle enclosed by the circle segment (Jalilian 2019; Ogata 1998).

Similarly, an anisotropic spatial kernel can be integrated analytically over an anisotropic region $S_i(\tilde{R})$ with maximum distance $\tilde{R}$ to the extended rupture. Due to the non-circular shape of region $S_i(R)$ for anisotropic triggers, radial partitioning can be only performed at both ends of the rupture segment. As Fig. 2b illustrates, in a similar way we use "bin partitioning" in the space orthogonal to the rupture. Unfortunately, in the anisotropic case, the weights $\phi/360°$ of the circle segments at both ends of the rupture only relate to the part of the integral that is estimated by radial partitioning. Similarly, the weight of a bin of size Δl is $\frac{\Delta l}{2 l_i}$ relative to only the orthogonal space on both sides of the rupture segment. In each iteration of the parameter estimation, the shares of the radial and the orthogonal integral parts change and need to be determined numerically before each iteration. This comes at the computational cost of approximately doubled run-time, given that only the minority of strong earthquakes with magnitude $M \geq M_{aniso}$ are modelled anisotropically.

3 Application

We carry out three forecasting experiments to check the quality of the previously defined models in predicting the observed Ridgecrest M6.4 and M7.1 sequences. Each forecasting experiment consists of three main steps, represented as blue boxes in Fig. 3:

- **Parameter Estimation:** Estimate model parameters for a specified event sub-set of southern Californian earthquake data
- **Forward Simulation:** Use the fitted model parameters to conduct 10,000 forward simulations of the Ridgecrest M6.4 or M7.1 sequence, respectively.
- **Evaluation:** Analyze simulated sequences and compare to the observation.

In the following, we first introduce the basic earthquake event set for California underlying this study, and define the time-space windows used to obtain the event sub-sets applied for parameter estimation. Next, we describe the three forecasting experiments, rigorously defining the magnitude-time-space windows applied for parameter estimation and forward simulations. Each forecasting experiment is repeated for five versions of the models introduced in the *Methods* section, which are defined in detail. Finally, we specify the forward simulation process and attributes and measures to assess the quality of the forecasts. Here, we also give an example of the estimation of strike angles and rupture positions for the Ridgecrest M6.4 and M7.1 events.

3.1 Data

As our basic event set, we downloaded the earthquake catalog from the Southern California Earthquake Data Center (SCEDC, Hauksson et al. 2012). The entire dataset comprises 699,175 event occurrences between January 1, 1981, and December 31, 2019. Because magnitudes seem to be clustered at values with one decimal, we round all

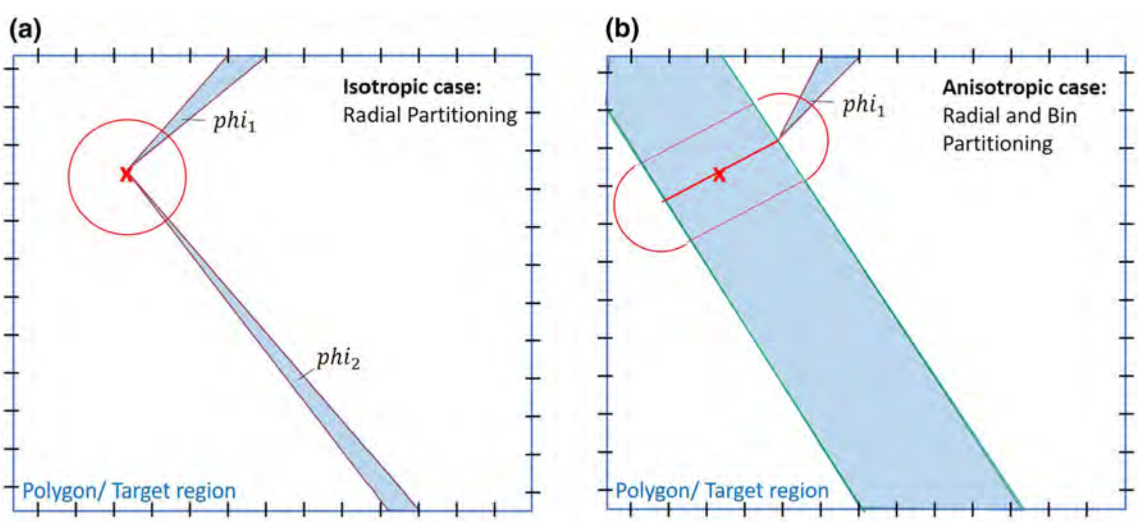


Fig. 2 Visualization of the spatial integral estimation needed for computing the log-likelihood function (Eq. 5) for **a** isotropic triggers and **b** anisotropic triggers. The box represents the boundary of the spatial target region (polygon), gridded into small intervals. Red crosses symbolize the epicenters of the triggering events. In **a**, the red circle around the event represents the contour lines of an isotropic spatial kernel and the shaded segments illustrate the radial partitioning method. In (b), the red box and semi-circles symbolize the contour lines of the anisotropic spatial kernel, and the shaded segments illustrate the radial and bin partitioning method

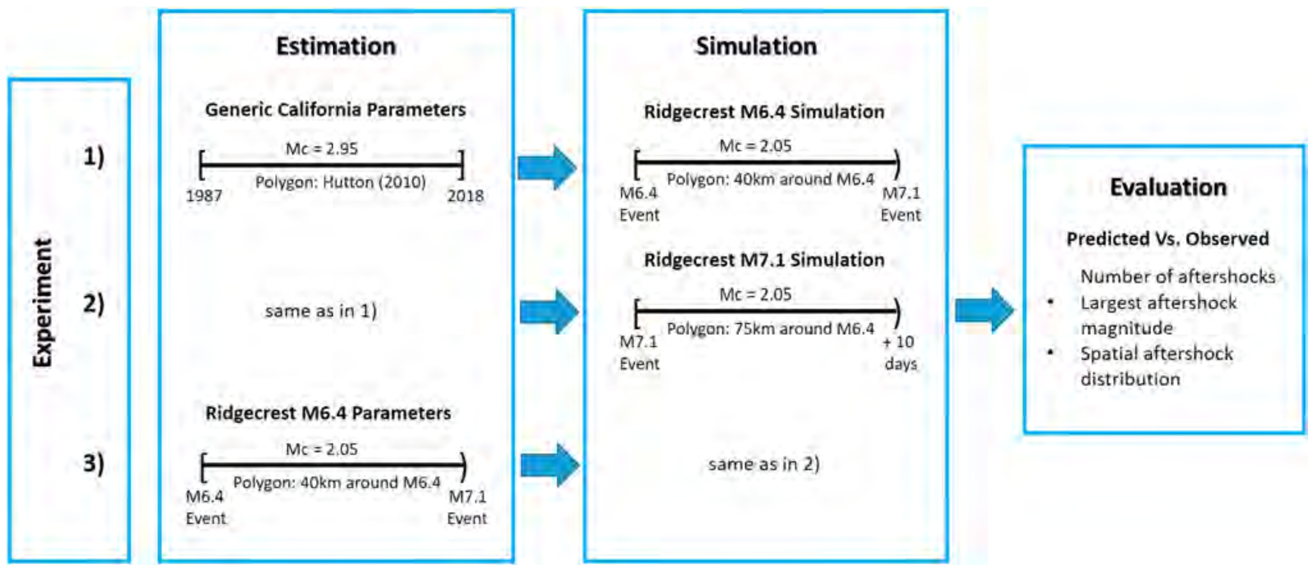


Fig. 3 Summary of the forecasting experiments (from left to right): The five model versions, listed in Table 1, are fitted to a long-term California event sub-set (*Experiments 1* and *2*) and to the local M6.4 Ridgecrest sequence (*Experiment 3*). The estimated parameters are applied to forward simulations of the Ridgecrest M6.4 sequence (*Experiment 1*) and the Ridgecrest M7.1 sequence (*Experiments 2* and *3*). The predicted sequences are compared to the observed ones with respect to three attributes, further described in the *Attributes and Measures* section

magnitudes to one decimal and use the cut-off magnitude $M_c = 2.05$ (Hutton et al. 2010; Hainzl 2021). We remove events at depths larger than 40 km to avoid completeness issues.

3.2 Forecasting experiments

Here, we describe in detail the design of the forecasting experiments, summarized in Fig. 3.

3.2.1 Experiment 1

We estimate generic, long-term California model parameters within the hexagonal polygon of southern California defined in Hutton et al. (2010). In order to mitigate computational costs, we restrict the time window to the period between January 1, 1987, and December 31, 2018, including the five prominent earthquake sequences before Ridgecrest as named in the *Introduction* section, and choose the larger cut-off magnitude $M_c = 2.95$. The cut-off magnitude is a trade-off between too large and too small event record sizes that ensures reasonable run-time and statistical robustness of parameter estimates. Additionally, it avoids potentially biased estimates of the blind time parameter T_b in large spatial areas due to simultaneous clustering. The magnitude-time-space window contains 7,215 fitted *target* events. We account for external triggering impact by including *complementary* events that occurred after January 1, 1986, and in the surrounding of 0.5° geographical degrees of the fitted area.

The estimated models are then applied to forecast the Ridgecrest M6.4 foreshock sequence above cut-off

Table 1 Overview of the model variants tested in this paper

Name	Model version	M_{aniso}	R_i (isotropic triggers)	R_i (anisotropic triggers)
ETAS conventional	ETAS	-	∞	-
ETAS iso-r	ETAS	-	1 RL_i	-
ETAS aniso-r	ETAS	6.0	1 RL_i	0.5 RL_i
ETASI iso-r	ETASI	-	1 RL_i	-
ETASI aniso-r	ETASI	6.0	1 RL_i	0.5 RL_i

Non applicable cases are filled with "-". Spatial restrictions R_i of event i are denoted in terms of the estimate rupture length (RL_i)

magnitude $M_c = 2.05$, within a circular polygon with radius 40 km centered in the M6.4 event location. The simulated time window starts in the moment of the M6.4 event (July 4, 2019) and ends at the M7.1 mainshock event time (July 6, 2019), thus it has a duration of approximately 34 hours. We initialize triggering seismicity by the event history from June 1, 2019.

3.2.2 Experiment 2

In the second experiment, we use the same set of generic, long-term California parameters, but apply it in a forecast of the Ridgecrest M7.1 mainshock sequence above cut-off magnitude $M_c = 2.95$, starting at the M7.1 event time for a duration of ten days. The spatial simulation window is defined by a disk with radius of 75 km, centered in the M7.1 event location. Again, we initialize triggering seismicity by the event history from June 1, 2019, here until the M7.1 event time.

3.2.3 Experiment 3

In the third experiment, we simulate Ridgecrest M7.1 sequences with the same settings as for *Experiment 2*, but based on parameter estimates fitted over the immediately preceding M6.4 foreshock sequence. For the parameter estimation, we use the same magnitude-time-space target window as for the M6.4 sequence simulations in *Experiment 1*. We account for external triggering by including complementary events that occurred after June 1, 2019, and within a disk with increased radius of 50 km.

3.3 Fitted models

Each forecasting experiment is carried out for five different versions of the model introduced in the *Methods* section, summarized in Table 1. The "ETAS conventional" model serves as our benchmark and uses the most standard set-up of an ETAS model (e.g. Ogata 1998; Zhuang et al. 2002; Jalilian 2019). It applies an isotropic spatial kernel with infinite spatial extent to all triggers. The "ETAS iso-r" model applies an isotropic kernel, but restricts the spatial extent to one rupture length for all events, according to Eq. (7). In the "ETAS aniso-r" model, all events with magnitudes $m_i \geq M_{aniso} = 6.0$ are modeled as an anisotropic trigger source with a spatial restriction to half a rupture length (Eqs. 6 and 7). The other events are modeled as isotropic triggers, restricted to one rupture length. The "ETASI iso-r" and "ETASI aniso-r" models combine the spatial kernel settings of the latter models with the ETASI approach accounting for STAI.

3.4 Simulation process

For each forecasting experiment and model version, we carry out 10,000 realizations of synthetic sequences to obtain statistically stable results. At the beginning of each simulation, we distribute the Poisson-sampled number of background events, scaled by the size of the spatial area, uniformly over space and time. The assumption of an uniform spatio-temporal background event distribution appears plausible within the short and small space-time simulation windows.

Next, we sample the numbers of offsprings for the initiating event history and the background events. The number of offsprings, depending on trigger magnitude m, is drawn from a Poisson distribution with expected value

$$N(m) = k(m)\,\frac{1}{1-p}\left((T+c)^{1-p} - c^{1-p}\right). \tag{11}$$

where $k(m)$ is the aftershock productivity function in Eq. (2) and the latter term is the integral from $t = 0$ to a maximum trigger duration $t = T$ (in days) over the Omori-Utsu function in Eq. (1). We need to rescale the aftershock productivity to obtain the expected number of offsprings within T days, because the Omori-Utsu law is not normalized (no pdf) and, therefore, typically does not integrate to one. Thus, it interacts with the scaling parameter A of the productivity function.

Each triggered event is then assigned an event time and location according to inversion sampling from the (rescaled) Omori-Utsu law and the spatial kernel. The magnitude is sampled by the inversion method from the estimated FMD in Eq. (3), applying a maximum magnitude of 7.5 for California. The aftershock sampling routine is repeated for every newly triggered event in the simulated time-space window until no more events are sampled.

In order to make fair comparisons of simulated sequences with the observed ones, we need to consider the implications of STAI in the forecasts. The ETASI models account for incomplete records in the parameter estimation and therefore forecast the "true", i.e. complete, aftershock sequence. According to its definition of event detectability, we would need to delete all events that occurred within the blind time T_b of an earlier event with larger magnitude.

For the sake of transparency and consistency with the observations, we used an alternative approach and manually fitted empirical magnitude completeness functions

$$M_c(t) = \begin{cases} -1.4\,log_{10}(t) + M - M_c - 4.75, & \text{(Ridgecrest M6.4)}, \\ -0.99\,log_{10}(t) + M - M_c - 3.75, & \text{(Ridgecrest M7.1)}. \end{cases} \tag{12}$$

to the logarithmic event time-magnitude scatter data of the observed Ridgecrest M6.4 and Ridgecrest M7.1 sequences in Fig. 1c and d.

In the forecasts generated by the ETASI iso-r and aniso-r models, we delete all events that fall in the supposedly undetected range below the line of the simulated sequence. In contrast, ETAS models estimate STAI-biased aftershock productivities and therefore lead to predictions of the incomplete, rather than the "true" size of the sequence. Therefore, in forecasts generated by these models we do not delete events.

3.4.1 Attributes and measures

For each model version and experiment, we want to assess the quality of the forecasts with respect to three attributes, in comparison with the observed sequence evaluated over the same magnitude-time-space window.

We compute the predicted cumulative distribution function (cdf) of the number of aftershocks and the predicted pdf of the largest aftershock magnitude out of the 10,000 forecasted sequences. As a quantitative measure of the fit, we determine the exceedance probability that the predicted distribution would forecast a larger or the observed value. Extreme exceedance probabilities, either close to 0 or 1, indicate an inadequate prediction of the attribute.

To test the spatial distribution of aftershock locations, we define a 1km $\times$ 1km spatial grid over the spatial simulation window of the forecasting experiment and count the number of aftershocks in each simulation run, that occurred closest to the respective grid points. We determine the spatial distribution D_{ij} of the i-th simulation run by dividing the number of events occurred at each grid point j, N_{ij}, by the number of events in the i-th simulation run, N_i, i.e.

$$D_{ij} = N_{ij}/N_i.$$

By repeating the same procedure for each simulation run, we obtain 10,000 simulated spatial distributions D_{ij} for each model version. Finally, we average the individual simulated distributions to obtain the predicted probability P_j that an event occurs at grid point j.

The more events of the observed sequence have occurred at grid points with high predicted probabilities P_j, the better is the forecast. Therefore, we quantify the goodness of the spatial fit with the likelihood $L_{space} = \prod_{grid\ points\ j} P_j^{N_j^{obs}}$ where N_j^{obs} is the number of observed events at grid point j with corresponding log-likelihood

$$LL_{space} = \sum_{grid\ points\ j} N_j^{obs}\, ln(P_j).$$

We compute the information gain of the models' spatial predictions relative to the ETAS conventional model by

$$IG = \frac{LL_{space} - LL^0_{space}}{N_{obs}}$$

where LL^0_{space} is the benchmark result for the ETAS conventional model (Hainzl 2021; Rhoades et al. 2014).

3.4.2 Strike and rupture position estimates

For anisotropic models, both the parameter estimation and the forward simulations of the Ridgecrest M6.4 and M7.1 sequences require estimates of the strike angle and rupture position of all events with magnitude $M > 6.0$.

Figure 4a demonstrates the methodology, described in the *Methods* section, for the Ridgecrest M6.4 foreshock. The forward trigger rate contribution (y axis) from Eq. (10) is plotted against the strike sample (x axis) and the sample of relative rupture positions (red lines). The curves clearly show a bi-modal shape, with peaks at strikes 34° and 132° and relative rupture positions 0.76 and 0.77. Fig. 4c visualizes the optimized rupture orientation and position as a fit through the cloud of potential aftershocks within the first hour (red) or within 30 hours (yellow). It confirms the earlier mentioned particularity of two almost orthogonally ruptured faults. The strike 34° rupture segment does not perfectly fit the aftershock alignment, as segment must go through the fixed M6.4 epicenter location which seems to be slightly off the ruptured fault. Apparently, later aftershocks have a very similar spatial distribution as the events occurred within the first hour. For larger Δt, the M6.4 strike estimates would vary by only 1° or 2°, respectively.

Figure 4b shows the analogous analysis for the M7.1 Ridgecrest mainshock. Here, the maximizing properties are strike 142° and a relatively central rupture position 0.55. The M7.1 event ruptured a single fault, resulting in an unimodal shape of the forward trigger contribution curves.

4 Results and discussion

In this section, we analyze and discuss the results of the three forecasting experiments, summarized in Fig. 3. We use the attributes and measures presented in the *Application* section to evaluate the quality of the forecasts, compared to the observed sequences. The model parameter estimation results of both the generic California and the Ridgecrest M6.4 sequence parameter fits are listed in Table 2 and will help us to understand and explain features in the forecasts. After a rigorous discussion of the forecasting results, we will mention some sensitivity tests that we applied to check the robustness of our findings.

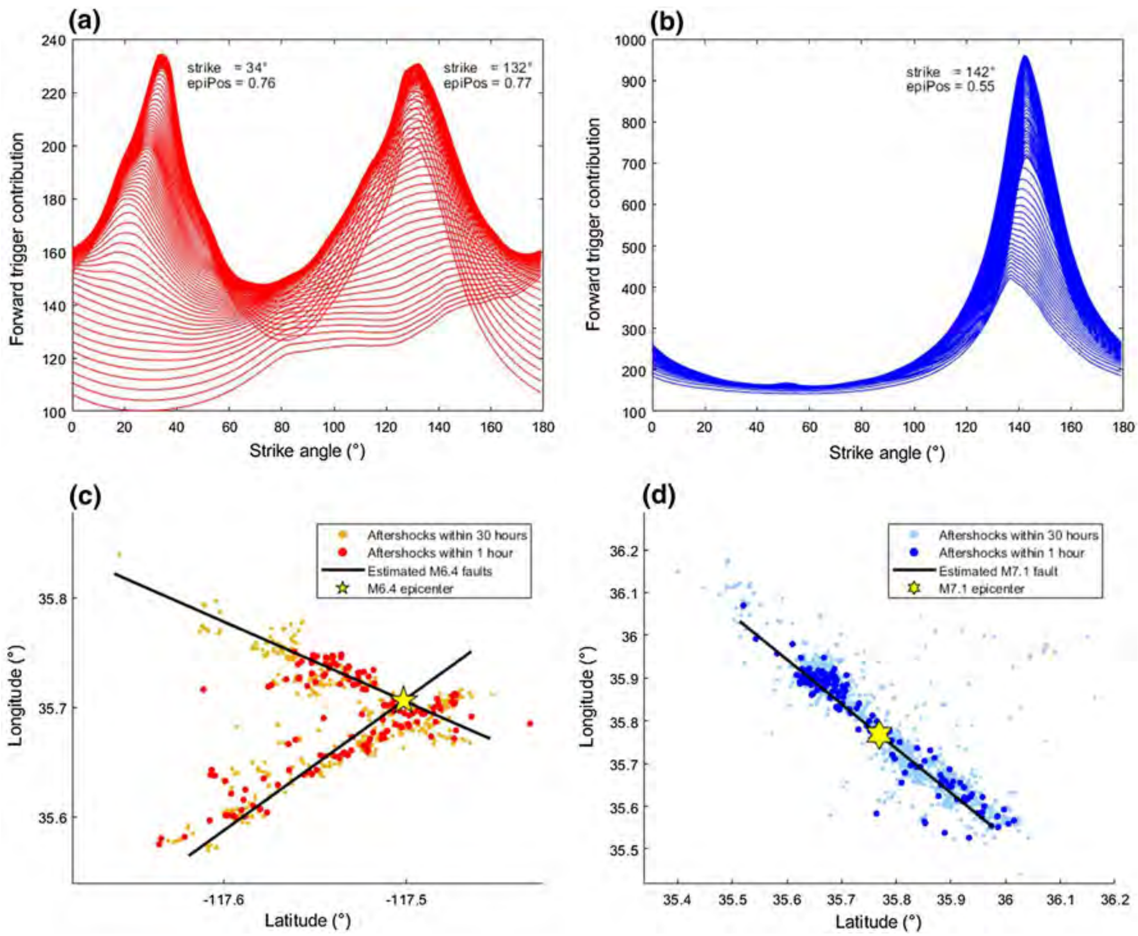


Fig. 4 Strike and relative rupture position optimization using initial ETAS parameter guesses $D = 0.0025, \gamma = 1.78, q = 1.71$. **a**, **b**: Sum of forward trigger rate contributions to events within one hour against strike sample (x axis) and relative rupture position sample (curves) for **a** the M6.4 Ridgecrest foreshock and **b** the M7.1 Ridgecrest mainshock. Text boxes show strike and relative rupture position estimates at the curve maxima. **c**, **d**: Fitted rupture segments through cloud of aftershocks after **c** the M6.4 Ridgecrest foreshock and **d** the M7.1 Ridgecrest mainshock. Darker red and blue points represent aftershocks within the first hour after the respective trigger event, yellow and lighter blue points represent aftershocks within the first 30 hours. Yellow pentagram symbolizes Mw6.4 foreshock, and yellow hexagram marks Mw7.1 mainshock. Thick black lines represent estimated rupture locations according to the strikes and relative rupture positions estimated in **a** and **b**.

4.1 Forecasting experiment 1

In the first forecasting experiment, we simulated the Ridgecrest M6.4 sequence, starting at the time of the M6.4 earthquake occurrence, based on generic parameters, fitted on a long-term and spacious Californian event set. The simulation period covers the 34 hours until (but non-including) the occurrence of the M7.1 mainshock.

4.1.1 Predicted aftershock productivity

Figure 5a shows the predicted cdfs of the number of aftershocks for each model, compared to the observed M6.4 sequence, which produced 633 events in the same time-space window. Evidently, the ETAS conventional model (with isotropic, unlimited spatial kernel) provides a very inappropriate estimate, as it does not reach the observed number in any of the 10,000 simulations. According to the ETAS iso-r and ETAS aniso-r models, the observed number of events would be a possible, but rather unlikely scenario, with approximately 2.4 and 3.7% probability to exceed the observed value. The ETASI models tend to only moderately (ETASI iso-r) or slightly (ETASI aniso-r) underestimate the observed number.

To explain this observation, we consider that the size of this relatively short sequence is predominantly influenced by the amount of direct aftershocks of the initial M6.4 trigger event. According to the model parameter estimates in Table 2 and Eq. (11), the M6.4 trigger event would only produce approximately 17 direct aftershocks in the ETAS conventional model, compared to 46 (ETAS iso-r), 49 (ETAS aniso-r), 66 (ETASI iso-r) and 74 (ETASI aniso-r)

Table 2 Overview of model results for generic (long-term) California and Ridgecrest M6.4 parameter estimation

Parameter		Generic California Estimates					Ridgecrest M6.4 Estimates				
		ETAS			ETASI		ETAS			ETASI	
		conv	iso-r	aniso-r	iso-r	aniso-r	conv	iso-r	aniso-r	iso-r	aniso-r
μ	$\frac{1}{days}$	0.16	0.21	0.21	0.21	0.21	0.11	0.30	0.29	0.18	0.30
A		0.027	0.012	0.011	0.010	0.009	0.052	0.024	0.022	0.022	0.019
α	$\frac{1}{mags}$	1.30	1.87	1.92	1.98	2.05	1.13	1.71	1.75	1.76	1.83
c	$\frac{1}{days}$	0.004	0.010	0.010	0.005	0.005	0.008	0.015	0.014	0.010	0.007
p		1.06	1.08	1.08	1.09	1.09	1.16	1.09	1.06	1.07	1.04
D	Km^2	0.085	0.037	0.110	0.037	0.107	0.135	0.085	0.469	0.080	0.399
γ	$\frac{1}{mag}$	1.60	1.86	2.09	1.88	2.10	1.15	1.43	1.55	1.44	1.57
q		1.51	1.03	2.14	1.07	2.20	1.93	1.73	8.98	1.72	8.79
T_b	sec			112.8	114.0				18.1	21.1	
b		0.98	0.98	0.98	1.01	1.01	0.72	0.72	0.72	0.77	0.79
LL		20,806	17,478	18,209	16,321	17,107	6524	6322	6433	6013	6131
ν_{branch}		0.73	0.60	0.59	0.61	0.60	1.38	1.76	1.89	1.54	1.52

in the other models. The larger the parameter α, the more direct aftershocks are expected for the M6.4 event.

As argued in the *Methods* section, the local restriction of the spatial kernels prevents a disproportionate triggering power of small events and in return increases the direct aftershock productivity of the stronger events, characterized by a considerable increase of the parameter α in the non-conventional models (Grimm et al. 2021). Besides, the application of the ETASI model accounts for missing aftershock records after strong trigger events and corrects for the biased, underestimated aftershock productivity, leading to an additional increase of α (Hainzl 2021). Finally, we note that the majority of the $M > 6$ mainshocks included in the estimation time window from 1987 until 2018, are characterized by anisotropic aftershock patterns. Consequently, more events are associated as direct aftershocks of the strong mainshocks when we estimate the parameters with the ETAS aniso-r or the ETASI aniso-r model.

4.1.2 Predicted largest aftershock magnitude

Figure 5b shows the predicted pdfs of the largest aftershock magnitude in the synthetic sequences, assuming that the Gutenberg-Richter distribution holds over the entire magnitude range up to the largest values. For each of the five models, a kernel density function was computed for the 10,000 largest magnitude samples. In all models, the observed M7.1 event would have been an extremely rare case, with exceedance probabilities $P(largest\ magnitude \geq 7.1) \leq 0.43\%$. Even the second largest, observed aftershock magnitude ($M = 5.4$) was not reached in approximately 75% of the simulations of the best model (ETASI aniso-r).

To interpret this result, think of the largest aftershock magnitude as the largest order statistic of the sample of simulated events in a simulation run. Then, the expected value of the sample maximum (i.e. the largest aftershock) increases if (1) the sequence size becomes larger or (2) if the magnitudes in the sample are distributed in a way that they favor high values.

The underestimations of the observed sequence size, shown in Fig. 5a and discussed earlier, cannot sufficiently explain the miss-match of the predicted largest aftershock magnitudes. Even the observed sample size (633 events) would produce a M7.1 event with a probability of less than 1%, given the generic California estimates for the FMD with $b = 0.98$ (ETAS models) or $b = 1.01$ (ETASI models; see Table 2). If $b = 1$, then each magnitude increment by 1 leads to a 10 times smaller probability of occurrence. Therefore, one M7.1 event is only obtained, on average, for a sequence with 100,000 aftershocks.

According to the results in Table 2, all models estimated significantly smaller values $b < 0.8$ for the observed Ridgecrest M6.4 sequence, which favors the occurrence of strong events. Note that the b estimates of the three ETAS models are biased, because they are fitted for the "true" FMD using the evidently short-term incomplete M6.4 sequence event record (see Fig. 1c). The ETASI models account for the missing smaller magnitudes and therefore lead to corrected, larger b values.

If we would simulate the Ridgecrest M6.4 sequence using its own estimation results (note that this is not a valid forecasting experiment, but used for illustration purposes),

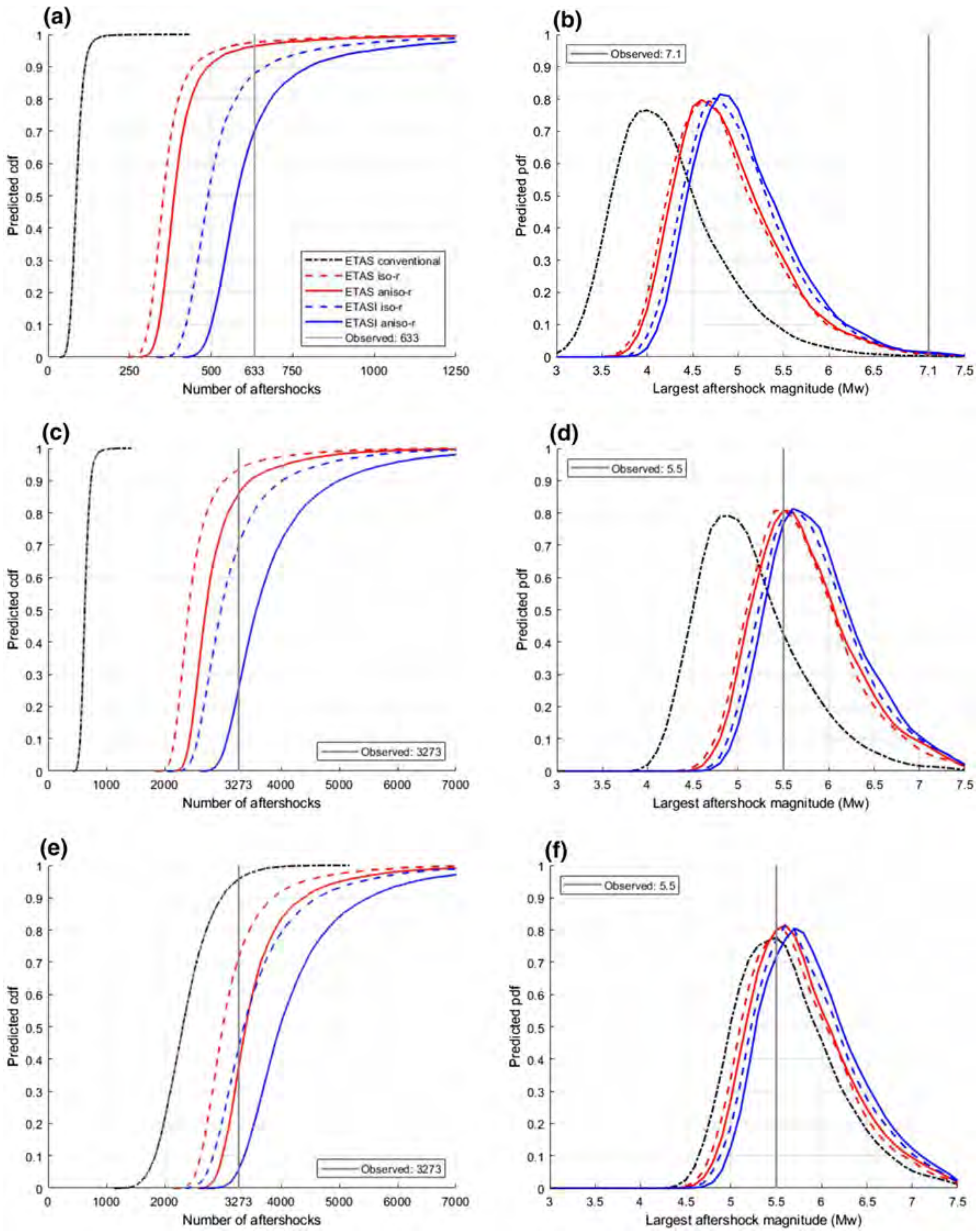


Fig. 5 Predicted cdfs of the number of aftershocks (**a**, **c**, **e**) and predicted pdfs of the largest aftershock magnitude (**b**, **d**, **f**) for *Experiment 1* (**a**, **b**), *Experiment 2* (**c**, **d**) and *Experiment 3* (**e**, **f**). Each predicted distribution is based on 10,000 simulated forecasts of the Ridgecrest M6.4 sequence (**a**, **b**) and the Ridgecrest M7.1 sequence (**c**–**f**), using the models indicated in the legend in the top left figure. Vertical gray lines show the value of the observed sequence

we would obtain an $M \geq 7.1$ event with 10.0% (ETAS conventional), 25.9% (ETAS iso-r), 53.7% (ETAS aniso-r), 15.6% (ETASI iso-r) and 25.3% (ETASI aniso-r) chance.

4.1.3 Criticality

The branching ratios ν_{branch}, i.e. the average number of direct aftershocks per event, clearly exceed 1 in each model (see Table 2). According to these models, an earthquake would trigger more than one direct aftershock on average, which would cause the sequence to be unstable, with exponentially increasing activity. The large branching ratios are mainly driven by the small *b* values, which substantially increase the occurrence probability of the more productive, strong earthquakes.

The instability of the M6.4 sequence could be interpreted as an indication of an imminent, strong mainshock. On the other hand, it is unlikely that the instability is based on a model error, e.g. due to a substantial misfit of the b-value due to few magnitude outliers. First, the FMD is estimated accounting for all earthquakes at equal weights, regardless of their magnitude. Therefore, the *b* value

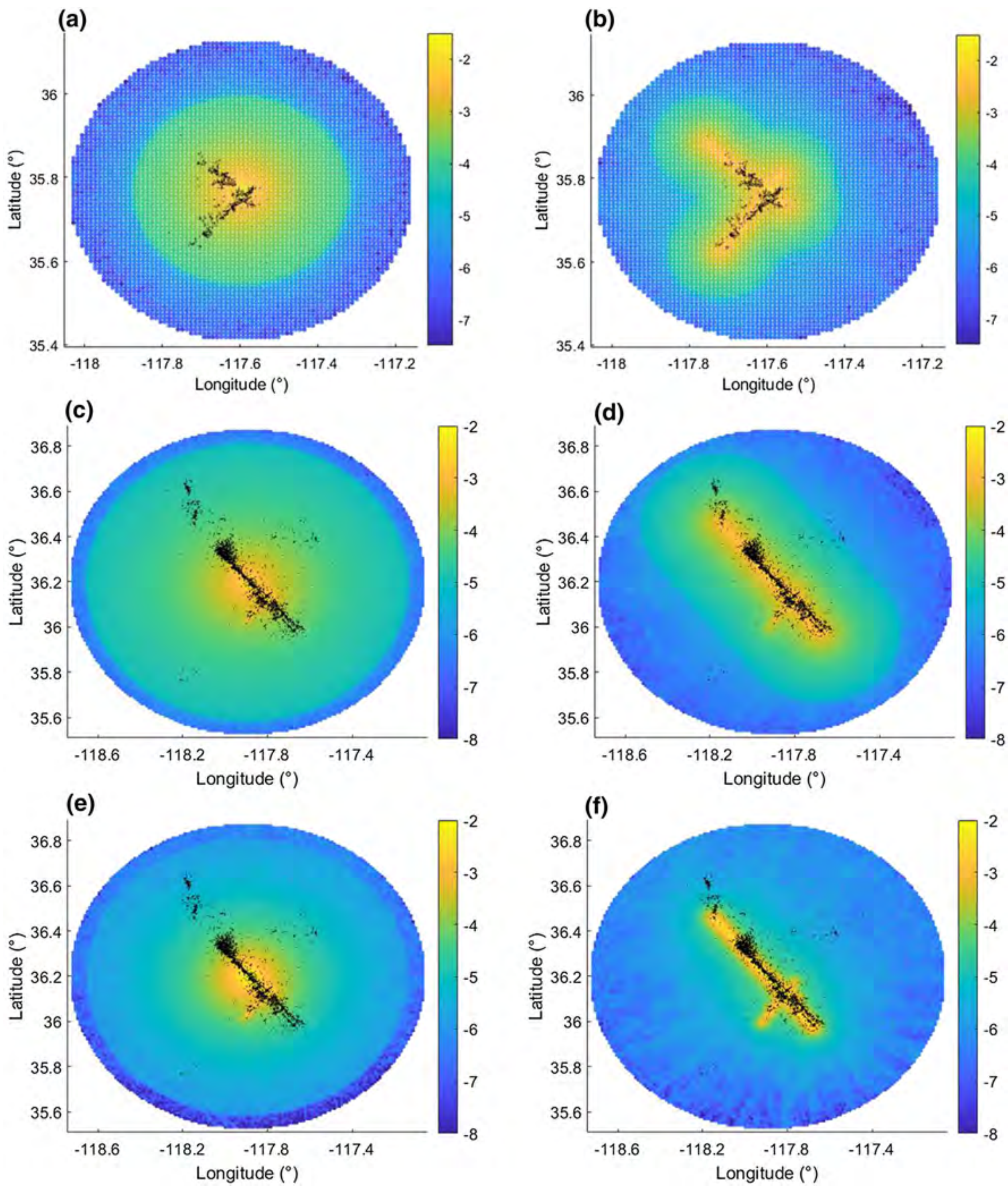


Fig. 6 Predicted spatial event distributions for *Experiment 1* (**a**, **b**), *Experiment 2* (**c**, **d**) and *Experiment 3* (**e**, **f**). Each predicted distribution is averaged over 10,000 simulated forecasts of the Ridgecrest M6.4 sequence (**a**, **b**) and the Ridgecrest M7.1 sequence (**c**–**f**), based on the ETASI iso-r model (**a**, **c**, **e**) and the ETASI aniso-r model (**b**, **d**, **f**) . The color bar indicates the predicted, logarithmic probability that an event occurs at the respective grid point

estimate is primarily controlled by the more numerous, small magnitudes. Secondly, the M7.1 event magnitude was *not* included in the *b* value estimation.

In summary, the generic California parameters are fitted to a long-term catalog mainly consisting of stable earthquake sequences and seismically quiet periods. Therefore, it cannot adequately predict the magnitude distribution of the M6.4 foreshock sequence, which is characterized by instability due to a particularly flat FMD.

4.1.4 Spatial distribution

Figure 6a and b show the predicted spatial event distributions, averaged over the 10,000 simulation runs and evaluated on the 1 km $\times$ 1 km grid described in the *Application* section, for the ETASI iso-r model (in (a)) and the ETASI aniso-r model (in (b)). We overlay the observed event locations to the logarithmic heat map of grid cell probabilities. At first glance, the anisotropic spatial forecast in (b) fits the observed, and clearly non-isotropical event distribution much better than the isotropic counterpart in (a).

In the isotropic model, all direct aftershocks are distributed point-symmetrically around the M6.4 trigger event. Subsequent secondary triggering then takes place around the new initiators. In the anisotropic model, the direct aftershocks are distributed around the fitted rupture segments of the two orthogonal faults (see Fig. 4). Subsequent trigger generations then spread isotropically (if $M < M_{aniso}$) or anisotropically (if $M \geq M_{aniso}$) around their new initiators. In both plots, we can see a pronounced boundary from green to blue color, indicating the spatial restriction to one rupture length (isotropic model) and half a rupture length (anisotropic model) around the trigger source, according to Eq. (7). Spatial grid cells outside of this boundary can only be activated by secondary triggering or background seismicity.

To quantify the quality of the spatial forecasts, we computed the information gains relative to the ETAS conventional model as described in the *Application* section. Figure 7c shows the results for *Experiment 1* in the left group of bars. Both anisotropic models lead to a pronounced improvement, which confirms the visual impression in Fig. 6a and b. The ETAS and ETASI iso-r models, which differ from the conventional approach in terms of the local spatial restriction, show a small information gain. As we can see in Fig. 6a, none of the observed events occurred outside of the spatial restriction. Therefore, the restriction leads to a slightly more pronounced accumulation of simulated event locations closer to the M6.4 trigger, which coincides better with the observation.

4.2 Forecasting experiment 2

In the second forecasting experiment, we simulated the Ridgecrest M7.1 sequence for a duration of 10 days based on the same generic California parameters as used for *Experiment 1*.

4.2.1 Predicted aftershock productivity

Figure 5c compares the number of aftershocks, predicted by the five models, to the observed number of 3,273 events. The forecasts show a very similar setup of curves as in *Experiment 1* (see Fig. 5a). The ETAS conventional model clearly underestimates the observed number of events. The ETAS iso-r and aniso-r models reach the observation in

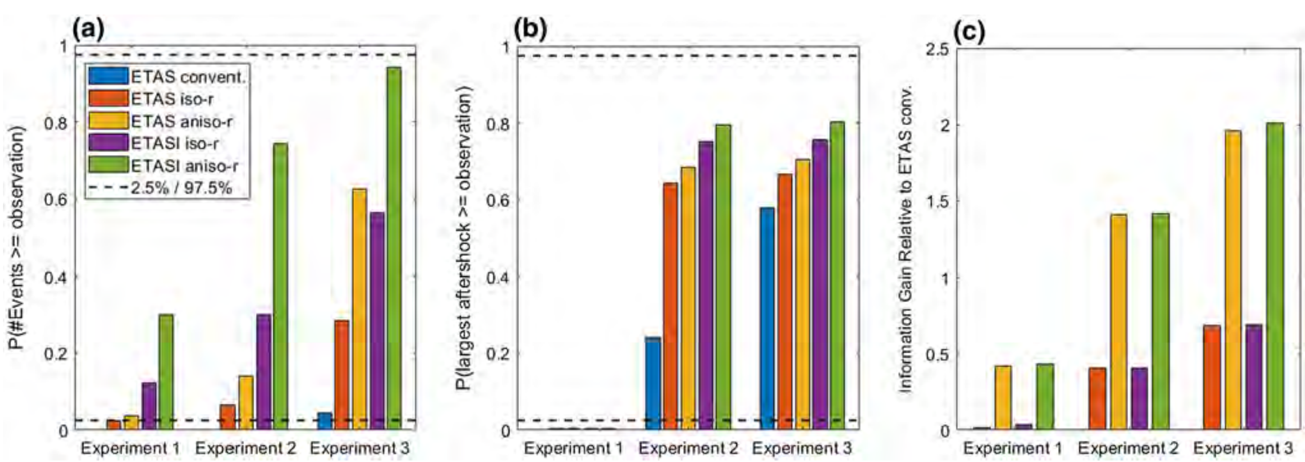


Fig. 7 Summary plots of forecasting results. Predicted probabilities per model that **a** the number of aftershocks exceeds the observation (633 for Ridgecrest M6.4; 3,273 for Ridgecrest M7.1) and **b** the largest aftershock magnitude exceeds the observation (7.1 for Ridgecrest M6.4; 5.5 for Ridgecrest M7.1). Dashed horizontal lines represent 2.5% and 97.5% quantiles. **c** Spatial information gains relative to the ETAS conventional model prediction for the same experiment. Legend in **a** holds for all plots

about 6.5 and 14.1% of the simulation runs. Again, the ETASI models provide the best approximations.

According to Eq. (11), the M7.1 trigger event would on average trigger only roughly 53 direct aftershocks in the ETAS conventional model, compared to 219 in the ETAS iso-r, 242 in the ETAS aniso-r, 328 in the ETASI iso-r and 387 in the ETASI aniso-r model. As explained in detail for *Experiment 1*, the reason is found in the parameter estimate for α.

4.2.2 Predicted largest aftershock magnitude

Figure 5d shows the predicted pdfs for the largest aftershock magnitude of the Ridgecrest M7.1 sequence. In contrast to *Experiment 1*, all but the conventional model provide very good forecasts, indicating that the generic, long-term California estimates of the FMD with $b \approx 1$ coincide well with the FMD of the Ridgecrest M7.1 sequence and the instability of the sequence ended with the occurrence of the M7.1 mainshock. The moderate underestimation by the ETAS conventional model can be explained by the underestimated sequence size, which substantially reduces the sample size of event magnitudes.

4.2.3 Spatial distribution

Figure 6c and d show the predicted spatial distributions of aftershock locations, again for the ETASI iso-r and aniso-r model. The visual impression, that the anisotropic model provides a substantially better forecast, is confirmed by the bar plot in Fig. 7c. The information gain by the anisotropic models is more pronounced for the Ridgecrest M7.1 sequence, because it has a longer rupture extension ($\sim 68km$ by Wells and Coppersmith 1994) than the M6.4 event and it did not rupture two orthogonal faults, which can be approximated more easily by an isotropic kernel.

4.3 Forecasting experiment 3

In the third forecasting experiment, we simulated the 10-days Ridgecrest M7.1 sequence based on the parameters fitted to the local Ridgecrest M6.4 foreshock sequence. Since the instability of the sequence would lead to exploding forecasts, we assumed the long-term estimated FMD with $b = 1$.

4.3.1 Predicted aftershock productivity

Figure 5e shows that the number of aftershocks is predicted much more similarly by the five models than in *Experiments 1* and *2*. It suggests that the particular features of the model versions play a smaller role in the estimation over a closed, local sequence than in the generic fit over a long-term catalog with several sequences and seismically quiet periods in between. The ETAS conventional model reaches the observation in 4.4% of the simulation runs, the ETASI aniso-r even overestimates the size of the sequence in 94.1% of the simulations. The other models show very good predictions.

4.3.2 Predicted largest aftershock magnitude

According to Fig. 5f, our manual choice of $b = 1$ led to very realistic predictions of the largest aftershock magnitude. Together with the results for the number of aftershocks, it shows that the Ridgecrest left the unstable state after the M7.1 event by returning to the generic FMD, while retaining a similar structure of aftershock productivity.

4.3.3 Spatial distribution

Finally, Fig. 6e and f suggests that, compared to *Experiment 2*, the spatial kernels fitted over the Ridgecrest M6.4 sequence are much narrower than those coming from the generic, long-term model fit. This is confirmed by the larger estimates of q and the smaller estimates of γ in Table 2. Figure 7c shows that the narrower spatial distribution leads to a more pronounced information gain by the local restriction and the anisotropy, relative to the ETAS conventional model.

Note that, to some extent, the predicted spatial distributions show traces of late or secondary aftershocks triggered along the orthogonal M6.4 Ridgecrest fault, in contrast to very few observed events in that area. This might be an indication of an underestimated Omori parameter p or an overestimated c, favoring pronounced triggering over a longer time period.

4.4 Summary of forecast quality

Figure 7 shows a summary of the quality measures for the three experiments, with respect to the predicted number of aftershocks in Fig. 7a, largest aftershock magnitude in Fig. 7b and spatial aftershock distribution in Fig. 7c. The conventional model scores worst in each category. It confirms the results in Grimm et al. (2021), who argued that the isotropic and unlimited spatial kernel assumes an implausibly far trigger reach and leads to underestimated cluster sizes.

According to Fig. 7a, the ETASI models seem to predominantly estimate more realistic aftershock productivties than the ETAS models when fitted over the long-term Californian catalog (see *Experiments 1* and *2*). When fitted over the specific Ridgcrest M6.4 sequence, the bias of an underestimated aftershock productivity seems to be balanced out by not cutting out undetected events. Anisotropic models always lead to larger predicted sequence sizes, in the case of *Experiment 3* even to a substantial overestimation.

The predictions of the largest aftershock magnitude, shown in Fig. 7b, are reasonable for all but *Experiment 1*. Apparently, the short-term incompleteness bias in the ETAS models is of much less consequence for the FMD than for the aftershock productivity.

According to Fig. 7c, as expected, the anisotropic models predict more realistic spatial event distributions. The spatial restriction leads to a much smaller improvement.

4.5 Sensitivity of results

As a sensitivity study, we forecasted the Ridgecrest M7.1 sequence over a duration of 50 days. In a longer time window, direct aftershock productivity has less dominance, and is being displaced more and more by secondary triggering. The underestimation of direct aftershock productivity (e.g. in the ETAS conventional model) typically goes along with more pronounced secondary triggering, characterized by larger estimates of the productivity constant A, see Table 2. Therefore, we observed that the ETAS conventional model caught up the missing events over time. On the other hand, this indicates a temporal distribution of aftershocks which is not in agreement with the observations. Other sensitivity tests, such as the model estimation with varying cut-off magnitudes M_c or different time windows for the generic California estimates showed generally stable results.

5 Conclusion

In this article, we combined an ETAS approach with generalized anisotropic and locally restricted spatial kernels (Grimm et al. 2021) with the ETASI time model of Hainzl (2021). The new features have the objective to solve the three major biases of the conventional ETAS model, which are the isotropic and spatially unlimited kernel as well as the neglection of short-term incompleteness in the fitted event records.

We estimated five different versions of the new ETASI time-space model to a generic, long-term Californian event set and to the specific Ridgecrest M6.4 foreshock sequence. Then, we applied the fitted model parameters to generate synthetic forecasts of the Ridgecrest M6.4 and the M7.1 sequences, which we analyzed regarding the predicted size of the sequence, largest aftershock magnitude and spatial aftershock distribution.

The results indicate that the ETAS conventional model leads to a substantial underestimation of the number of aftershocks due to its disproportionately small estimates of the direct aftershock productivity for the M6.4 and M7.1 trigger events. The locally restricted ETAS models without ETASI-extension provide more realistic, but still underestimated predictions, as they are affected by the bias of short-term incomplete event sequences in the fitted event set. The combination of ETASI model with locally restricted spatial kernel seems to solve the bias and provides the most robust predictions in the forecasting experiments. The anisotropy of the spatial kernel has a positive impact on the model estimation, however, shows its strength more clearly in the prediction of the spatial event distribution of aftershocks.

More as a by-product, we find that the Ridgecrest M6.4 foreshock sequence showed instable behavior, favoring strong aftershocks by a small Gutenberg-Richter parameter $b < 0.8$. The instability of the foreshock sequence can be interpreted as an indication of an imminent strong mainshock. In consequence, models fitted on the long-term, stable Californian event records cannot adequately predict the magnitude distribution of this sequence.

The new model provides a better understanding of spatio-temporal earthquake clustering and solves three major biases of the conventional ETAS model at once. Particularly, it leads to better estimates of the aftershock productivity and to improved forecasts of the size of a sequence and the spatial distribution of aftershocks. These improvements may be of major interest for short-term risk assessment during an on-going aftershock sequence, particularly for the risk of a second, damaging earthquake following the initial trigger event. The anisotropic spatial forecast of aftershock locations enables desaster response managers to take actions in areas at risk where particularly high aftershock activity is expected.

Future work should test the forecast quality for other earthquake sequences. It would be interesting to address the question whether the ETASI time-space model can be used to reliably detect the instability of a live sequence, which could have positive impacts on emergency management during on-going sequences. An evaluation of the goodness of fit for the temporal event distribution should be included into such analyses.

6 Data and resources

The earthquake event set for California has been downloaded from the Southern California Earthquake Data Center (https://scedc.caltech.edu/data/alt-2011-dd-hauksson-yang-shearer.html, last accessed on October 25, 2021). Results and figures were produced using Matlab software. The source code for model estimation and simulation is made freely available by the first author in the Github repository https://github.com/ChrGrimm/ETASanisotropic.

Acknowledgements We thank the anonymous reviewers for their helpful and constructive feedback that has significantly improved the work. Particularly, we thank Marco Pagani for valuable discussions during the on-going research. Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich. S.H. was supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre 1294 (Data Assimilation – The seamless integration of data and models, project B04) and the European Unions Horizon 2020 research and innovation program under Grant Agreement Number 821115, real-time earthquake risk reduction for a resilient Europe (RISE).

Authors contribution CG (first author) derived and programmed the model, conducted simulation experiments, prepared figures and wrote the paper. SH provided theory of ETASI time model, consulted in programming and conducted comparative calculations. CG, SH and MK designed the simulation experiments and interpreted the results. HK consulted on statistical questions and was involved in the concept of the study.

Funding Open Access funding enabled and organized by Projekt DEAL. Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich (e.g. travel costs, conference and journal fees).

Declarations

Conflict of interest The authors acknowledge that there are no relevant financial or non-financial interests to disclose.

Human or animal rights This article does not contain any studies involving human participants or animals performed by any of the authors.

References

de Arcangelis L, Godano C, Lippiello E (2018) The overlap of aftershock coda waves and short-term postseismic forecasting. J Geophys Res Solid Earth 123:5661–5674. https://doi.org/10.1029/2018JB015518 (**ISSN: 21699356**)

Grimm C, Käser M, Hainzl S, Pagani M, Küchenhoff H (2021) Improving earthquake doublet frequency predictions by modified spatial trigger kernels in the epidemic-type aftershock sequence (ETAS) model. Bull Seismol Soc Am. https://doi.org/10.1785/0120210097 (**ISSN: 0037-1106**)

Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. Bull Seismol Soc Am 34:185–188. https://doi.org/10.1038/156371a0 (**ISSN: 00280836**)

Hainzl S (2016) Rate-dependent incompleteness of earthquake catalogs. Seismol Res Lett 87(2A):337–344. https://doi.org/10.1785/0220150211 (**ISSN: 19382057**)

Hainzl S (2016) Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. J Geophys Res Solid Earth 121(9):6499–6509. https://doi.org/10.1002/2016JB013319 (**ISSN: 21699356**)

Hainzl S (2021) ETAS-approach accounting for short-term incompleteness of earthquake catalogs. Bull Seismol Soc Am. https://doi.org/10.1785/0120210146 (**ISSN: 0037-1106**)

Hainzl S, Christophersen A, Enescu B (2008) Impact of earthquake rupture extensions on parameter estimations of point-process models. Bull Seismol Soc Am 98(4):2066–2072. https://doi.org/10.1785/0120070256

Hainzl S, Zakharova O, Marsan D (2013) Impact of aseismic transients on the estimation of aftershock productivity parameters. Bull Seismol Soc Am 103(3):1723–1732. https://doi.org/10.1785/0120120247 (**ISSN: 00371106**)

Hainzl S, Zakharova O, Marsan D (2013) Impact of aseismic transients on the estimation of aftershock productivity parameters. Bull Seismol Soc Am 103(3):1723–1732. https://doi.org/10.1785/0120120247 (**ISSN: 00371106**)

Harte DS (2016) Model parameter estimation bias induced by earthquake magnitude cut-off. Geophys J Int 204(2):1266–1287. https://doi.org/10.1093/gji/ggv524 (**ISSN: 1365246X**)

Hauksson E, Yang W, Shearer PM (2012) Waveform relocated earthquake catalog for Southern California (1981 to June 2011). Bull Seismol Soc Am 102(5):2239–2244. https://doi.org/10.1785/0120120010 (**ISSN: 00371106**)

Hutton K, Woessner J, Hauksson E (2010) Earthquake monitoring in Southern California for seventy-seven years (1932–2008). Bull Seismol Soc Am 100(2):423–446. https://doi.org/10.1785/0120090130 (**ISSN: 00371106**)

Jalilian A (2019) ETAS: An R package for fitting the space-time ETAS model to earthquake data. *J Stat Softw*. https://doi.org/10.18637/jss.v088.c01. ISSN: 15487660

Lippiello E, Cirillo A, Godano G, Papadimitriou E, Karakostas V (2016) Real-time forecast of aftershocks from a single seismic station signal. Geophys Res Lett 43(12):6252–6258. https://doi.org/10.1002/2016GL069748 (**ISSN: 19448007**)

Marsan D, Ross ZE (2021) Inverse migration of seismicity quiescence during the 2019 Ridgecrest sequence. J Geophys Res Solid Earth 126(3):1–13. https://doi.org/10.1029/2020JB020329 (**ISSN: 21699356**)

Marsan D, Lengliné O (2008) Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079. https://doi.org/10.1126/science.1148783. ISSN: 00368075

Mizrahi L, Nandan S, Wiemer S (2021) The effect of declustering on the size distribution of mainshocks. Seismol Res Lett. https://doi.org/10.1785/0220200231 (**ISSN: 23318422**)

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27 (**ISSN: 0162-1459**)

Ogata Y (1998) Space-time point-process models for earthquake occurrences. Ann Inst Stat Math 50(2):379–402 (**ISSN: 00203157**)

Ogata Y (2011) Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. Earth

Planets Space 63(3):217–229. https://doi.org/10.5047/eps.2010.09.001 (**ISSN: 18805981**)

Ogata Y, Zhuang J (2006) Space-time ETAS models and an improved extension. Tectonophysics 413(1–2):13–23. https://doi.org/10.1016/j.tecto.2005.10.016 (**ISSN: 00401951**)

Omi T, Ogata Y, Hirata Y, Aihara K (2013) Forecasting large aftershocks within one day after the main shock. Sci Rep 3:1–7. https://doi.org/10.1038/srep02218 (**ISSN: 20452322**)

Omi T, Ogata Y, Hirata Y, Aihara K (2014) Estimating the ETAS model from an early aftershock sequence. Geophys Res Lett 41:850–857. https://doi.org/10.1002/2013GL058958 (**ISSN: 19448007**)

Page MT, van Der Elst N, Hardebeck J, Felzer K, Michael AJ (2016) Three ingredients for improved global aftershock forecasts: tectonic region, time-dependent catalog incompleteness, and intersequence variability. Bull Seismol Soc Am 106(5):2290–2301. https://doi.org/10.1785/0120160073 (**ISSN: 19433573**)

Rhoades DA, Gerstenberger MC, Christophersen A, Zechar JD, Schorlemmer D, Werner MJ, Jordan TH (2014) Regional earthquake likelihood models II: information gains of multiplicative hybrids. Bull Seismol Soc Am 104(6):3072–3083. https://doi.org/10.1785/0120140035 (**ISSN: 19433573**)

Seif S, Mignan A, Zechar JD, Werner MJ, Wiemer S (2017) Estimating ETAS: the effects of truncation, missing data, and model assumptions. J Geophys Res Solid Earth 122(1):449–469. https://doi.org/10.1002/2016JB012809 (**ISSN: 21699356**)

Utsu T, Ogata Y, Matsu'ura RS (1995) The centenary of the Omori formula for a decay law of aftershock activity. J Phys Earth 43:1–33

Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. Bull Seismol Soc Am 84(4):974–1002 (**ISSN 0037-1106**)

Zhang L, Werner MJ, Goda K (2018) Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. Bull Seismol Soc Am 108(6):3313–3335. https://doi.org/10.1785/0120180126

Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380. https://doi.org/10.1198/016214502760046925 (**ISSN: 01621459**)

Zhuang J, Ogata Y, Wang T (2017) Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters. Earth Planets Space 69:36. https://doi.org/10.1186/s40623-017-0614-6 (**ISSN: 18805981**)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.