


Are Regionally Calibrated Seismicity Models More Informative than Global Models? Insights from California, New Zealand, and Italy

José A. Bayona¹, William H. Savran^{2,3}, Pablo Iturrieta⁴, Matthew C. Gerstenberger⁵, Kenny M. Graham⁵, Warner Marzocchi⁶, Danijel Schorlemmer⁴, and Maximilian J. Werner¹

Abstract

Earthquake forecasting models express hypotheses about seismogenesis that underpin global and regional probabilistic seismic hazard assessments (PSHAs). An implicit assumption is that the comparatively higher spatiotemporal resolution datasets from which regional models are generated lead to more informative seismicity forecasts than global models, which are however calibrated on greater datasets of large earthquakes. Here, we prospectively assess the ability of the Global Earthquake Activity Rate (GEAR1) model and 19 time-independent regional models to forecast M 4.95+ seismicity in California, New Zealand, and Italy from 2014 through 2021, using metrics developed by the Collaboratory for the Study of Earthquake Predictability (CSEP). Our results show that regional models that adaptively smooth small earthquake locations perform best in California and Italy during the evaluation period; however, GEAR1, based on global seismicity and geodesy datasets, performs surprisingly well across all testing regions, ranking first in New Zealand, second in California, and third in Italy. Furthermore, the performance of the models is highly sensitive to spatial smoothing, and the optimal smoothing likely depends on the regional tectonic setting. Acknowledging the limited prospective test data, these results provide preliminary support for using GEAR1 as a global reference M 4.95+ seismicity model that could inform eight-year regional and global PSHAs.

Cite this article as Bayona, J. A., W. H. Savran, P. Iturrieta, M. C. Gerstenberger, K. M. Graham, W. Marzocchi, D. Schorlemmer, and M. J. Werner (2023). Are Regionally Calibrated Seismicity Models More Informative than Global Models? Insights from California, New Zealand, and Italy, *The Seismic Record*, 3(2), 86–95, doi: [10.1785/0320230006](https://doi.org/10.1785/0320230006).

Supplemental Material

Introduction

The increasing availability and quality of geophysical datasets, including earthquake catalogs, geological records, and interseismic strain rates, has enabled the creation of regional and global earthquake forecasting models, some of which underpin probabilistic seismic hazard assessments (e.g., [Pagani et al., 2020](#); [Danciu et al., 2021](#)). Regional models provide detailed approximations of seismic sources and sophisticated earthquake forecasts, given the comparatively high spatiotemporal resolution of the datasets on which they are based (e.g., the Uniform California Earthquake Rupture Forecast, Version 3; [Field et al., 2017](#)). However, the relative lack of large earthquakes makes it difficult to assess whether regional models can translate the greater data availability and resolution into more skillful forecasts, or whether they might be overfitting their calibration

datasets and providing lower skill than expected. Global models, in contrast, offer greater testability against the large damaging earthquakes due to the more frequent seismic activity observed around the world, but are necessarily coarser, and information

1. School of Earth Sciences, University of Bristol, Bristol, United Kingdom,  <https://orcid.org/0000-0002-3026-0327> (JAB);  <https://orcid.org/0000-0002-2430-2631> (MJW); 2. Southern California Earthquake Center, University of Southern California, Los Angeles, California, U.S.A.,  <https://orcid.org/0000-0001-5404-0778> (WHS); 3. Now at Nevada Seismological Laboratory, University of Nevada, Reno, Nevada, U.S.A.; 4. GFZ German Research Centre for Geosciences, Potsdam, Germany,  <https://orcid.org/0000-0002-4787-1343> (PI);  <https://orcid.org/0000-0003-3969-1059> (DS); 5. GNS Science, Lower Hutt, New Zealand,  <https://orcid.org/0000-0003-3645-6344> (KMG); 6. University of Naples Federico II, Naples, Italy,  <https://orcid.org/0000-0002-9114-1516> (WM)

*Corresponding author: jose.bayona@bristol.ac.uk

© 2023. The Authors. This is an open access article distributed under the terms of the CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

coverage can vary considerably in space. Thus, regional models are largely expected to be far superior; however, model evaluations must be done prospectively, that is, against future earthquakes, as regional models may have the advantage of better fitting the past. In this study, we test the assumption that regionally calibrated seismicity models are more informative than global models by benchmarking the predictive skills of a set of regional and global stationary earthquake-rate models that were submitted to the Collaboratory for the Study of Earthquake Predictability (CSEP; e.g., Michael and Werner, 2018) for truly prospective evaluation.

Starting in 2007, the Working Group of the Regional Earthquake Likelihood Models (Field, 2007) launched a prospective earthquake forecasting experiment in California that involved multiple datasets, models, and tests formalizing different hypotheses about seismogenesis. This experiment sparked a series of subsequent forecast experiments in other seismically active regions, including Japan (Tsuruoka *et al.*, 2012), New Zealand (Gerstenberger and Rhoades, 2010), and Italy (Marzocchi *et al.*, 2010), with the goal of developing and prospectively evaluating regional earthquake forecasting models. On a global scale, Strader *et al.* (2018) evaluated four time-invariant seismicity models against two years of prospective data, finding that the hybrid Global Earthquake Activity Rate (GEAR1; Bird *et al.*, 2015) model was the most informative. GEAR1 results from a multiplicative combination of smoothed seismicity and interseismic strain rates, globally calibrated to M_w 5.767+, shallow (≤ 70 km) earthquakes reported from 1977 to 2013 in the Global Centroid Moment Tensor catalog (i.e., Ekström *et al.*, 2012). Compared to GEAR1, regional models that participated in the aforementioned CSEP experiments rely on parameters that might be better constrained by higher-resolution regional information, or indeed they are based on datasets that may never be uniformly available on a global scale, such as historical earthquake catalogs (e.g., the EBEL, HAZGRIDX, and TRIPLES-HYBRID models), adaptive smoothed M 2+ seismicity (e.g., HKJ, HRSS-CSI), detailed geological fault maps (e.g., NZHM), spatial variations in the Gutenberg–Richter b -value (e.g., Asperity Likelihood Model [ALM], HALM, and ALM-IT), tectonic zonations (e.g., MPS04-AFTER), and kinematic models of surface velocities (e.g., NEOKINEMA). In Table 1 and Figure 1, we provide a brief description of the main features of the time-independent mainshock + aftershock models (i.e., models forecasting all earthquakes) that we consider in this study. For more details about these models, see Field (2007), Gerstenberger and

Rhoades (2010), Marzocchi *et al.* (2010), and the references therein.

Prospective Data

Prior to evaluating any seismicity model, CSEP provides strict definitions of the testing region, period, magnitude range, target earthquakes, and testing methods (e.g., Schorlemmer *et al.*, 2010). Accordingly, we prospectively evaluate the forecasting skills of regional and global models using M 4.95+ earthquakes reported in California, New Zealand, and Italy from 1 January 2014 to 1 January 2022 in the Comprehensive Earthquake Catalog of the Advanced National Seismic System (ANSS; Guy *et al.*, 2015), the GeoNet catalog (Ristau, 2013), and the Italian seismic bulletin (Bollettino Sismico Italiano; Amato *et al.*, 2006), respectively. In California and Italy, the target earthquakes are within the 0–30 km depth range, whereas, in New Zealand, these are confined to the upper 40 km of the crust (see Fig. 2).

Model Evaluation

We project GEAR1 forecasts onto the CSEP-California, New Zealand, and Italy testing regions to create three globally calibrated regional seismicity models (see Fig. 1). For the sake of simplicity, we extrapolate the original GEAR1 M_w 5.95+ earthquake rates to a magnitude threshold of 4.95, assuming a global b -value of 1 (e.g., Petruccioli *et al.*, 2019). In the case of Italy, we divide GEAR1 rates by a factor of 1.602 to correct for differences in magnitude scales used in model calibration (moment magnitude) and model testing (local magnitude; Schorlemmer *et al.*, 2010). Furthermore, we assume that, at a magnitude threshold of 4.95, the contribution of earthquakes from within the depth ranges of 30–70 km in California and Italy, and 40–70 km in New Zealand to the long-term seismicity rate is minimal, so that the comparison between GEAR1 (which extends to a depth of 70 km) and the regional models is more compatible. In the specific case of New Zealand, only 6% of the M 4.95+ seismicity recorded instrumentally from 2003 to 2021 is within the 40–70 km depth range (Ristau, 2013), whereas none of the M 4.95+ earthquakes observed between 1985 and 2021 in Italy are located deeper than 30 km.

Using the CSEP's pyCSEP toolkit (Savran, Bayona, *et al.*, 2022; Savran, Werner, *et al.*, 2022), we then conduct the paired T -test of Rhoades *et al.* (2011) to comparatively evaluate the performance of the regional models with that of GEAR1. The T -test is based on the Information Gain per Earthquake (IGPE), here obtained by a (regional) model A over a (global)

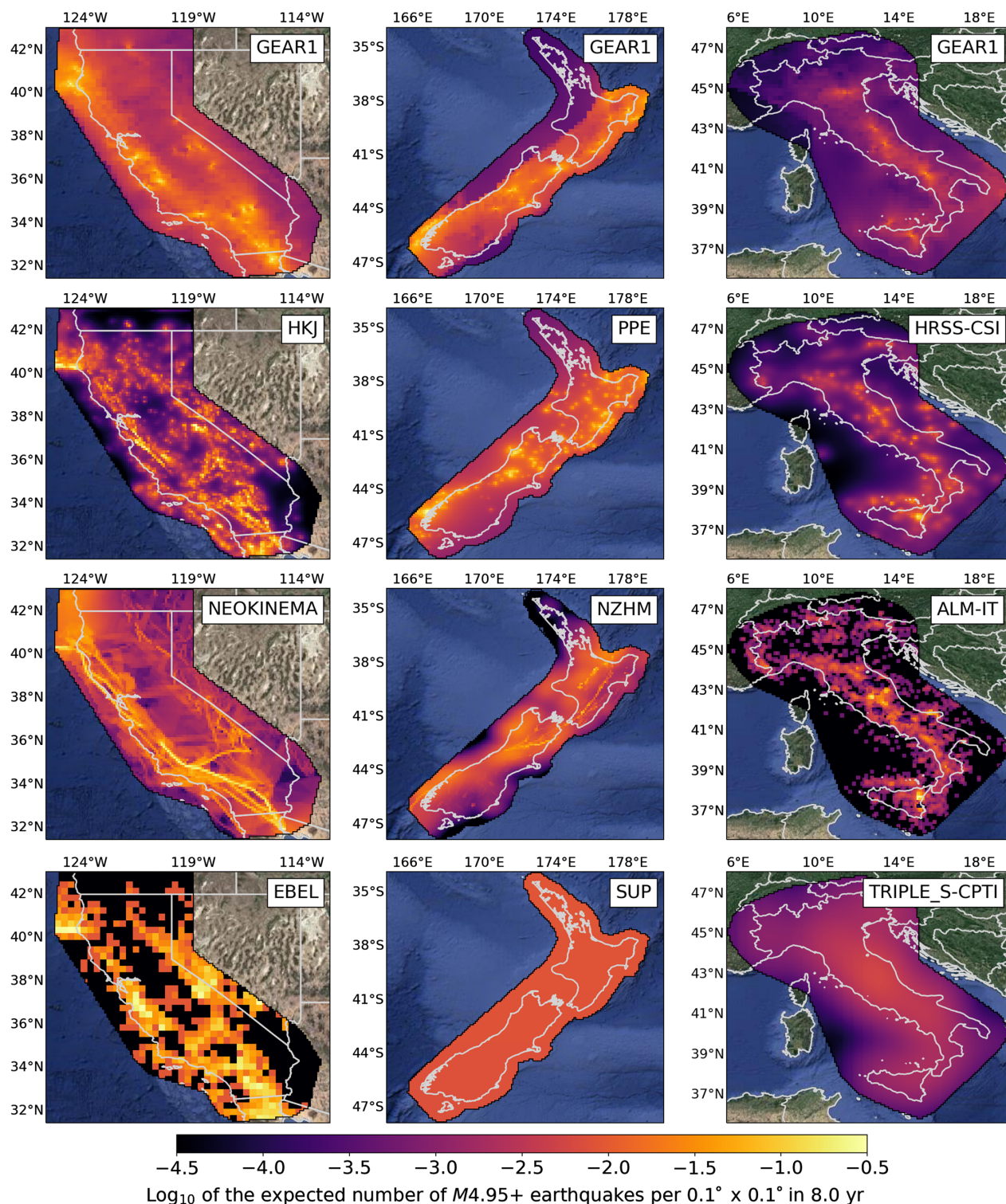


Figure 1. Global and regional time-invariant earthquake forecasting models for California, New Zealand, and Italy. The first row displays three globally calibrated the Global Earthquake Activity Rate (GEAR1) $M_{4.95+}$ seismicity forecasts, obtained from projecting GEAR1 onto the testing regions previously defined the Collaboratory for the Study of Earthquake

Predictability (CSEP). Rows 2–4 show some of the regional models that participate in regional CSEP forecast experiments. Expected numbers of $M_{4.95+}$ earthquakes from 1 January 2014 to 1 January 2022 are shown per $0.1^\circ \times 0.1^\circ$ cell. Yellow–orange colors denote high expected seismicity rates, whereas purple–black colors indicate the opposite.

Table 1

Summary of Globally and Regionally Calibrated Time-Independent Models Forecasting Non-Declustered M 4.95+ Seismicity in California, New Zealand, and Italy

Model	Region	Input Data and Main Features	Calibration Period
GEAR1	Globe	Multiplicative combination of M 5.767+ smoothed seismicity and interseismic strain rates	1977–2013
HKJ	California	Adaptive smoothing of instrumental M 2+ seismicity	1981–2005
EBEL	California	Extrapolation of $M \geq 5.0$ nondeclustered historical seismicity	1932–2004
NEOKINEMA	California	Earthquake rates inferred from plate tectonics, geodetic information, and fault kinematic data, integrated through finite element modeling	1900–2002
PI	California	Variations in seismicity associated to a preparatory phase of large earthquakes	1950–2005
SUP	NZ	Spatially uniform Poisson model averaging $M \geq 5.0$ instrumental seismicity. Earthquake rate densities vary with latitude	1965–2006
PPE	NZ	Smoothing of $M \geq 5.0$ instrumental seismicity	1965–2006
NZHM	NZ	National hazard model combining M 6.5+ historical earthquakes with M 4+ instrumental seismicity and geological data	1840–1997
ALM	Italy	Spatial variations in b -values	1981–2009
HALM	Italy	Spatial variations in b -values considering a geological zonation	1981–2009
ALM-IT	Italy	Spatial variations in b -values obtained from a declustered catalog	1950–2009
MPS04-AFTER	Italy	National hazard model defining different tectonic subregions with distinctive seismicity parameters	217BC–2003
HAZGRIDX	Italy	Smoothing of historical and instrumental M 4+ seismicity	217 B.C.–2003
HZATI	Italy	Combination of smoothed instrumental seismicity, Coulomb stress changes, and rate-and-state friction	1985–2008
RI	Italy	Variations in earthquake activity relative to a background level	1985–2009
HRSS-CSI	Italy	Adaptive smoothing of instrumental M 2.95+ seismicity	1984–2009
HRSS-HYBRID	Italy	Adaptive smoothing of M 2.95+ seismicity	1901–2006
TRIPLES-CPTI	Italy	Smoothing of historical and instrumental M 4.7+ earthquakes	1901–2006
TRIPLES-CSI	Italy	Smoothing of instrumental M 2.95+ seismicity	1981–2002
TRIPLES-HYBRID	Italy	Smoothing of historical and instrumental seismicity	1981–2006

ALM, Asperity Likelihood Model; GEAR1, Global Earthquake Activity Rate; NZHM, New Zealand Hazard Model; PI, pattern informatics; PPE, proximity to past earthquakes; RI, relative intensity; and SUP, spatially uniform poisson.

model B, and its 95% confidence intervals estimated from the Student's t -distribution (see equations 17 and 18 in Rhoades *et al.*, 2011). Each IGPE can be used as a measure of the relative amount of information about future target earthquakes that two competing models can provide. Thus, a regional model can be considered statistically more informative than GEAR1 if the IGPE is positive and its confidence interval does not include zero; whereas, if the IGPE is negative and the confidence interval

does not include zero, it can be considered that the regional model is statistically less informative than the global model. Alternatively, if the confidence interval does contain zero, the two models can be considered statistically equally informative, that is, the null hypothesis of equal forecasting performance cannot be rejected at a 0.05 significance level.

We also evaluate the spatial performance of the models by conducting the spatial S -test of Zechar *et al.* (2010) and the

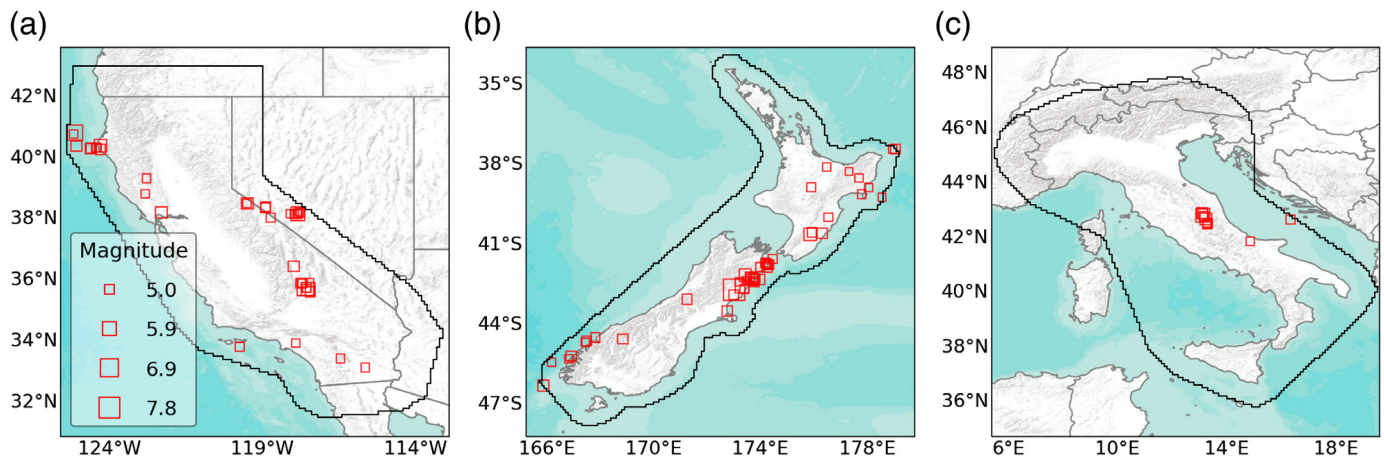


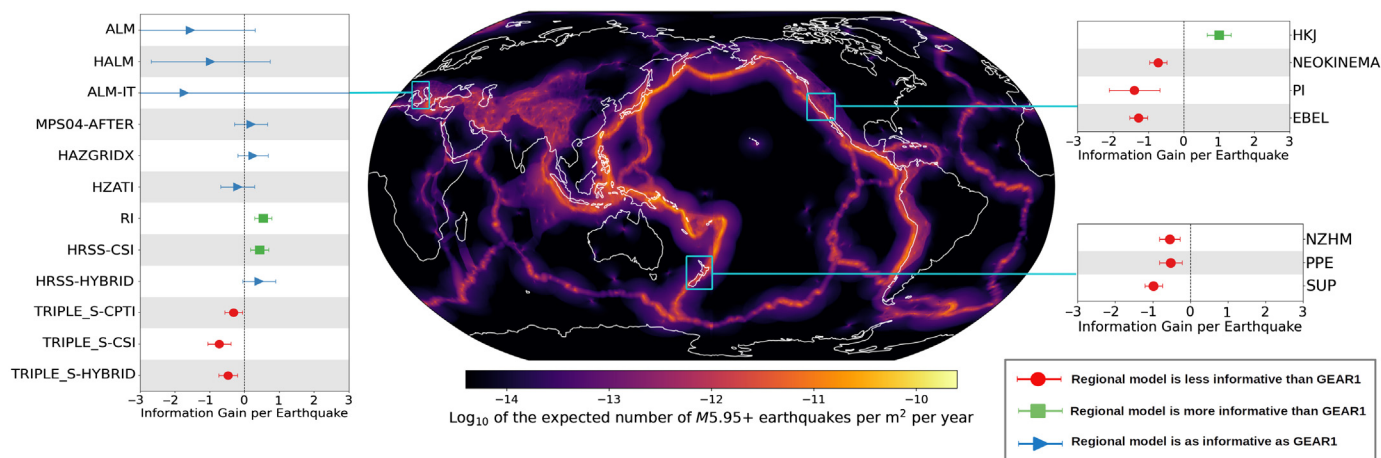
Figure 2. Maps showing the spatial distribution of prospective M 4.95+ earthquakes reported in the CSEP (a) California, (b) New Zealand, and (c) Italy testing regions by their respective authoritative earthquake catalogs from 1 January 2014 to 1 January 2022. Target earthquakes are confined to depth ranges of 0–30 km in California and Italy, and 0–40 km in New Zealand. The testing datasets contain 38 earthquakes in California, including the M_w 7.1 mainshock of the 2019 Ridgecrest sequence, 47 earthquakes in New Zealand, including the 2016 M_w 7.8 Kaikōura earthquake, and 11 target events in Italy, including the 2016 M_w 6.6 Norcia earthquake.

binary S -test of Bayona *et al.* (2022). Similar to the S -test, the binary S -test assumes that earthquakes in individual cells are independent; however, this test is based on the likelihood of observing either 0 or 1+ earthquakes in each cell, rather than observing a specific number of events ω , which makes the assumption by Zechar *et al.* (2010) that earthquakes are spatially Poisson distributed less critical. Given the differences in the number of $0.1^\circ \times 0.1^\circ$ cells in each testing region (7682 in California, 6343 in New Zealand, and 8993 in Italy) and the number of target events observed during the 8 yr evaluation period (38 in California, 47 in New Zealand, and 11 in Italy), we calculate joint log-likelihood scores per earthquake and active cell, that is, the sum of log-likelihood scores obtained by each forecast in each individual cell divided by the total number of observed earthquakes and activated cells, respectively, to analyze and compare the spatial forecasting capabilities of all the models (see equations 5 and 12 in Bayona *et al.*, 2022). In addition, we compute Gini coefficients (G_c ; Breiman *et al.*, 1984) for each model to measure how localized or smooth their predicted earthquake rates are and thus better interpret our spatial test results. The G_c is equal to twice the area between the receiver operating characteristic (ROC; e.g., Molchan and Keilis-Borok, 2008) curve and the diagonal line, with the ROC curve being the normalized cumulative forecast rate in cells sorted in decreasing order, and takes values in the interval [0,1). A G_c that tends to zero indicates that the earthquakes are expected to be homogeneously distributed throughout the testing region, whereas a G_c that tends to one denotes very localized seismicity (see the TRIPLE_S-CPTI [$G_c \approx 0.5$] and ALM-IT [$G_c \approx 0.8$] models for Italy in Fig. 1 as two contrasting examples). Finally, we apply the number N -test (i.e., equations 5–8 in Zechar *et al.*, 2010) to assess the consistency between the observed and expected

number of earthquakes and share these results in the supplemental material to this article (see [Data and Resources](#)).

Results and Discussion

Prospective T -test results for California show that the regional HKJ model is the most informative, obtaining an IGPE of approximately 1.0 over GEAR1, whereas GEAR1, ranking second, exceeds the predictive skill of the NEOKINEMA, Pattern Informatics (PI), and EBEL models (see Fig. 3). These results provide additional evidence that small earthquakes and geodetic data are useful for delineating areas in California where moderate-to-large events may occur (e.g., Kafka and Levin, 2000; Zeng *et al.*, 2018), as HKJ and GEAR1 use M 2+ earthquakes and interseismic strain rates to forecast seismicity, respectively. Furthermore, these results suggest that California is currently in a period of relatively low on-fault seismicity, as NEOKINEMA—a kinematic model of surface velocities based on fault information, markedly overestimates the observed seismicity rate and poorly explains the spatial distribution of earthquakes (see Fig. S1, available in the supplemental material to this article and the additional resources referenced in [Data and Resources](#)). Similarly, PI and EBEL



perform poorly, because the EBEL model significantly overpredicts the number of M 4.95+ earthquakes, and because PI is particularly sensitive to the unlikely occurrence of the 2019 Ridgecrest sequence (Bayona *et al.*, 2022).

In New Zealand, T -test results show that GEAR1 is the most skillful model to forecast M 4.95+ seismicity during the target period, obtaining IGPEs of about 0.5 over the 2002 NZHM (Stirling *et al.*, 2002) and Proximity to Past Earthquakes (PPE) (see Fig. 3). These results are due in part to the fact that the NZHM was designed to forecast mainshocks only; nonetheless, the model provides a number of target earthquakes that is consistent with the observations (see Fig. S1). A more important factor influencing these results is the spatial dimension of the models. Residual analyses of (Poisson) spatial-likelihood scores obtained by GEAR1 and NZHM in individual cells show that the NZHM outperforms GEAR1 in forecasting the occurrence of the M_w 7.8 Kaikōura earthquake and some of its surrounding aftershocks, whereas GEAR1 outperforms the NZHM in forecasting the $M \sim 5$ events that nucleated about 170 km northeast from the mainshock (see Fig. 2b and Figs. S2a,b). These differences can be explained considering that the NZHM contains fault information near the Kaikōura rupture area that adds predictive skill, whereas GEAR1, through its seismicity model component (i.e., the KJSS model), incorporates earthquake information derived from the 2013 Cook Strait sequence (Hamling *et al.*, 2014), which is not present in the regional models, because they were trained with data through 2006. In addition, GEAR1, through its geodesy model component (i.e., the SHIFT_GSRM2f model), includes interseismic strain rates, which appear to be useful for mapping the location of the “closest” aftershocks of the M_w 7.8 Kaikōura earthquake (see how the relative performance of GEAR1 to NZHM is superior to that of KJSS in Fig. S2).

Figure 3. Prospective T -test results for globally and regionally calibrated seismicity models for California, New Zealand, and Italy. We show Information Gain per Earthquake (IGPE) obtained by 19 regional models over GEAR1, along with their 95% confidence intervals shown as bars. Green squares denote regional models that can be considered statistically more informative than GEAR1, blue triangles show regional models that can be considered as informative as GEAR1, and red circles display regional models that are less informative than GEAR1. We include a global forecast map showing M_w 5.95+, $d \leq 70$ km estimates of seismicity per square meter per year, originally provided by the GEAR1 model.

The results of the T -test for Italy show that only relative intensity (RI) and the adaptively smoothed HRSS-CSI seismicity model can be considered statistically more informative than GEAR1, as they obtain IGPEs of about 0.7 and 0.6 over the global model, respectively (see Fig. 3). As in California, these results suggest that small (M 3+) earthquakes might be useful for forecasting the occurrence of larger events in Italy. GEAR1, on the other hand, outperforms the ALM and TRIPLE_S model families, and can be considered as informative as the HZATI, HAZGRIDX, and MPS04-AFTER models, the latter being the official time-independent model of the 2004 National Seismic Hazard Model for Italy (MPS Working Group, 2004). These results differ in part from the results obtained by Taroni *et al.* (2018), who prospectively found that HZATI, MPS04-AFTER, and TRIPLE_S-CPTI were the best-performing time-invariant models during the 2009–2014 period. These discrepancies are mainly due to the absence of the 2012 Emilia earthquake sequence in our testing catalog, which, with 11 M 4.95+ events, heavily influenced the results of the 5 yr experiment. Thus, these results reflect the variability in model performance that stems from a relatively limited number of prospective target earthquakes, so we strongly

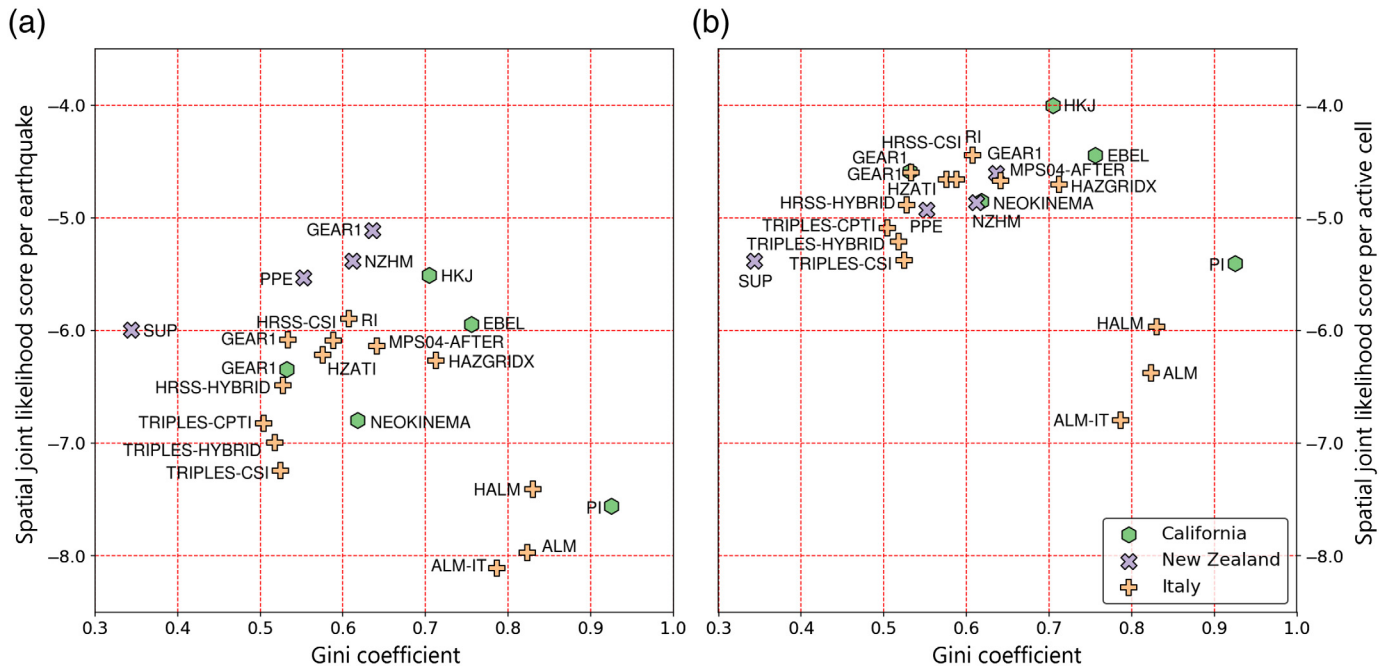


Figure 4. Spatial joint likelihood scores per (a) earthquake and (b) active cell obtained by global and regional models for California (green hexagones), New Zealand (purple crosses), and Italy (orange pluses). These scores are based on a Poisson and binary likelihood function, respectively. The higher the probability score, the more consistent the model is with the spatial distribution of observed epicenters. We include calculated Gini coefficients for each seismicity model to assess the dependence of model performance on spatial smoothing. Low Gini coefficients indicate spatially distributed forecasted seismicity, whereas high Gini coefficients denote the opposite. The spatially uniform Poisson SUP model has a nonzero G_c , because its earthquake rate densities vary with latitude.

recommend further prospective evaluations in these regions to investigate the timescale over which test results are reasonably stable.

Nonetheless, the prospective performance of GEAR1 is surprising, considering that California, New Zealand, and Italy are some of the best-instrumented and studied earthquake-prone regions in the world; so one might expect regionally calibrated models to be better informed and therefore more informative than GEAR1. Yet, GEAR1, using global datasets compiled no earlier than 1977, performs just as well (e.g., California and Italy) and sometimes better (e.g., New Zealand) than regional models, raising questions around which regional datasets make forecast models truly informative, and to what extent other factors, such as spatial smoothing or calibration period, add predictive skill. Overall, our comparative test results indicate that smoothing the locations of small earthquakes and combining geodetic data with smoothed $M 5.767+$ seismicity in a multiplicative fashion are two useful methods for forecasting moderate-to-large ($M 4.95+$) earthquakes in California, New Zealand, and Italy over a period of eight years. Furthermore, our analysis of the spatial dimension of the models shows that the regional models that provide forecasts that are too smooth (low G_c 's) or too localized (high G_c 's) perform poorly (see Fig. 4). The spatially uniform Poisson (SUP) and TRIPLE_S models, for example, indicate when more smoothing leads to systematic information loss, whereas the PI and ALM models, based on great specificity, illustrate overconfident localized forecasts. In

contrast, models with G_c 's between ~ 0.6 and 0.75 , such as HKJ, GEAR1 (New Zealand), and NZHM, obtain the highest likelihood scores, showing the range of smoothing in which spatial forecasting, in the context of time-independent forecasts over a period of eight years, appears optimal. Thus, we recommend earthquake forecasters consider “intermediate” smoothing procedures, for example, adaptive smoothing, to add predictive skill to their time-invariant seismicity models.

Finally, we note that the 47 earthquakes in New Zealand appear to be spatially more predictable than the 38 target events in California, and the 11 prospective earthquakes in Italy, respectively, as the New Zealand models obtain the highest spatial probability scores per earthquake (see Fig. 4a). In fact, some of the target earthquakes in New Zealand nucleated in the Cook Strait region in 2016, only a few years after the occurrence of three $M 5.7+$ earthquakes (Hamling *et al.*,

2014). Interestingly, this pattern is not observed when estimating spatial-likelihood scores per active cell, as the California and Italian models appear to “benefit” more than the New Zealand models from the reduced sensitivity of the binary S-test to clustering (see Fig. 4b). These results may stem from temporary fluctuations in seismic activity, as various target earthquakes in New Zealand nucleated on known faults, but not many on-fault events occurred in California and Italy during the testing period. Alternatively, these observations may be due to fundamental regional differences, likely related to the frequency and localization of earthquakes in each tectonic setting (e.g., Petruccelli *et al.*, 2019).

Conclusions

We evaluate and compare the performance of regional and global time-independent seismicity models for California, New Zealand, and Italy, using eight years of prospective data. Our comparative test results show that a regional model based on adaptive smoothed $M 2+$ seismicity is the most informative in California, whereas GEAR1, based on global datasets of $M 5.767+$ earthquakes and interseismic strain rates, ranks second. In New Zealand, GEAR1 outperforms all the models, including the former 2002 New Zealand National Seismic Hazard Model of Stirling *et al.* (2002), mainly due to the incorporation of seismicity information that is not present in the regional models. In Italy, a regional model based on relative seismic intensities and a model using adaptively smoothed $M 3+$ seismicity are the most informative, whereas the globally calibrated model, ranking third, can be considered as informative as the model informing the Italian National Seismic Hazard Model of 2004 (i.e., MPS Working Group, 2004). The performance of GEAR1 is unexpected, considering the comparatively lower resolution of its global training datasets. Thus, these results suggest that the adaptive smoothing of small earthquake locations and the multiplicative blend of geodetic data with smoothed seismicity are useful methods for forecasting the occurrence of moderate earthquakes in these regions over a period of nearly a decade.

Upon analyzing differences in model performance, we find that some regional models appear to underfit or overfit their calibration datasets, which may result in uninformative and overconfident forecasts, respectively. Specifically, models that are too smooth or too localized perform poorly, whereas models that apply “intermediate” smoothing procedures, such as adaptive smoothing, show better performance in forecasting the spatial distribution of observed seismicity. Hence, we recommend earthquake forecasters consider adaptive smoothing

techniques to add spatial predictive skill to their time-invariant models. In addition, we find that the global and regional models for New Zealand obtain higher spatial-likelihood scores per earthquake than the California and Italy models, respectively, meaning that some target events in New Zealand, past and prospective, occurred in similar locations. These observations might be due to temporal variations in on-fault and off-fault seismicity, or could be related to the frequency and spatial localization of earthquakes that are characteristic of each tectonic environment.

Acknowledging that the number of $M 4.95+$ earthquakes in the evaluation datasets is limited, and that earthquake clustering is likely to influence the model ranking, these prospective test results, together with the results of prospective global earthquake forecasting experiments (e.g., Strader *et al.*, 2018; Bayona *et al.*, 2021), provide preliminary support for the use of GEAR1 as a reference model forecasting moderate-to-large seismicity in California, New Zealand, Italy, and presumably elsewhere around the globe over an eight-year period. GEAR1 could be used as a benchmark against seismicity models in future CSEP forecast experiments in these and other testing regions, or against earthquake models informing global and regional seismic hazard assessments, including the 2020 European Seismic Hazard Model (Danciu *et al.*, 2021) and the 2018 Global Earthquake Model (Pagani *et al.*, 2020). All forecasts, data, and tests in this article are openly available for anyone to do so (see Data and Resources).

Data and Resources

The Comprehensive Earthquake Catalog of the Advanced National Seismic System (ANSS; Guy *et al.*, 2015) is available at <https://earthquake.usgs.gov/earthquakes/search/>, the GeoNet catalog (Ristau, 2013) is available at <https://www.geonet.org.nz/earthquake>, and the Italian seismic bulletin (Amato *et al.*, 2006) can be found at <http://terremoti.ingv.it/en/bsi>. The forecast files are openly accessible on Zenodo at <https://zenodo.org/record/7116221>, whereas the reproducibility package for this article, which also contains consistency test results for all earthquake forecasting models, can be found on GitHub at https://github.com/bayonato89/reproducibility_global_vs_regional. The pyCSEP software is freely available on GitHub at <https://github.com/SCECcode/pycsep>, and the documentation can be found at <https://cseptest.org/>. Two additional figures showing prospective number test results for all the models and residuals between spatial log-likelihood scores obtained by the Global Earthquake Activity Rate (GEAR1) model in New Zealand

and New Zealand Hazard Model (NZHM), and between the seismicity model component of GEAR1 and NZHM, are provided in the supplemental material to this article. All websites were last accessed in January 2023.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The authors sincerely thank editor Keith D. Koper and two anonymous reviewers for their thorough reading and constructive comments to improve the quality of this work. The authors also thank David Rhoades for valuable comments that helped them improve this article. José A. Bayona, Maximilian J. Werner, Danijel Schorlemmer, and Warner Marzocchi were supported by the European Union H2020 program (Grant Number 821115, Real-time earthquake risk reduction for a resilient Europe [RISE]). Maximilian J. Werner also received support from the Natural Environment Research Council (NERC; NE/R017956/1, “EQUIPT4RISK”). In addition, this research was supported by the Southern California Earthquake Center (SCEC, Contribution Number 12706). The SCEC is funded by the National Science Foundation (NSF) Cooperative Agreement EAR-1600087 and U.S. Geological Survey (USGS) Cooperative Agreement G17AC00047.

References

- Amato, A., L. Badiali, M. Cattaneo, A. Delladio, F. Doumaz, and F. Mele (2006). The real-time earthquake monitoring system in Italy, *Geosciences-Revue du BRGM*, available at <http://hdl.handle.net/2122/2574> (last accessed January 2023).
- Bayona, J., W. Savran, D. Rhoades, and M. Werner (2022). Prospective evaluation of multiplicative hybrid earthquake forecasting models in California, *Geophys. J. Int.* **229**, no. 3, 1736–1753, doi: [10.1093/gji/ggac018](https://doi.org/10.1093/gji/ggac018).
- Bayona, J., W. Savran, A. Strader, S. Hainzl, F. Cotton, and D. Schorlemmer (2021). Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information, *Geophys. J. Int.* **224**, no. 3, 1945–1955, doi: [10.1093/gji/ggaa554](https://doi.org/10.1093/gji/ggaa554).
- Bird, P., D. Jackson, Y. Kagan, C. Kreemer, and R. Stein (2015). GEAR1: A global earthquake activity rate model constructed from geodetic strain rates and smoothed seismicity, *Bull. Seismol. Soc. Am.* **105**, no. 5, 2538–2554, doi: [10.1785/0120150058](https://doi.org/10.1785/0120150058).
- Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). Classification and regression trees, *Wadsworth Int. Group* **37**, no. 15, 237–251, doi: [10.1785/0120150058](https://doi.org/10.1785/0120150058).
- Danciu, L., S. Nandan, C. Reyes, R. Basili, G. Weatherill, C. Beauval, A. Rovida, S. Vilanova, K. Şeşetyan, P. Bard, and F. Cotton (2021). The 2020 update of the European seismic hazard model: Model overview, *EFEHR Technical Rept. 001, v1. 0.0*, available at <http://hdl.handle.net/2122/15520> (last accessed January 2023).
- Ekström, G., M. Nettles, and A. Dziewoński (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. In.* **200**, 1–9, doi: [10.1016/j.pepi.2012.04.002](https://doi.org/10.1016/j.pepi.2012.04.002).
- Field, E. (2007). Overview of the working group for the development of regional earthquake likelihood models (RELM), *Seismol. Res. Lett.* **78**, no. 1, 7–16, doi: [10.1785/gssrl.78.1.7](https://doi.org/10.1785/gssrl.78.1.7).
- Field, E., T. Jordan, M. Page, K. Milner, B. Shaw, T. Dawson, G. Biasi, T. Parsons, J. Hardebeck, A. Michael, and R. Weldon (2017). A synoptic view of the third Uniform California Earthquake Rupture Forecast (UCERF3), *Seismol. Res. Lett.* **88**, no. 5, 1259–1267, doi: [10.1785/0220170045](https://doi.org/10.1785/0220170045).
- Gerstenberger, M., and D. Rhoades (2010). New Zealand earthquake forecast testing centre, in *Seismogenesis and Earthquake Forecasting: The Frank Evison*, M. K. Savage, E. G. C. Smith, D. A. Rhoades, M. C. Gerstenberger, and D. Vere-Jones (Editors), Vol. II, 23–38, Springer, Basel, Switzerland, doi: [10.1007/978-3-0346-0500-7_3](https://doi.org/10.1007/978-3-0346-0500-7_3).
- Guy, M., J. Patton, J. Fee, M. Hearne, E. Martinez, D. Ketchum, C. Worden, V. Quitoriano, E. Hunter, G. Smoczyk, and S. Schwarz (2015). National earthquake information center systems overview and integration, *U.S. Department of the Interior, US Geol. Surv.*, Reston, Virginia, doi: [10.3133/ofr20151120](https://doi.org/10.3133/ofr20151120).
- Hamling, I., E. D’Anastasio, L. Wallace, S. Ellis, M. Motagh, S. Samsonov, N. Palmer, and S. Hreinsdóttir (2014). Crustal deformation and stress transfer during a propagating earthquake sequence: The 2013 Cook Strait sequence, central New Zealand, *J. Geophys. Res.* **119**, no. 7, 6080–6092, doi: [10.1002/2014JB011084](https://doi.org/10.1002/2014JB011084).
- Kafka, A., and S. Levin (2000). Does the spatial distribution of smaller earthquakes delineate areas where larger earthquakes are likely to occur? *Bull. Seismol. Soc. Am.* **90**, no. 3, 724–738, doi: [10.1785/0119990017](https://doi.org/10.1785/0119990017).
- Marzocchi, W., D. Schorlemmer, and S. Wiemer (2010). Preface to the special volume “An earthquake forecast experiment in Italy”, *Ann. Geophys.* **53**, no. 3, III–VIII, doi: [10.4401/ag-4851](https://doi.org/10.4401/ag-4851).
- Michael, A., and M. Werner (2018). Preface to the focus section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New results and future directions, *Seismol. Res. Lett.* **89**, no. 4, 1226–1228, doi: [10.1785/0220180161](https://doi.org/10.1785/0220180161).
- Molchan, G., and V. Keilis-Borok (2008). Earthquake prediction: Probabilistic aspect, *Geophys. J. Int.* **173**, no. 3, 1012–1017, doi: [10.1111/j.1365-246X.2008.03785.x](https://doi.org/10.1111/j.1365-246X.2008.03785.x).
- MPS Working Group (2004). Redazione della Mappa di Pericolosità Sismica prevista dall’Ordinanza PCM 3274 del 20 marzo 2003, *Rapporto conclusivo per il Dipartimento della Protezione Civile, INGV, Milano-Roma, aprile 2004* (MPS04), 65–70, available at <http://zonesismiche.mi.ingv.it> (last accessed November 2022) (in Italian).
- Pagani, M., J. Garcia-Pelaez, R. Gee, K. Johnson, V. Poggi, V. Silva, M. Simionato, R. Styron, D. Viganò, L. Danciu, and D. Monelli (2020).

- The 2018 version of the Global Earthquake Model: Hazard component, *Earthq. Spectra* **36**, no. 1_suppl, 226–251, doi: [10.1177/8755293020931866](https://doi.org/10.1177/8755293020931866).
- Petruccioli, A., D. Schorlemmer, T. Tormann, A. 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**, 115791, doi: [10.1016/j.epsl.2019.115791](https://doi.org/10.1016/j.epsl.2019.115791).
- Rhoades, D., D. Schorlemmer, M. Gerstenberger, A. Christophersen, J. Zechar, and M. Imoto (2011). Efficient testing of earthquake forecasting models, *Acta Geophys.* **59**, no. 4, 728–747, doi: [10.2478/s11600-011-0013-5](https://doi.org/10.2478/s11600-011-0013-5).
- Ristau, J. (2013). Update of regional moment tensor analysis for earthquakes in New Zealand and adjacent offshore regions, *Bull. Seismol. Soc. Am.* **103**, no. 4, 2520–2533, doi: [10.1785/0120120339](https://doi.org/10.1785/0120120339).
- Savran, W., J. Bayona, P. Iturrieta, K. Asim, H. Bao, K. Bayliss, M. Herrmann, D. Schorlemmer, P. Maechling, and M. Werner (2022). PyCSEP: A python toolkit for earthquake forecast developers, *Seismol. Res. Lett.* **93**, no. 5, 2858–2870, doi: [10.1785/0220220033](https://doi.org/10.1785/0220220033).
- Savran, W., M. Werner, D. Schorlemmer, and P. Maechling (2022). PyCSEP: A python toolkit for earthquake forecast developers. *J. Open Source Softw.* **7**, no. 69, 3658, doi: [10.21105/joss.03658](https://doi.org/10.21105/joss.03658).
- Schorlemmer, D., A. Christophersen, A. Rovida, F. Mele, M. Stucchi, and W. Marzocchi (2010). Setting up an earthquake forecast experiment in Italy, *Ann. Geophys.* doi: [10.4401/ag-4844](https://doi.org/10.4401/ag-4844).
- Stirling, M., G. Verry, and K. Berryman (2002). A new seismic hazard model for New Zealand, *Bull. Seismol. Soc. Am.* **92**, no. 5, 1878–1903, doi: [10.1785/0120010156](https://doi.org/10.1785/0120010156).
- Strader, A., M. Werner, J. Bayona, P. Maechling, F. Silva, M. Liukis, and D. Schorlemmer (2018). Prospective evaluation of global earthquake forecast models: 2 yrs of observations provide preliminary support for merging smoothed seismicity with geodetic strain rates, *Seismol. Res. Lett.* **89**, no. 4, 1262–1271, doi: [10.1785/0220180051](https://doi.org/10.1785/0220180051).
- 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](https://doi.org/10.1785/0220180031).
- Tsuruoka, H., N. Hirata, D. Schorlemmer, F. Euchner, K. Nanjo, and T. Jordan (2012). CSEP Testing center and the first results of the earthquake forecast testing experiment in Japan, *Earth Planets Space* **64**, no. 8, 661–671, doi: [10.5047/eps.2012.06.007](https://doi.org/10.5047/eps.2012.06.007).
- Zechar, J., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. Maechling, and T. Jordan (2010). The collaboratory for the study of earthquake predictability perspective on computational earthquake science, *Concurrency Comput.* **22**, no. 2, 1836–1847, doi: [10.1002/cpe.1519](https://doi.org/10.1002/cpe.1519).
- Zeng, Y., M. Petersen, and Z. Shen (2018). Earthquake potential in California-Nevada implied by correlation of strain rate and seismicity, *Geophys. Res. Lett.* **45**, no. 4, 1778–1785, doi: [10.1002/2017GL075967](https://doi.org/10.1002/2017GL075967).

Manuscript received 7 February 2023

Published online 6 April 2023