

Bayona, J. A., Savran, W., Strader, A., Hainzl, S., Cotton, F., Schorlemmer, D. (2021): Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information. - Geophysical Journal International, 224, 3, 1945-1955.

<https://doi.org/10.1093/gji/ggaa554>

Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information

J.A. Bayona,^{1,2,*} W. Savran,³ A. Strader,¹ S. Hainzl,^{1,2} F. Cotton^{1,2} and D. Schorlemmer¹

¹GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. E-mail: jose.bayona@bristol.ac.uk

²Institute of Geosciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany

³Southern California Earthquake Center, University of Southern California, 90089-0742 Los Angeles, CA, USA

Accepted 2020 November 8. Received 2020 October 13; in original form 2020 May 20

SUMMARY

Global seismicity models provide scientific hypotheses about the rate, location and magnitude of future earthquakes to occur worldwide. Given the aleatory variability of earthquake activity and epistemic uncertainties in seismicity forecasting, the veracity of these hypotheses can only be confirmed or rejected after prospective forecast evaluation. In this study, we present the construction of and test results for two updated global earthquake models, aimed at providing mean estimates of shallow ($d \leq 70$ km) seismicity for seismic hazard assessment. These approaches, referred to as the Tectonic Earthquake Activity Model (TEAM) and the World Hybrid Earthquake Estimates based on Likelihood scores (WHEEL) model, use the Subduction Megathrust Earthquake Rate Forecast (SMERF2), an earthquake-rate model for subduction zones constrained by geodetic strain measurements and earthquake-catalogue information. Thus, these global ensemble seismicity models capture two independent components necessary for long-term earthquake forecasting, namely interseismic crustal strain accumulation and sudden lithospheric stress release. The calibration period for TEAM and WHEEL extends from 1977 January 1 to 2013 December 31. Accordingly, we use $m \geq 5.95$ earthquakes recorded during the 2014–2019 period to pseudo-prospectively evaluate the forecasting skills of these earthquake models, and statistically compare their performances to that of the Global Earthquake Activity Rate (GEAR1) model. As a result, GEAR1 and WHEEL are the most informative global seismicity models during the pseudo-prospective test period, as both rank with the highest information scores among all participant earthquake-rate forecasts. Nonetheless, further prospective evaluations are required to more accurately assess the abilities of these global ensemble seismicity models to forecast long-term earthquake activity.

Key words: Probabilistic forecasting; Earthquake interaction, forecasting and prediction; Seismicity and tectonics; Statistical seismology.

1 INTRODUCTION

Until recently, there has been some skepticism within the earthquake and engineering community about the capacity of earthquake-rate models to reliably characterize seismicity patterns. This lack of credulity primarily stems from the strongly stochastic nature of earthquakes and the relatively limited disposal of target data to independently calibrate and test seismicity models (Werner *et al.* 2011). Accordingly, only prospective forecast evaluations are considered rigorous enough to objectively describe the forecasting skills of

any seismicity model (Taroni *et al.* 2014). However, a drawback of prospective forecast tests is the time that one might have to wait to sample a statistically representative number of large events, given the usual level of global activity. Hence, multiple efforts have been made to reduce prospective forecast evaluation periods by expanding earthquake forecasting areas, which has led to the generation of global seismicity models. As a result, global $m \geq 5.8$ earthquake forecasts offer prospective test results that can be obtained in only 1–8 yr (Bird *et al.* 2015).

Short- and long-term global earthquake forecasts rely on homogenized seismicity catalogues, which currently provide hypocentral locations and magnitude estimates for thousands of earthquakes around the world. The Kagan–Jackson smoothed-seismicity (KJSS;

* Now at: School of Earth Sciences, University of Bristol, Queens Road, BS8 1QU Bristol, United Kingdom

Kagan & Jackson (2011) model, for instances, averages out the exceedance rates of 30 000 earthquakes reported in the 1977–2008 global centroid moment tensor Project (CMT; Dziewonski *et al.* 1981; Ekström *et al.* 2012) and the preliminary determinations of epicentres (PDE) seismicity catalogues to estimate global earthquake activity. Moreover, this model smoothes the location of each tectonic event with an anisotropic kernel function and uses a tapered Gutenberg–Richter relation to compute the spatial and magnitude distributions of forecasted seismicity. Thus, KJSS provides earthquake rate densities per unit area, time and magnitude that depend exclusively on instrumentally recorded seismicity.

The increasing availability of geodetic measurements along plate boundaries over the last two decades has enabled the creation of high-resolution models of plate motion and strain rate, which can be translated in present day into global maps of predicted seismicity. The Seismic Hazard Inferred from Tectonics based on the Global Strain Rate Map (SHIFT_GSRM, Kreemer *et al.* 2003; Bird & Liu 2007; Bird *et al.* 2010) model converts interseismic strain rates into long-term rates of seismic moment using global ‘geodesy-to-seismicity’ parameters, such as fault-dipping angles, coupled seismogenic thicknesses and elastic shear moduli (i.e. Bird & Kagan, 2004). Depending on the coupling, or the fraction of geodetic moment to be seismically released, seismicity rates are thereupon obtained by assuming a tapered Gutenberg–Richter frequency–magnitude distribution, which varies spatially among different tectonic regimes. According to retrospective test results for the 1977–2009 evaluation period, the raw SHIFT_GSRM forecast is consistent with the observations in continental convergent boundaries. Nevertheless, it significantly underpredicts earthquake rates in subduction zones, presumably due to geometrical effects inappropriately captured by the global average subduction dip angle employed in the formulation. Thus, the SHIFT_GSRM authors empirically corrected the forecast as the last step in constructing this seismicity model.

A few years later, Bird & Kreemer (2015) presented a revised version of SHIFT_GSRM: the SHIFT_GSRM2f earthquake model. SHIFT_GSRM2f is based on the updated Global Strain Rate Map (GSRM2.1; Kreemer *et al.* 2014), a global velocity gradient tensor field and continuous strain-rate model obtained from 22 415 GPS interseismic velocities. Among other improvements, SHIFT_GSRM2f incorporates a spatial smoothing of model strain rates around off-shore plate boundaries to provide high-resolution estimates of earthquake activity. However, the uncorrected SHIFT_GSRM2f forecast underestimates global earthquake rates during the 2005–2012 retrospective evaluation period, mainly due to underpredictions of subduction-zone seismicity. As a result, the SHIFT_GSRM2f modellers applied empirical calibration factors to improve the forecast in subduction zones by assuming that discrepancies between this new earthquake-rate forecast and the observations are primarily derived from the absence of specific fault-dipping angles in the moment-rate balance equation.

SHIFT_GSRM2f was then combined with KJSS to generate the Global Earthquake Activity Rate (GEAR1; Bird *et al.* 2015) model. Specifically, GEAR1 results from a multiplicative log-linear blend of smoothed-earthquake data and geodetic strain rates, expressed as:

$$H_{ij} = c \cdot \max[(S_{ij}^d \cdot T_{ij}^{1-d}), f]. \quad (1)$$

In this equation, H_{ij} represents the hybrid seismicity forecast in each grid cell centred on longitude i and latitude j , c is a normalization factor adjusting the global predicted earthquake rate

to the global rate of observed seismicity according to the 1977–2004 global CMT catalogue, \max stands for the maximum value, S_{ij} denotes the smoothed seismicity KJSS parent forecast, d is an optimization parameter to be determined, T_{ij} refers to the tectonic SHIFT_GSRM2f earthquake component and f symbolizes a baseline seismicity rate defined as the $\min[\min(S_{ij}), \min(T_{ij})]$. The optimized combination between parent forecast components for GEAR1 was determined by maximizing the I_1 (success) information score of Kagan (2009) from actual 2005–2012 seismicity. In this manner, Bird *et al.* (2015) found that the most informative GEAR1 forecast is derived from a multiplicative combination of parent components, with exponent $d = 0.6$ on KJSS (Seismicity) and $1 - d = 0.4$ on SHIFT_GSRM2f (Tectonics). Based on this, the GEAR1 modellers enhanced the preferred hybrid seismicity model by recomputing the Seismicity and Tectonics parent forecasts using the complete 1977–2013 global CMT catalogue. As a result, the update of such a preferred hybrid earthquake-rate model is what they refer to as the GEAR1 seismicity forecast.

GEAR1 and its individual components were submitted to the Collaboratory for the Study of Earthquake Predictability (CSEP; Jordan, 2006; Schorlemmer & Gerstenberger 2007; Zechar *et al.* 2010) testing centre for independent evaluation. According to 2-yr prospective test results, GEAR1 significantly outperforms both of its parent forecasts, providing preliminary support to combine geodetic strain rates with smoothed-seismicity data for long-term earthquake forecasting (Strader *et al.* 2018). In particular, the total earthquake number, spatial and magnitude distributions forecasted by GEAR1 were all consistent with observed seismicity. Nonetheless, both geodetic-based models SHIFT_GSRM and SHIFT_GSRM2f failed the spatial evaluation, in spite of the high spatial strain-rate resolution offered by GSRM and GSRM2.1.

During the 20th century, subduction interface seismicity released almost 90 per cent of the global seismic moment rate, and comprised approximately 60 per cent of the earthquake activity observed worldwide (Pacheco & Sykes 1992; Bird *et al.* 2010). Based on this, Bayona-Viveros *et al.* (2019) constructed the Subduction Megathrust Earthquake Rate Forecast (SMERF) as a complementary approach to the raw SHIFT_GSRM and SHIFT_GSRM2f forecasts to improve their forecasted number, and potentially spatial, distributions in subduction zones. Similar to SHIFT_GSRM2f, SMERF computes long-term budgets of seismic moment from interseismic strain rates and instrumentally recorded seismicity. Nevertheless, SMERF uses regional—not global—geodesy-to-seismicity conversion parameters to account for the great diversity of earthquake patterns among subduction margins (i.e. Heuret *et al.* 2011; Kagan & Jackson 2016). As a result, SMERF holds the initial capacity to properly forecast subduction-zone seismicity, according to retrospective and pseudo-prospective number-test results for the 1977–2014 and 2015–2018 evaluation periods. A limitation of SMERF, however, is the relatively large number of seismicity parameters it needs to be constructed and the limited availability of earthquake-catalogue information to individually calibrate them. Hence, we present in this study, a revised version of SMERF that depends on a reduced amount of earthquake parameters, referred to as the SMERF2 model. Moreover, we integrate SMERF2 estimates in subduction zones with SHIFT_GSRM2f computations outside of these tectonic convergent margins to produce the global Tectonic Earthquake Activity Model (TEAM). Furthermore, we combine KJSS with this updated tectonic earthquake model to create the World Hybrid Earthquake Estimates based on Likelihood scores (WHEEL) model; an alternative hybrid seismicity approach to GEAR1. Finally, we use $d \leq 70$ km, $m \geq 5.95$ earthquakes recorded during the

Table 1. Mean estimates of seismic coupling coefficient c_H for 14 (Kagan & Jackson 2016) subduction interfaces.

Number	Heuret's subduction zone	KJ's subduction zone	c_H
1	Northern Tonga	Kermadec-Tonga-Samoa	0.12 [0.10–0.14]
	Southern Tonga		
	Northern Kermadec		
	Southern Kermadec		
2	Southern New-Hebrides	New Hebrides Is.	0.82 [0.56–1.00]
	D'Entrecasteux		
	Northern New-Hebrides		
3	Solomon	Bismarck-Solomon Is.	0.56 [0.38–0.83]
	Bougainville		
	New Britain		
4	Java	Sunda Arc	0.16 [0.11–0.22]
5	Sumatra	Andaman Is.-Sumatra	0.16 [0.12–0.21]
	Andaman		
6	Taiwan-Manila	Taiwan	0.21 [0.15–0.30]
7	Southern Ryukyu	S.E. Japan-Ryukyu Is.	0.26 [0.18–0.37]
	Northern Ryukyu		
	Nankai		
8	Tohoku	Japan-Kamchatka	0.47 [0.33–0.66]
	Southern Kuril		
	Northern Kuril		
	Kamchatka		
9	Marianas	Guam-Japan	0.10 [0.09–0.12]
	Izu-Bonin		
10	Southwestern Aleutians	Alaska-Aleutian Arc	0.36 [0.26–0.50]
	Central Aleutians		
	Eastern Aleutians		
	Western Alaska		
	Eastern Alaska		
11	Mexico	Mexico-Guatemala	0.47 [0.32–0.70]
	Cocos		
12	Costa Rica	Central America	0.21 [0.15–0.30]
	Colombia-Ecuador		
13	Antilles	Caribbean Loop	0.17 [0.11–0.24]
14	Northern Peru	Andean S.America	0.79 [0.52–1.00]
	Southern Peru		
	Northern Chile		
	Southern Chile		

Notes: Based on eq. (14) in Bayona-Viveros *et al.* (2019), we empirically calibrate these values to construct the SMERF2 seismicity model; In bold, we present mean estimates of seismic coupling coefficient c_H ; In square brackets, we report uncertainties associated with c_H computations.

2014–2019 period to pseudo-prospectively evaluate the consistency of these global ensemble seismicity models with the observations, and quantitatively compare their forecasting skills with GEAR1, our benchmark global earthquake-rate model.

2 NEW SEISMICITY MODELS

2.1 The updated Subduction Megathrust Earthquake Rate Forecast

One of the most basic ways that SMERF differs from SHIFT_GSRM2f is that SMERF identifies 37 subduction margins, according to variations of seismicity parameters (i.e. Heuret *et al.* 2011; Kagan & Jackson 2016). In contrast, SHIFT_GSRM2f distinguishes one subduction zone from other tectonic plate boundaries based on discontinuities of relative plate velocity (i.e. Bird *et al.* 2009). Thus, SMERF possesses a relatively large number

of degrees of freedom, specifically regional seismic coupling coefficients, to be individually constrained by available earthquake-catalogue data. Therefore, we present a revised version of SMERF that depends on a reduced number of seismicity parameters, referred to as the SMERF2 model. Same as SMERF, SMERF2 uses the trench segmentation model of Heuret *et al.* (2011) to assign geometric parameters, such as seismogenic thicknesses, trench lengths and fault-dipping angles to 37 subduction areas. Nonetheless, this updated model version employs only one average elastic shear modulus $\bar{\mu}_s = 68$ GPa for subduction interfaces, derived from the results of Bayona-Viveros *et al.* (2019). Moreover, SMERF2 utilizes the subduction segmentation model of Kagan & Jackson (2016) to re-calibrate mean regional seismic coupling coefficients c_H from estimates of corner magnitude computed for each subduction zone in such a study. For this aim, we group together some of the Heuret *et al.* (2011) subduction zones (e.g. Northern Peru, Southern Peru, Northern Chile and Southern Chile) to relate them to the Kagan & Jackson (2016) subduction segments (e.g. Andean S. America). As

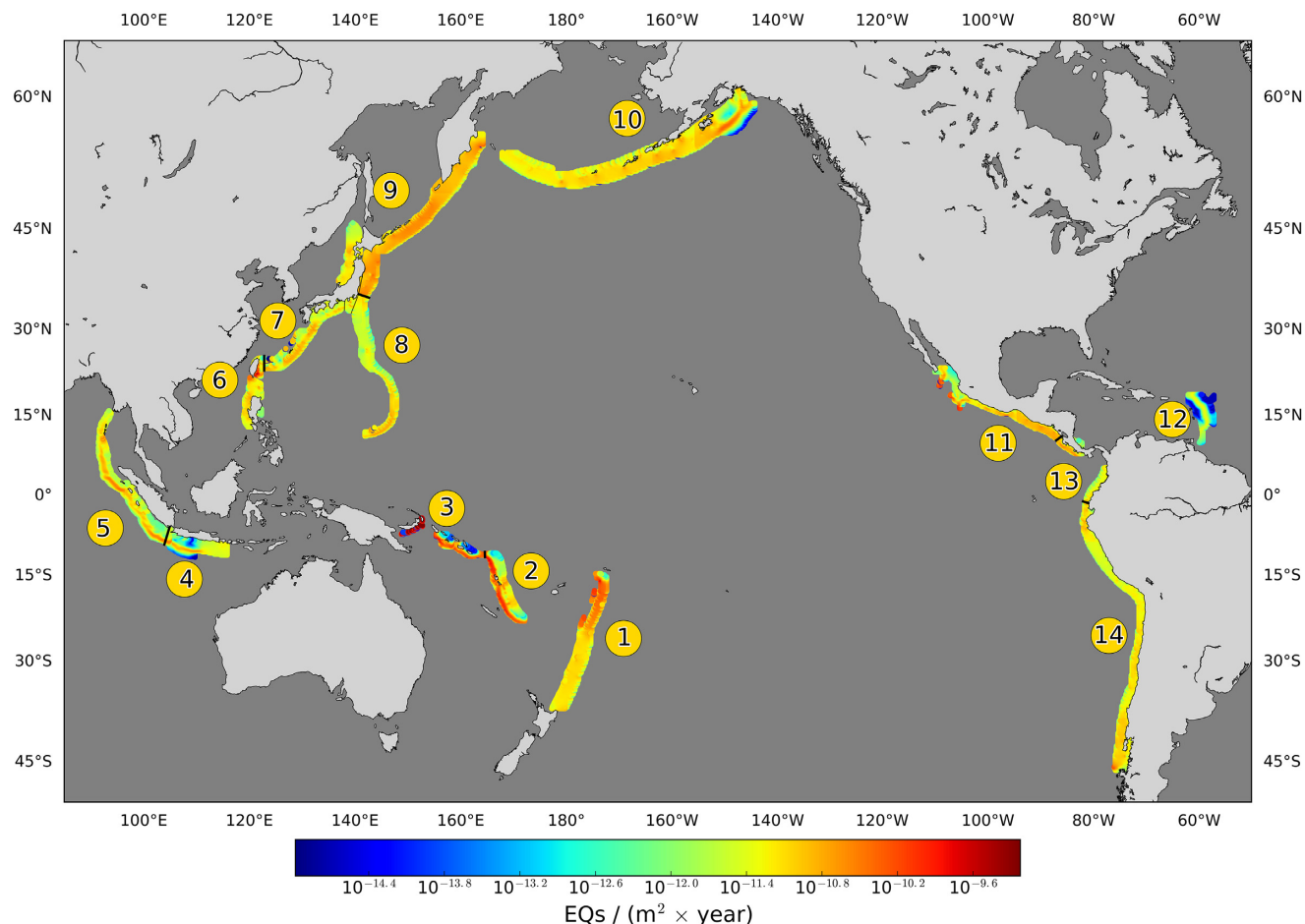


Figure 1. Mean annual density rates of shallow $d \leq 70$ km, $m \geq 5.95$ interface seismicity, provided by the updated SMERF2.

as a result, SMERF2 reduces the number of mean seismic coupling coefficients from 37 to 14 (see Table 1) to provide high-resolution estimates of subduction-zone seismicity (see Fig. 1).

2.2 The Tectonic Earthquake Activity Model

Strictly speaking, SHIFT_GSRM2f and SMERF2 are hybrid seismicity models, because both unavoidably employ earthquake information to translate geodetic strain rates into high-resolution estimates of shallow seismicity. Particularly, the predicted number of earthquakes is calibrated from observed seismicity, and the spatial earthquake distribution is provided by interseismic strain measurements. Nonetheless, we assume SHIFT_GSRM2f and SMERF2 in this work to be geodetic-based earthquake models. Thus, we combine SMERF2 computations of subduction-zone seismicity with SHIFT_GSRM2f estimates of earthquake activity outside of the SMERF2 subduction margins to generate the global TEAM (see Figs 2 and 3). To achieve it, we first preserve the annual earthquake rate for subduction zones predicted by SMERF2, as seismic coupling coefficients c_H employed in its formulation are empirically calibrated from actual 1977–2013 interface seismicity (see eq. 14 in Bayona-Viveros *et al.* 2019). Then, we multiply SHIFT_GSRM2f earthquake rate densities in non-SMERF2 subduction areas by a factor of approximately 0.8 to adjust the mean global rate to approximately 176 $m \geq 5.767$ earthquakes per year, provided by the 1977–2013 global CMT catalogue.

2.3 The World Hybrid Earthquake Estimates based on Likelihood scores model

We finally blend the new TEAM seismicity model with KJSS to create the WHEEL model. Same as GEAR1, this hybrid earthquake-rate model uses a multiplicative log-linear combination of earthquake parent components, with exponent $d = 0.6$ on the Seismicity forecast (KJSS) and $1 - d = 0.4$ on the Tectonic constituent (now TEAM). As a result, WHEEL serves as an alternative approach to GEAR1 to compute global earthquake activity, with special focus on subduction-zone seismicity (see Figs 4 and 5).

3 EARTHQUAKE MODEL EVALUATIONS

SHIFT_GSRM2f, KJSS, GEAR1, TEAM and WHEEL provide scientific hypotheses about where and how frequently earthquakes can occur (see Table 2). Nonetheless, the veracity of these hypotheses can only be confirmed or rejected after prospective forecast experiments. In this study, we evaluate the initial consistency of these global seismicity forecasts with observations recorded during the 2014 January 1–2019 December 31 pseudo-prospective test period. Moreover, we statistically compare the performance of each participant global earthquake approach with GEAR1, our selected benchmark seismicity model. To do so, we first divide the global study region into spatio-magnitude bins with increments of 0.1 units in longitude, latitude and magnitude. Within each bin, we then specify expected numbers of earthquakes during the evaluation period,

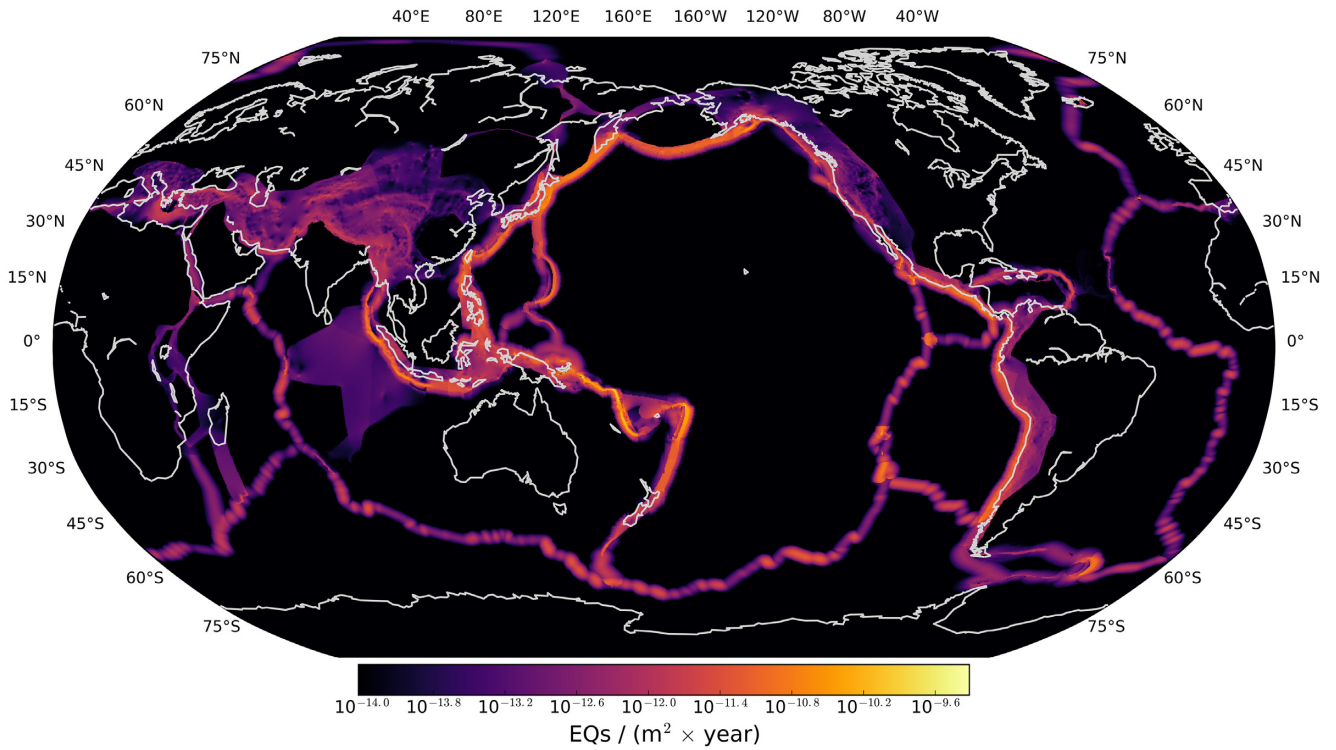


Figure 2. Forecast map showing annual $d \leq 70$ km, $m \geq 5.95$ earthquake rate densities (per m²), derived from the global TEAM.

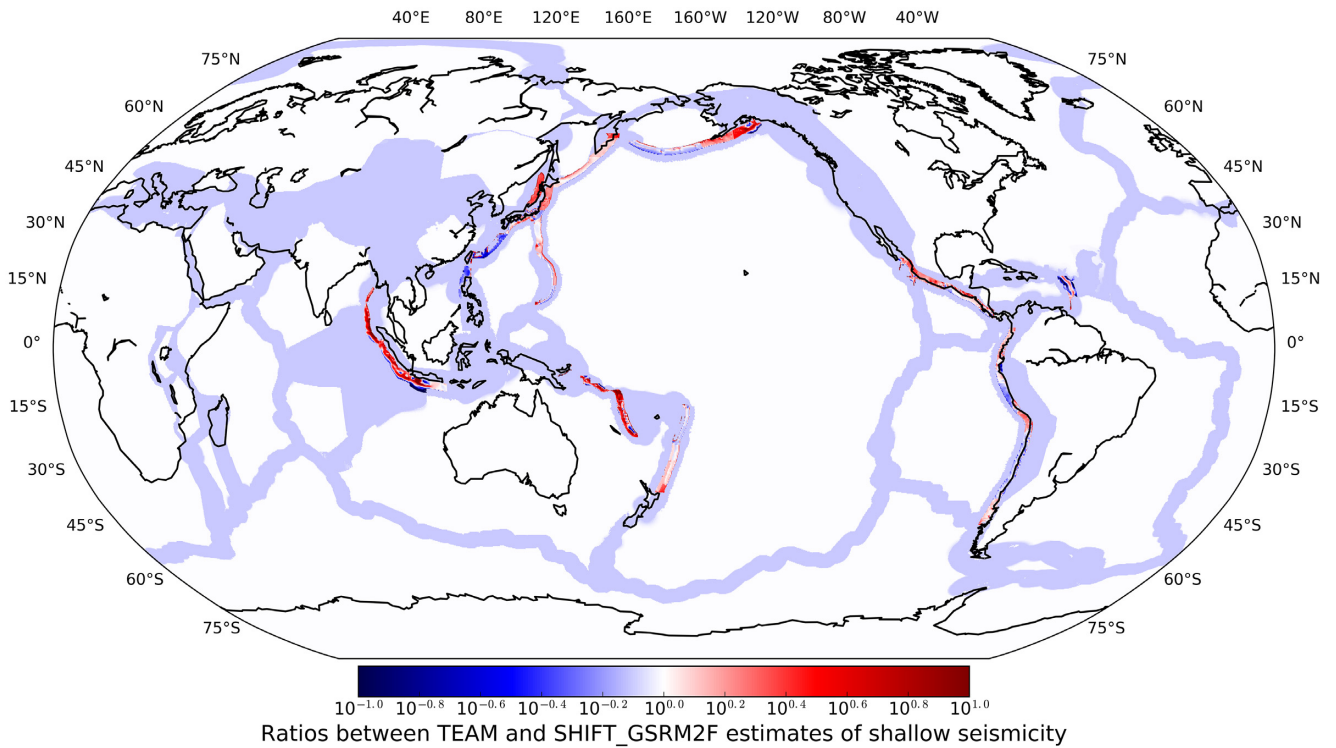


Figure 3. Variations in $m \geq 5.95$ annual earthquake rate densities forecasted by TEAM and SHIFT_GSRM2f. In red regions, TEAM estimates a larger number of earthquakes per year than SHIFT_GSRM2f. On the contrary, blue points denote locations where SHIFT_GSRM2f computes larger seismicity rates than TEAM.

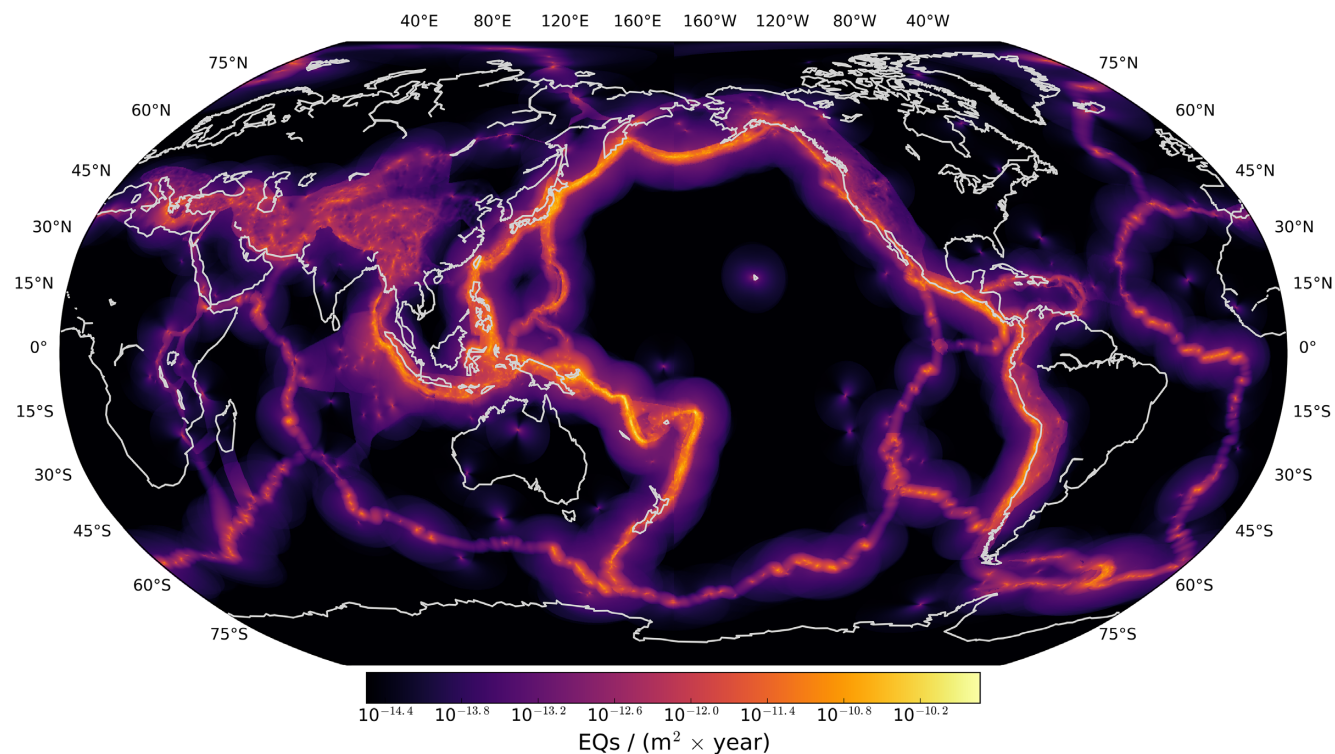


Figure 4. Mean estimates of annual $d \leq 70$ km, $m \geq 5.95$ earthquake rate densities (per m^2), computed by the WHEEL model.

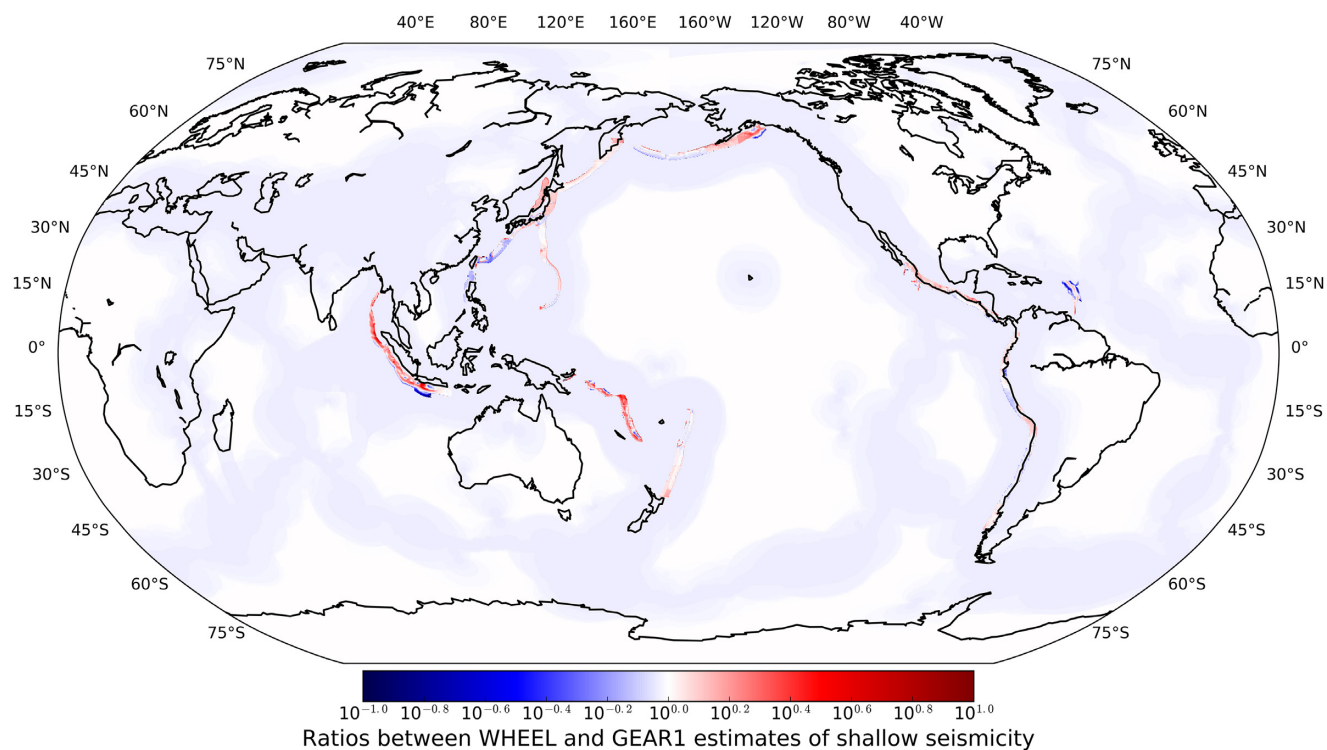


Figure 5. Variations in $m \geq 5.95$ annual seismicity rates forecasted by the WHEEL and GEAR1 earthquake models. In red, WHEEL estimates a larger number of earthquakes than GEAR1. In blue, oppositely, GEAR1 computes larger seismicity rates than WHEEL.

which are assumed to be Poisson distributed (Schorlemmer *et al.* 2007; Zechar, Gerstenberger & Rhoades 2010). Also, we implement the comparative and consistency tests described below to assess the forecasting capacities of contestant global seismicity models.

3.1 Comparative tests

We evaluate the relative performance of forecast pairs by measuring the rate-corrected information gain per earthquake of one forecast over another (Rhoades *et al.* 2011). For each global forecast pair,

Table 2. Summary of seismicity models described in this study. In bold, we highlight global earthquake-rate forecasts undergoing pseudo-prospective evaluation. We do not include SHIFT_GSRM, as SHIFT_GSRM2f is its successor earthquake-rate model.

Model	Nature	Coverage	Based on	Reference
SHIFT_GSRM	Tectonic	Global	GSRM and CMT catalogue	Bird <i>et al.</i> (2010)
KJSS	Seismicity	Global	CMT catalogue	Kagan & Jackson (2011)
SHIFT_GSRM2f	Tectonic	Global	GSRM2.1 and CMT catalogue	Bird & Kreemer (2015)
GEAR1	Hybrid	Global	KJSS and SHIFT_GSRM2f	Bird <i>et al.</i> (2015)
SMERF	Tectonic	Subduction zones	GSRM2.1 and CMT catalogue	Bayona-Viveros <i>et al.</i> (2019)
SMERF2	Tectonic	Subduction zones	GSRM2.1 and CMT catalogue	This issue
TEAM	Tectonic	Global	SMERF2 and SHIFT_GSRM2f	This issue
WHEEL	Hybrid	Global	KJSS and TEAM	This issue

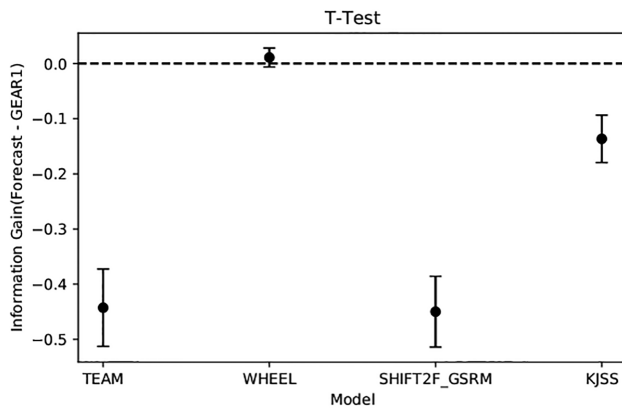


Figure 6. Comparative T -test results for contestant global seismicity models during the 2014–2019 pseudo-prospective evaluation period. We present the mean information gain per earthquake as circles, and the 95 percent confidence interval with vertical lines. GEAR1 and WHEEL are the most informative models among all participant earthquake-rate approaches, as both rank with the highest information gain scores. Although not displayed, W -test results corroborate all T -tests.

we apply the Student's paired T -test. This comparative test is based on the null hypothesis that two forecasts perform similarly, and the alternate hypothesis that one forecast significantly outperforms the other. As a result, one forecast is considered more informative than the other if the mean information gain significantly differs from the scaled difference in forecasted earthquake numbers between two forecasts.

If the number of target earthquakes is large, the T -test relies on the approximate normality of the mean of the information gain scores at observed earthquake locations, which is guaranteed by the central limit theorem. In contrast, for small a number of target earthquakes, the Student's paired T -test requires the assumption that the information gain scores at observed earthquake locations are normally distributed. In this case, the non-parametric W -test, evaluating the median information gain per earthquake rather than the mean, needs to be applied to strengthen a T -test result. This test only requires that the information gain distribution is symmetric, and increases in power with increasing numbers of observed earthquakes (Rhoades *et al.* 2011).

3.2 Consistency tests

Using a suite of likelihood consistency tests, we evaluate the consistency of forecasted and observed seismicity during the experiment's testing period. These tests are based on the likelihood of observed seismicity patterns, given forecasted earthquake numbers. A forecast's log-likelihood score is a metric, based on the Poisson

distribution, used to evaluate the consistency of forecasted seismicity patterns with observed earthquakes (Schorlemmer *et al.* 2010). Greater log-likelihood scores indicate greater consistency, corresponding to a higher probability of the forecast generating a seismicity distribution similar to observations. This in turn implies a greater ability of the seismicity model to forecast earthquakes.

Consistency with observed seismicity, expressed as the log-likelihood score, can be decomposed into three dimensions: number of earthquakes, magnitude and spatial distributions. So, we apply tests of consistency for each of these dimensions (N -, M - and S -tests, respectively), which are directly derived from the likelihood, or L -test (Zechar *et al.* 2010). A forecast's log-likelihood score is most impacted by the number of earthquakes. Therefore, we also apply the conditional likelihood (CL) test, which provides information about a forecast's spatial and magnitude distribution while removing information regarding the total number of earthquakes (Werner *et al.* 2011).

4 RESULTS AND DISCUSSION

T -test results show that GEAR1 and WHEEL are the most informative seismicity forecasts during the 2014–2019 pseudo-prospective evaluation period, as they possess the highest information gain scores among all participant earthquake-rate forecasts (see Fig. 6 and Table 3). In agreement with Strader *et al.* (2018), the outperformance of these hybrid seismicity models over their individual forecast parent components indicate that the combination of instrumentally recorded seismicity and geodetic strain information is suitable for long-term earthquake modelling. As formerly explained, the multiplicative log-linear blend of seismicity forecasts is based on the maximization of the Kagan (2009) I_1 success information score during the 2005–2012 retrospective period. Thus, comparative T -test results support the selection of the preferred GEAR1 model over a similar test period, suggest its stability over time, as described by Bird (2018), and serve as new evidence to promote the creation and development of ensemble seismicity models, as concluded by Marzocchi *et al.* (2012) and Akinci *et al.* (2018).

Comparative T -test results also exhibit that KJSS exceeds the forecasting skills of TEAM and SHIFT_GSRM2f during the 6-yr pseudo-prospective assessment period. Additionally, we observe that TEAM and SHIFT_GSRM2f are equally informative during the same testing period, as the difference between their information gain scores is statistically insignificant. Nonetheless, these latter seismicity models might perform better than KJSS in the long term due to the incorporation of interseismic strain data into their formulations. Plate motion is assumed to lead to constant strain rates that can be estimated over fairly short interseismic time periods assuming stationary coupling properties (Haines & Holt 1993; Savage & Simpson 1997; Kreemer *et al.* 2014). The earthquake probabilities

Table 3. Pseudo-prospective N -, M -, S - and CL-test statistics for participant global seismicity models.

Model	δ_1	δ_2	κ	ζ	ξ	IG (over GEAR1)
GEAR1	0.96	0.04	0.55	1.00	1.00	0.00
WHEEL	0.96	0.04	0.57	0.98	1.00	0.01 [-0.02 – 0.04]
KJSS	0.96	0.04	0.50	0.98	1.00	-0.14 [-0.22 – -0.09]
SHIFT_GSRM2f	0.96	0.05	0.59	0.00	0.00	-0.45 [-0.55 – -0.32]
TEAM	0.96	0.05	0.62	0.00	0.00	-0.44 [-0.56 – -0.31]

Notes: The N -test metrics δ_1 and δ_2 describe the probabilities of observing at least and at most the actual number of earthquakes, respectively. If $\delta_1 < 0.025$, the forecast significantly underpredicts observed seismicity; if $\delta_2 < 0.025$, the forecast importantly overestimates actual seismicity; if $\delta_1 > 0.025$ and $\delta_2 > 0.025$ the forecast is consistent with the observations. The magnitude κ , spatial ζ , and magnitude-spatial ξ statistics provide the percentage of simulated seismicity catalogues, derived from each earthquake model, with a lower log-likelihood score than the score computed for the observed earthquake catalogue. At a 0.05 confidence level, if these metrics are greater than 0.025, the forecasted seismicity distributions are consistent with actual earthquake activity. Thus, we indicate in bold values inconsistencies of earthquake-rate forecasts with the observations. Finally, we display information gain scores per event IG and confidence intervals (in brackets) over GEAR1, obtained for each contestant earthquake model during the 2014–2019 pseudo-prospective test period.

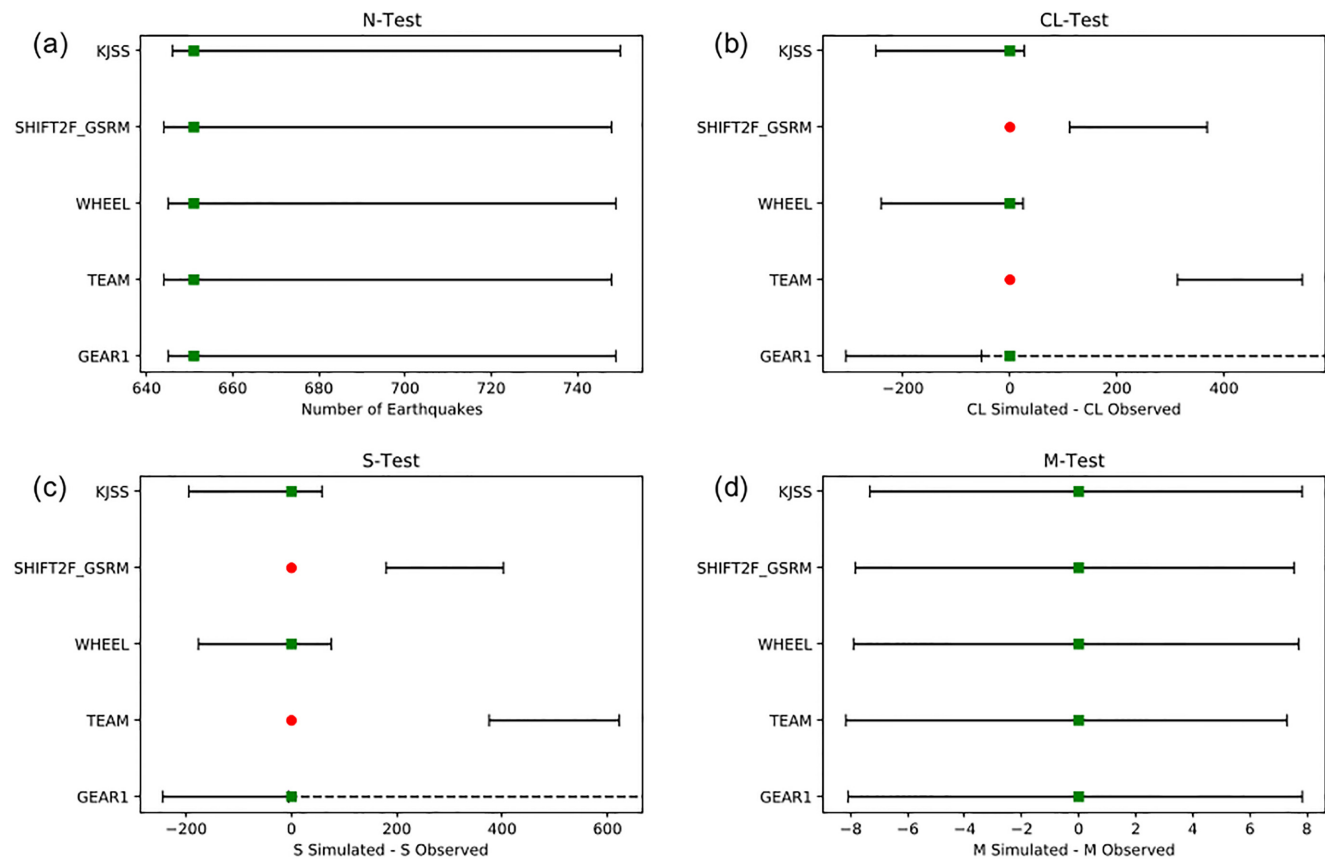


Figure 7. Consistency evaluation results for participant global earthquake-rate models during the 2014 January 1–2019 December 31 pseudo-prospective testing period. We show (a) N -, (b) CL-, (c) S - and (d) M -test results. Green squares indicate that earthquake forecasts are consistent with the observations. In contrast, red circles denote that models forecast earthquake patterns inconsistent with actual seismicity. Horizontal lines stand for the 95 per cent confidence intervals. Dashed lines represent the one-sided form of the test distribution. Thus, seismicity models with a quantile score greater than the distribution upper limit are not rejected. For the N -test, the x -axis exhibits the number of earthquakes. For all other evaluations, the x -axis displays log-likelihood scores.

are expected to be proportional to the stress rates which, in turn, can be related to the estimated strain rates. Thus, geodetic-based seismicity models like TEAM and SHIFT_GSRM2f can be expected to represent the long-term probabilities better than earthquake-based models like KJSS, which attempt to estimate probabilities based on relatively limited earthquake-catalogue information. Accordingly, we recommend further prospective evaluations to more accurately

describe the forecasting potential of these global geodetic-based earthquake models.

N -test results reveal that all global seismicity forecasts are consistent with observations recorded during the 2014–2019 evaluation period (see Fig. 7). This is important, as the N -test is the most powerful consistency test implemented in CSEP, but it is unsurprising because all contestant global forecasts were normalized to compute

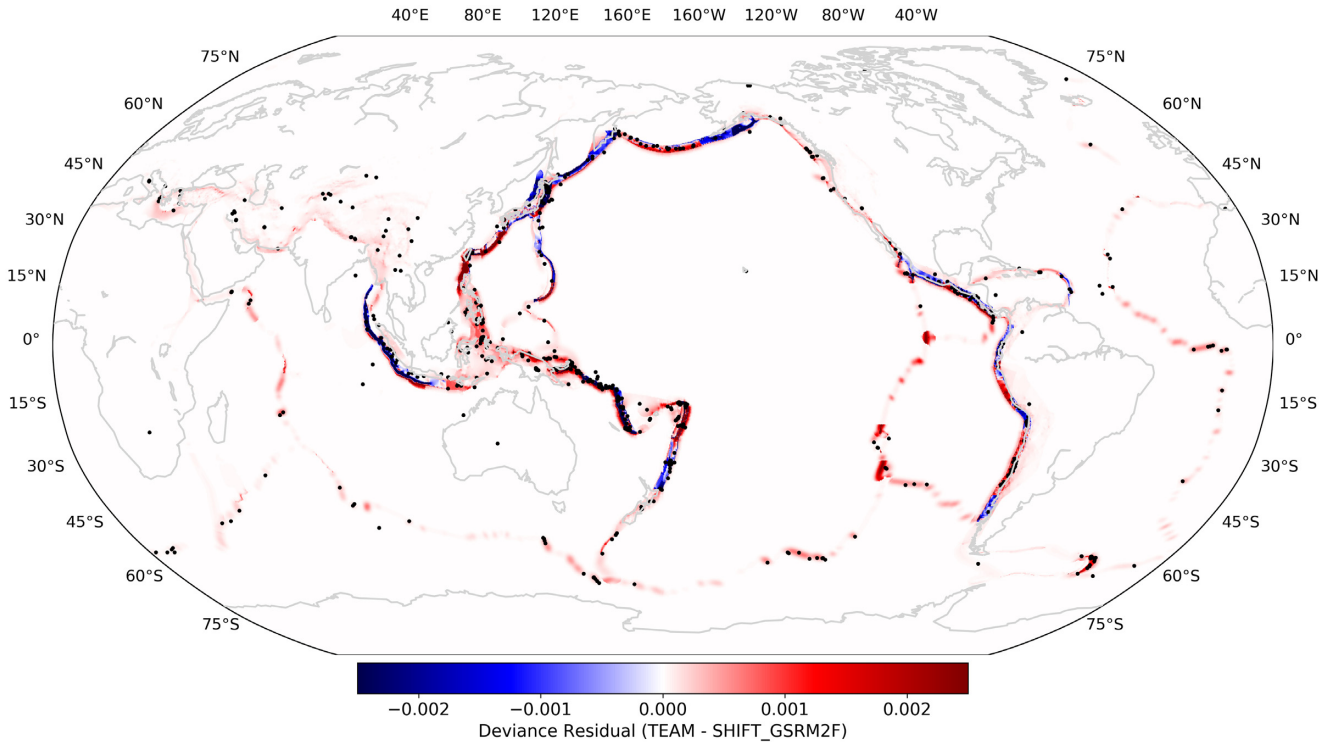


Figure 8. Residuals between spatial log-likelihood scores forecasted by TEAM and SHIFT_GSRM2f during the pseudo-prospective test period. In red areas, ratios between spatial likelihood scores obtained by these models are positive, indicating that TEAM is spatially more informative than SHIFT_GSRM2f. In contrast, blue zones denote locations where the spatial distribution forecasted by SHIFT_GSRM2f is more similar to the observations than the spatial distribution computed by TEAM. We include as black circles $d \leq 70$ km, $m_w \geq 5.95$ earthquakes reported from 2014 January 1 to 2019 December 31 in the global CMT catalogue.

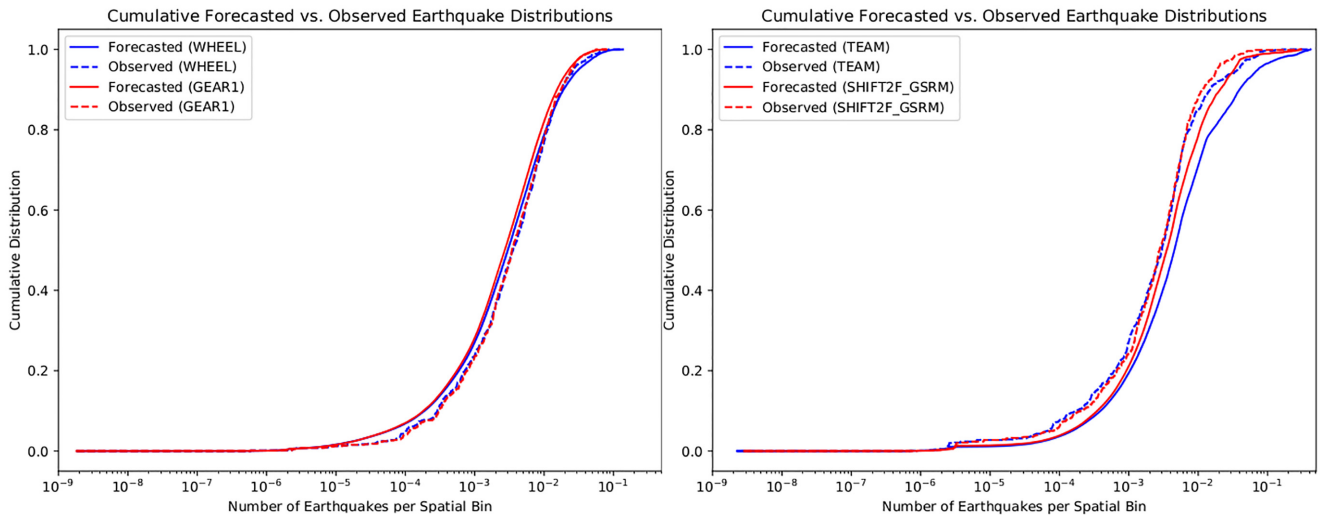


Figure 9. Concentration plots displaying the normalized cumulative distribution of observed seismicity (dashed lines), in comparison to that of the forecasted earthquake activity (solid lines), according to (left) the hybrid WHEEL (blue) and GEAR1 (red) models, and (right) the strain-based TEAM (blue) and SHIFT_GSRM2f (red) seismicity models.

the global annual seismicity rate provided by the 1977–2013 global CMT catalogue. In addition, we note that all contestant seismicity forecasts simultaneously pass the magnitude M -test during the pseudo-prospective evaluation period (see Fig. 7). These results are due to the fact that all models use unions of tapered Gutenberg–Richter functions to distribute their computed seismic moment rates among rates of forecasted seismicity. Distinctively, only the spatial earthquake distributions forecasted by WHEEL, GEAR1 and KJSS

are consistent with observed seismicity during the testing period at the 0.05 significance level, as both tectonic models, TEAM and SHIFT_GSRM2f, forecast spatial seismicity patterns inconsistent with the observations (see Fig. 7). As discussed in Strader *et al.* (2018) and Bayona-Viveros *et al.* (2019), SHIFT_GSRM2f spatial inconsistencies during the 2015–2017 period could be primarily due to initial underestimations of earthquake numbers in most subduction zones. Therefore, TEAM attempts to compensate for this issue

Table 4. Pseudo-prospective *N*- and *M*-test results for GEAR1 and WHEEL over 2-yr intervals, overlapping during the 2014–2019 period.

Evaluation start date	δ_1 (GEAR1)	δ_2 (GEAR1)	ζ (GEAR1)	δ_1 (WHEEL)	δ_2 (WHEEL)	ζ (WHEEL)
01-01-2014	0.56	0.47	1.00	0.56	0.47	0.99
01-01-2015	0.89	0.12	1.00	0.89	0.12	0.98
01-01-2016	0.99	0.01	0.69	0.99	0.01	0.40
01-01-2017	1.00	0.01	0.29	1.00	0.01	0.08
01-01-2018	0.73	0.29	0.98	0.73	0.29	0.91

Notes: These results show that *N*-test results are unstable during the 2016 January 1–2017 December 31 time frame, as GEAR1 and WHEEL overestimate rates of global seismicity (highlighted in bold).

by concentrating more of the forecasted seismicity density within subduction margins by applying localized geomechanical parameters, such as dip angles, corner magnitudes and coupling coefficients to these regions. Nevertheless, both tectonic forecasts obtain almost the same spatial likelihood score (−4858.75 for TEAM and −4857.09 for SHIFT_GSRM2f) during the testing period, despite substantial variations among their spatial earthquake distributions (see Fig. 8). Thus, these results suggest that spatial inconsistencies forecasted by TEAM and SHIFT_GSRM2f might rather be due to spatial undersmoothing or limitations of CSEP evaluations.

To better understand *S*-test results, we generate concentration plots for both hybrid and tectonic forecasts, showing their normalized cumulative observed and forecasted seismicity distributions (see Fig. 9). For the hybrid models, the predicted earthquake distributions are slightly shifted to the left of the actual seismicity distribution, which explains why the GEAR1 model ‘fails’ the two-sided *S*-test during the evaluation period (although the *S*-test is one-sided). For the strain-based models, the forecasted cumulative distributions are significantly shifted to the right of the observed earthquake distribution, with a greater difference observed for TEAM. This indicates that seismicity forecasted by TEAM and SHIFT_GSRM2f is too localized along plate boundaries compared to KJSS, GEAR1 and WHEEL. Consequently, if an earthquake occurs just outside of a subduction zone, the forecasted rate will abruptly decrease and cause a large drop in the *S*-likelihood score. As described above, CSEP tests rely on the assumption that earthquakes rates are independent and Poisson distributed. However, this assumption is particularly questionable in the presence of clusters of seismicity, as the Poisson distribution could be underdispersed with respect to the true earthquake distribution (Lombardi & Marzocchi 2010). Thus, *S*-test results may need to be revisited in future work using testing procedures that can more adequately capture the clustering nature of earthquakes.

Finally, a further pseudo-prospective analysis shows that the number and spatial test results for GEAR1 and WHEEL tend to remain stable over 2-yr intervals, overlapping during the 2014–2019 testing period (see Table 4). Interestingly, these forecasts overestimate global earthquake activity during the 2016 January 1–2017 December 31 period, despite the occurrence of intermediate-size tectonic events, such as the 2016 $m_w = 7.9$ Papua New Guinea, the 2016 $m_w = 7.8$ Solomon Islands and the 2017 $m_w = 8.2$ Mexico earthquakes. Nevertheless, we interpret these results to be due to temporal fluctuations in earthquake activity, which should be taken into account for time-dependent seismicity modelling.

5 CONCLUSION AND PROSPECTS

In this study, we describe the construction and pseudo-prospective evaluation of two updated global seismicity models, aimed at providing mean estimates of earthquake activity for seismic

hazard assessment. These approaches are based on SMERF2, an earthquake-rate model for subduction zones constrained by interseismic strain measurements and earthquake-catalogue information. TEAM is an ensembled tectonic seismicity model resulting from the combination of SMERF2 estimates in subduction margins with SHIFT_GSRM2f computations in non-SMERF2 interplate regions. WHEEL is a hybrid seismicity model created from the multiplicative blend of the tectonic parent TEAM and the KJSS seismicity forecast. Using actual earthquakes recorded during the 2014 January 1–2019 December 31 period, we pseudo-prospectively test the forecasting skills of these global ensemble earthquake models, and statistically compare their performances with respect to GEAR1. As a result, GEAR1 and WHEEL are the most informative seismicity forecasts during the pseudo-prospective test period, as they obtain the highest information gain scores among all contestant earthquake forecasts (see Fig. 6 and Table 3). Interestingly, the difference between their information gain scores is statistically so insignificant, that the two models can be considered equally informative. Thus, we strongly recommend further prospective evaluations of all contestant global seismicity models to objectively assess their capacities to forecast longer term earthquake activity.

The two global strain-based earthquake models are restricted to provide high-resolution estimates of shallow seismicity along tectonic plate boundaries. When compared, TEAM exceeds the spatial forecasting abilities of SHIFT_GSRM2f in subduction areas where both forecasted and observed earthquake rate densities are high (e.g. the western portion of the Aleutian arc, the coastline of central Chile and the New-Hebrides and Bismarck–Solomon subduction zones; see Fig. 8). In contrast, SHIFT_GSRM2f better characterizes spatial patterns than TEAM in interplate regions with low forecasted seismicity rates, hosting relatively few earthquakes during the testing period (e.g. the Andaman–Sumatra subduction margin, the middle American trench along southern Mexico, and the eastern sector of the Aleutian arc). We expect these results to converge towards a more stable solution in the future, as geodetic strain rates are thought to have longer term effects on earthquake forecasting than six years, given their intrinsic relationship with seismic coupling and plate motion. Accordingly, we will submit TEAM and WHEEL to the CSEP testing centre to assess the veracity of this hypothesis.

CSEP is currently transitioning into a new phase of earthquake forecast experiments designed to relax the Poisson assumption by accounting for the full distribution of the forecast (Nandan *et al.* 2019; Savran *et al.* 2020). The pseudo-prospective evaluation of earthquake-rate models presented in this study is the first testing exercise of this underconstruction project on a global scale. Until now, CSEP has been able to identify the most informative seismicity models among more than 400 participant earthquake approaches (Michael & Werner 2018; Schorlemmer *et al.* 2018). In this manner, seismicity models like GEAR1 and WHEEL have been proven to indeed possess the initial capacities to reliably characterize earthquake patterns worldwide, despite some skepticism.

6 DATA AND RESOURCES

The earthquake catalogue was obtained from the Global Centroid Moment Tensor (CMT; Dziewonski *et al.* 1981; Ekström *et al.* 2012) Project, available at www.globalcmt.org/CMTsearch.html (last accessed January 2020).

ACKNOWLEDGEMENTS

The authors genuinely thank the Editor Margarita Segou, the anonymous reviewers, and the reviewer Naoshi Hirata for their thorough reading and helpful comments in improving this manuscript. Additionally, we are grateful to David Jackson for enlightening discussions on the construction of these global ensemble seismicity models. Moreover, we warmly thank Peter Bird for providing valuable feedback to produce these models, as well as the KJSS and SHIFT_GSRM2f forecast grids to compare them with. This research was funded by the Consejo Nacional de Ciencia y Tecnología (CONACYT) in collaboration with the Deutscher Akademischer Austauschdienst (DAAD), under the funding program number 57177537.

REFERENCES

- Akinci, A., Moschetti, M.P. & Taroni, M., 2018. Ensemble smoothed seismicity models for the new Italian probabilistic seismic hazard map, *Seismol. Res. Lett.*, **89**(4), 1277–1287.
- Bayona-Viveros, J.A., von Specht, S., Strader, A., Hainzl, S., Cotton, F. & Schorlemmer, D., 2019. A regionalized seismicity model for subduction zones based on geodetic strain rates, geomechanical parameters, and earthquake-catalog data, *Bull. seism. Soc. Am.*, **109**(5), 2036–2049.
- Bird, P., 2018. Ranking some global forecasts with the Kagan information score, *Seismol. Res. Lett.*, **89**(4), 1272–1276.
- 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, *Bull. seism. Soc. Am.*, **105**(5), 2538–2554.
- Bird, P. & Kagan, Y.Y., 2004. Plate-tectonic analysis of shallow seismicity: apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. seism. Soc. Am.*, **94**(6), 2380–2399.
- Bird, P., Kagan, Y.Y., Jackson, D.D., Schoenberg, F.P. & Werner, M.J., 2009. Linear and nonlinear relations between relative plate velocity and seismicity, *Bull. seism. Soc. Am.*, **99**(6), 3097–3113.
- Bird, P. & Kreemer, C., 2015. Revised tectonic forecast of global shallow seismicity based on version 2.1 of the Global Strain Rate Map, *Bull. seism. Soc. Am.*, **105**(1), 152–166.
- Bird, P., Kreemer, C. & Holt, W.E., 2010. A long-term forecast of shallow seismicity based on the Global Strain Rate Map, *Seismol. Res. Lett.*, **81**(2), 184–194.
- Bird, P. & Liu, Z., 2007. Seismic hazard inferred from tectonics: California., *Seismol. Res. Lett.*, **78**(1), 37–48.
- Dziewonski, A.M., Chou, T.A. & Woodhouse, J.H., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.: Solid Earth*, **86**(B4), 2825–2852.
- Ekström, G., Nettles, M. & Dziewoński, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.*, **200**, 1–9.
- Haines, A.J. & Holt, W.E., 1993. A procedure for obtaining the complete horizontal motions within zones of distributed deformation from the inversion of strain rate data, *J. geophys. Res.: Solid Earth*, **98**(B7), 12057–12082.
- Heuret, A., Lallemand, S., Funicello, F., Piromallo, C. & Faccenna, C., 2011. Physical characteristics of subduction interface type seismogenic zones revisited, *Geochem. Geophys. Geosyst.*, **12**(1), 58–61.
- Jordan, T.H., 2006. Earthquake predictability, brick by brick, *Seismol. Res. Lett.*, **77**(1), 3–6.
- Kagan, Y.Y., 2009. Testing long-term earthquake forecasts: likelihood methods and error diagrams, *Geophys. J. Int.*, **177**(2), 532–542.
- Kagan, Y.Y. & Jackson, D.D., 2011. Global earthquake forecasts, *Geophys. J. Int.*, **184**(2), 759–776.
- Kagan, Y.Y. & Jackson, D.D., 2016. Earthquake rate and magnitude distributions of great earthquakes for use in global forecasts, *Geophys. J. Int.*, **206**(1), 630–643.
- Kreemer, C., Blewitt, G. & Klein, E.C., 2014. A geodetic plate motion and Global Strain Rate Model, *Geochem. Geophys. Geosyst.*, **15**(10), 3849–3889.
- Kreemer, C., Holt, W.E. & Haines, A.J., 2003. An integrated global model of present-day plate motions and plate boundary deformation, *Geophys. J. Int.*, **154**(1), 8–34.
- Lombardi, A.M. & Marzocchi, W., 2010. The assumption of Poisson seismic-rate variability in CSEP/RELM experiments, *Bull. seism. Soc. Am.*, **100**, 15A(8):2293–2300.
- Marzocchi, W., Zechar, J.D. & Jordan, T.H., 2012. Bayesian forecast evaluation and ensemble earthquake forecasting, *Bull. seism. Soc. Am.*, **102**(6), 2574–2584.
- 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, *Seismol. Res. Lett.*, **89**(4), 1226–1228.
- Nandan, S., Ouillon, G., Sornette, D. & Wiemer, S., 2019. Forecasting the full distribution of earthquake numbers is Fair, robust, and better, *Seismol. Res. Lett.*, **90**(4), 1650–1659.
- Pacheco, J.F. & Sykes, L.R., 1992. Seismic moment catalog of large shallow earthquakes, 1900 to 1989, *Bull. seism. Soc. Am.*, **82**(3), 1306–1349.
- Rhoades, D.A., Schorlemmer, D., Gerstenberger, M.C., Christophersen, A., Zechar, J.D. & Imoto, M., 2011. Efficient testing of earthquake forecasting models, *Acta Geophys.*, **4**(59), 1306–1349.
- Savage, J.C. & Simpson, R.W., 1997. Surface strain accumulation and the seismic moment tensor, *Bull. seism. Soc. Am.*, **87**(5), 1345–1353.
- Savran, W.H., Werner, M.J., Marzocchi, W., Rhoades, D.A., Jackson, D.D., Milner, K., Field, E. & Michael, A., 2020. Pseudoprospective evaluation of UCERF3–ETAS forecasts during the 2019 Ridgecrest Sequen, *Bull. seism. Soc. Am.*, **110**(4), 1799–1817.
- Schorlemmer, D. & Gerstenberger, M.C., 2007. RELM testing center, *Seismol. Res. Lett.*, **78**(1), 30–36.
- 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.
- Schorlemmer, D., Zechar, J.D., Werner, M.J., Field, E.H., Jackson, D.D. & Jordan, T.H., RELM Working Group, 2010. First results of the regional earthquake likelihood models experiment, in *Seismogenesis and Earthquake Forecasting: The Frank Evison Volume II*, pp. 5–22, Springer, Basel.
- Schorlemmer, D. *et al.*, 2018. The collaboratory for the study of earthquake predictability: achievements and priorities, *Seismol. Res. Lett.*, **89**(4), 1305–1313.
- 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, *Seismol. Res. Lett.*, **89**(4), 1262–1271.
- Taroni, M., Zechar, J.D. & Marzocchi, W., 2014. Assessing annual global M 6+ seismicity forecasts, *Geophys. J. Int.*, **196**(1), 422–431.
- Werner, M.J., Helmstetter, A., Jackson, D.D. & Kagan, Y.Y., 2011. High-resolution long-term and short-term earthquake forecasts for California, *Bull. seism. Soc. Am.*, **101**(4), 1630–1648.
- Zechar, J.D., Gerstenberger, M.C. & Rhoades, D.A., 2010. Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. seism. Soc. Am.*, **100**(3), 1184–1195.
- 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.: Pract. Exper.*, **22**(12), 1836–1847.