Deliverable 7.2

D7.2: Report on First Results of Key Hypothesis Testing

Deliverable information	
Work package	[WP7: Testing]
Lead	[GFZ Potsdam]
Authors	[Max Werner (University of Bristol), Danijel Schorlemmer (GFZ Potsdam)]
Reviewers	[Warner Marzocchi, University of Naples]
Approval	[Management Board]
Status	[Final]
Dissemination level	[Public]
Delivery deadline	[28.02.2022]
Submission date	[28.02.2022]
Intranet path	[DOCUMENTS/DELIVERABLES/]

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 reports on first results of tests of key hypotheses of seismogenesis, and one test of nonlinear ground motion amplification. Many hypotheses and models have been offered to control or explain the occurrence of earthquakes. For example: the seismic gap hypothesis (earthquakes are more likely in sections of seismic quiescence) remains embedded in media communications and some time-dependent seismic hazard models; the Coulomb-stress hypothesis remains the most popular mechanism for explaining aftershock occurrences; b-value variations (in time or space) contain information about impending earthquakes; and many more. To make progress on their validity and to exploit them for predictive purposes, RISE aimed to test some of these hypotheses, using both retrospective and prospective datasets and using, where possible, the framework of the Collaboratory for the Study of Earthquake Predictability (CSEP). This report provides eight snapshots of concluded and ongoing evaluations of various hypotheses, including the Coulomb-stress hypothesis, seismic gap theory, the interplay of aftershocks and the fault network, the role of afterslip in aftershock triggering, and predictive skill of nonlinear amplification terms in ground motion models. These studies have also led to advances in testing methodology and metrics.

1. Tests of the Elasto-Static Coulomb Stress Hypothesis

Tests on the 2019 Ridgecrest, California, earthquake sequence [Mancini et al., 2020]:

Operational earthquake forecasting protocols commonly use statistical models for their recognized ease of implementation and robustness in describing the short-term spatiotemporal patterns of triggered seismicity. However, recent advances on physics-based aftershock forecasting reveal comparable performance to the standard statistical counterparts with significantly improved predictive skills when fault and stress-field heterogeneities are considered. Mancini et al. (2020) performed a pseudoprospective forecasting experiment during the first month of the 2019 Ridgecrest (California) earthquake sequence. They developed seven Coulomb rate-and-state models that couple static stress-change estimates with continuum mechanics expressed by the rate-and-state friction laws. The model parameterization supports a gradually increasing complexity; they start from a preliminary model implementation with simplified slip distributions and spatially homogeneous receiver faults to reach an enhanced one featuring optimized fault constitutive parameters, finite-fault slip models, secondary triggering effects, and spatially heterogenous planes informed by pre-existing ruptures. The data-rich environment of southern California allows them to test whether incorporating data collected in near-real time during an unfolding earthquake sequence boosts our predictive power. They assess the absolute and relative performance of the forecasts by means of statistical tests used within the Collaboratory for the Study of Earthquake Predictability and compare their skills against a standard benchmark epidemic-type aftershock sequence (ETAS) model for the short (24 hr after the two Ridgecrest mainshocks) and intermediate terms (one month). Stress-based forecasts expect heightened rates along the whole near-fault region and increased expected seismicity rates in central Garlock fault. Their comparative model evaluation not only supports that faulting heterogeneities coupled with secondary triggering effects are the most critical success components behind physics-based forecasts, but also underlines the importance of model updates incorporating near-real-time available aftershock data reaching better performance than standard ETAS. In their publication, Mancini et al. explore the physical basis behind the results by investigating the localized shut down of pre-existing normal faults in the Ridgecrest near-source area.

Retrospective forecast model for Italy using the Coulomb-based rate-and-state framework [Cheng et al., 2022]:

During 2009-2014, the Collaboratory for the Study of Earthquake Predictability (CSEP) executed a state-wide rate-based forecast in Italy. Cheng et al. (2022) implement a retrospective study using the rate and-state framework and the Epidemic Type Aftershock Sequence (ETAS) method to forecast the spatiotemporal variation of earthquakes in a retrospective scenario. They test the hypothesis that an enhanced CRS framework involving improved source and fault characterization and model updates could improve the skill of forecasts on the Italy-wide scale for the 1-day interval. Cheng et al. (2022) also evaluate how our state-wide CRS models perform during specific earthquake sequences, namely the 2009 L'Aquila sequence and the 2012 Emilia sequence. The result indicates that adopting the finite slip models, spatially variable receiver faults, and including stress rearrangement from secondary triggering could increase the performance of the Italy-wide CRS forecast. Further developments will include using a multi-layer receiver model and testing earthquake datasets with lower completeness magnitude and more precise hypocentre location.

Publication outputs:

Mancini, Segou, Werner & Parsons (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

Cheng, Segou, Werner, Main & McCloskey (2022, in preparation): Retrospective state-wide forecast model using Coulomb based rate-and-state framework.

2. The Interplay of Aftershocks and the Fault Network

Savran et al. (2020) developed a tailored pseudo-prospective experiment to test the hypothesis that large supra-seismogenic aftershocks occur on (mapped) faults and control the overall aftershock patterns. The 2019 Ridgecrest sequence provided the first opportunity to evaluate the Uniform California Earthquake Rupture Forecast v.3 with epidemic-type aftershock sequences (UCERF3-ETAS) in a pseudoprospective sense. For comparison, they include a version of the model without explicit faults more closely mimicking traditional ETAS models (UCERF3-NoFaults). They evaluate the forecasts with new metrics developed within the Collaboratory for the Study of Earthquake Predictability (CSEP). The metrics consider synthetic catalogs simulated by the models rather than synoptic probability maps, thereby relaxing the Poisson assumption of previous CSEP tests. Their approach compares statistics from the synthetic catalogs directly against observations, providing a flexible approach that can account for dependencies and uncertainties encoded in the models. RISE developed bespoke software and testing methods that enabled comparing the simulated catalogs with the observed catalog, thereby circumventing the need for approximating likelihood functions. The testing methods were developed as part of the open-source community software toolkit PyCSEP, which is available from https://github.com/SCECcode/pycsep.

Savran et al. (2020) find that, to the first order, both UCERF3-ETAS and UCERF3-NoFaults approximately capture the spatiotemporal evolution of the Ridgecrest sequence, adding to the growing body of evidence that ETAS models can be informative forecasting tools. However, they also find that both models mildly overpredict the seismicity rate, on average, aggregated over the evaluation period. More severe testing indicates the overpredictions occur too often for observations to be statistically indistinguishable from the model. Magnitude tests indicate that the models do not include enough variability in forecasted magnitude-number distributions to match the data. Spatial tests highlight discrepancies between the forecasts and observations, but the greatest differences between the two models appear when aftershocks occur on modeled UCERF3-ETAS faults. Therefore, any predictability associated with embedding earthquake triggering on the (modeled) fault network may only crystalize during the presumably rare sequences with aftershocks on these faults. Accounting for uncertainty in the model parameters could improve test results during future experiments.

Publication outputs:

Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E. H. & Michael, A. (2020). Pseudoprospective Evaluation of UCERF3-ETAS Forecasts during the 2019 Ridgecrest Sequence. Bulletin of the Seismological Society of America, 110(4), 1799-1817. https://doi.org/10.1785/0120200026

3. Do Enhanced Seismicity Catalogs Improve Aftershock Forecasts?

Artificial intelligence methods are revolutionizing modern seismology by offering unprecedentedly rich seismic catalogs. Recent developments in short-term aftershock forecasting show that Coulomb rate-and-state (CRS) models hold the potential to achieve operational skills comparable to standard statistical Epidemic-Type Aftershock Sequence (ETAS) models, but only when the near real-time data quality allows to incorporate a more detailed representation of sources and receiver fault populations. In this framework, the high-resolution reconstructions of the seismicity patterns introduced by machine-learning-derived earthquake catalogs represent a unique opportunity to test whether they can be exploited to improve the predictive power of aftershock forecasts.

Mancini et al. (2022) present a retrospective forecast experiment on the first year of the 2016-2017 Central Italy seismic cascade, where seven M5.4+ earthquakes occurred between a few hours and five months after the initial Mw 6.0 event, migrating over a 60-km long normal fault system. As target dataset, they employ the best available high-density machine learning catalog recently released for the sequence, which reports ~1 million events in total (~22,000 with M ≥ 2).

First, they develop develop a CRS model featuring (1) rate-and-state variables optimized on 30 years of pre-sequence regional seismicity, (2) finite fault slip models for the seven mainshocks of the sequence, (3) spatially heterogeneous receivers informed by pre-existing faults, and (4) updating receiver fault populations using focal planes gradually revealed by aftershocks. The authors then test the effect of considering stress perturbations from the M2+ events. Using the same high-precision catalog, Mancini et al. produce a standard ETAS model to benchmark the stress-based counterparts. All models are developed on a 3D spatial grid with 2 km spacing; they are updated daily and seek to forecast the space-time occurrence of M2+ seismicity for a total forecast horizon of one year. Mancini et al. formally rank the forecasts with the statistical scoring metrics introduced by the CSEP and compare their performance to a generation of CRS and ETAS models previously published for the same sequence by Mancini et al. (2019), who used solely real-time data and a minimum triggering magnitude of M=3.

Mancini et al. find that considering secondary triggering effects from events down to M=2 slightly improves model performance. While this result highlights the importance of better seismic catalogs to model local triggering mechanisms, it also suggests that to appreciate their full potential future modelling efforts will likely have to incorporate also fine-scale rupture characterizations (e.g., smaller source fault geometries retrieved from enhanced focal mechanism catalogs) and introduce denser spatial model discretizations.

Publication outputs:

Mancini, Segou, Werner & Parsons (2022, in preparation): Do Enhanced Seismicity Catalogs Improve Aftershock Forecasts? A Test on the 2016-2017 Central Italy Earthquake Cascade.

4. A Test of the Seismic Gap Hypothesis

The seismic gap hypothesis has a long and controversial history, but continues to be popular and is frequently cited in the media. In particular, the seismic gap hypothesis has been widely cited in Mexico to predict the location of future earthquakes and to assess seismic hazard, specifically in the context of the so-called 'Guerrero gap'. However, no analysis of the outcome of any predictions of the hypothesis in Mexico has been done to-date. Husker, Bayona, Werner and Santoyo are preparing a manuscript that analyzes the outcome of the formal seismic gap prediction by Nishenko and Singh (1987). The prediction has well-defined probabilities, areas and timeframes that allow for its evaluation. Those timeframes were 5 years, 10 years and 20 years after 1986. The prediction relies on the precise repeat times of characteristic earthquakes to define segments, but the catalog that the authors use relies on an imprecise definition of characteristic earthquakes. Husker et al. discuss some of their decisions in building their catalog to explain how they analyze the outcome of the prediction. They create catalogs of earthquakes based on the probabilities of earthquake occurrence for each segment. They also generate null model earthquake catalogs using the average number of earthquakes that occur in the subduction zone, and randomly distribute these along the distance of the segments. They find that null model performed better than the seismic gap hypothesis prediction. The prediction over the longest time frame of 20 years correctly predicted the outcome in only 48% of the segments compared to 91% coinciding for the null model. The gap hypothesis also greatly over predicted the total number of segments with a characteristic earthquake. Ms ≥ 7.4 earthquakes were predicted to occur in 6 of the 11 segments over the 20-year timeframe, but only 1 actually occurred. That lone earthquake was a Mw 8.0 which occurred in a segment with a 0% chance of an earthquake in one of their models and 16% change in another. Husker et al. conclude that the gap hypothesis did not perform well at predicting earthquakes in Mexico and, in fact, its predictions were worse than predicting earthquakes by chance. There is thus no evidence to suggest earthquakes are overdue in the Guerrero gap, and therefore Husker et al. recommend taking special care in invoking the gap hypothesis to communicate earthquake hazards in Mexico.

Publication Outputs:

Husker, Bayona, Werner & Santoyo (2022, in preparation): A Test of the Earthquake Gap Hypothesis in Mexico: the case of the Guerrero Gap.

5. Does Aseismic Afterslip Control Aftershock Productivity?

Understanding the controls on aftershock triggering is key to skillful operational earthquake forecasting and short-term hazard assessment. Many studies suggest that aseismic afterslip plays a key role in driving aftershock sequences, often citing strong correlations in their spatio-temporal evolutions. Churchill, Werner, Fagereng and Biggs (2022a) previously showed that the amount of afterslip produced after an earthquake can vary greatly, from <1% to >300% of the coseismic moment. Thus, afterslip could feasibly account for some of the spatio-temporal complexity many aftershock sequences exhibit, which coseismic Coulomb static stress change alone struggles to explain. If this link is robustly established, including afterslip in frameworks such as ETAS (which currently assumes that every earthquake triggers aftershocks in a statistically identical way) may improve their predictive capabilities.

Churchill et al. (2022b) explore correlations between the relative afterslip moment of compiled earthquakes and the relative productivity, relative cumulative moment, b-value and Omori decay exponent (p) of the corresponding aftershock sequences. They select sequences from the global PDE catalog (MC=4.7) for comparability across different tectonic regions using three methods for robustness: 1) a 2D method based on empirical aftershock-zone scaling; 2) a 3D method using a volume around the ruptured fault planes that afterslip may reasonably activate via static Coulomb stress; and 3) a Nearest Neighbour Distance declustering algorithm. These methods select similar proportions of aftershocks for a given mainshock. Across different mainshocks, variation in the relative productivity and relative cumulative moment of sequences correlates weakly with relative afterslip moment, but b-value and p do not. It is surprising that relative afterslip moment does not correlate with aftershock decay, but they cannot rule out that afterslip decays do not co-vary with aftershock decays globally. However, Churchill et al. (2022b) cannot provide strong evidence on these global, statistical scales that afterslip drives aftershocks, but recognise that additional probing of decay behaviours is necessary.

Publication outputs:

Churchill, Werner, Fagereng & Biggs (2022a, under review): Afterslip Moment Scaling and Variability from a Global Compilation of Estimates, under review in the Journal of Geophysical Research: Solid Earth.

Churchill, Werner, Fagereng & Biggs (2022b, in preparation): Does Abundant Afterslip mean Productive Aftershock Sequences?

6. Do Hybrid Models Achieve Greater Predictive Skill?

The Regional Earthquake Likelihood Models (RELM) experiment, conducted within the Collaboratory for the Study of Earthquake Predictability (CSEP), showed that the smoothed seismicity (HKJ) model by Helmstetter et al. (2007) was the most informative time-independent earthquake model in California during the 2006–2010 evaluation period. The diversity of competing forecast hypotheses and geophysical datasets used in RELM was suitable for combining multiple models that could provide more informative earthquake forecasts than HKJ. Thus, Rhoades et al. (2014) created multiplicative hybrid models that involve the HKJ model as a baseline and one or more conjugate models. In retrospective evaluations, some hybrid models showed significant information gains over the HKJ forecast. Bayona et al. (2022) prospectively assess the predictive skills of 16 hybrids and 6 original RELM forecasts at a 0.05 significance level, using a suite of traditional and new CSEP tests that rely on a Poisson and a binary likelihood function. In addition, they include consistency test results at a Bonferroni-adjusted significance level of 0.025 to address the problem of multiple tests. Furthermore, they compare the performance of each forecast to that of HKJ. The evaluation dataset contains 40 target events recorded within the CSEP California testing region from 1 January 2011 to 31 December 2020, including the 2016 Hawthorne earthquake swarm in southwestern Nevada and the 2019 Ridgecrest sequence. Consistency test results show that most forecasting models overestimate the number of earthquakes and struggle to explain the spatial distribution of epicenters, especially in the case of seismicity clusters. The binary likelihood function significantly reduces the sensitivity of spatial log-likelihood scores to clustering, however; most models still fail to adequately describe spatial earthquake patterns. Contrary to retrospective analyses, our prospective test results show that none of the models are significantly more informative than the HKJ benchmark forecast, which they interpret to be due to temporal instabilities in the fit that forms hybrids. These results suggest that smoothing high-resolution, small earthquake data remains a robust method for forecasting moderate-to-large earthquakes over a period of five to fifteen years in California.

Publication outputs:

Bayona, Savran, Rhoades & Werner (2022), Prospective evaluation of multiplicative hybrid earthquake forecasting models in California, Geophysical Journal International, ggac018, https://doi.org/10.1093/gji/ggac018

7. Optimal forecast resolution – Quadtree

The GFZ group is specifically testing how many earthquakes are necessary for meaningful forecasts, essentially addressing the forecast uncertainty problem. They developed a new approach to a multi-resolution grid that allows to use differently-sized cells based on the number of earthquakes available for estimating the forecast rates (or any other criterion). The underlying technology is the tile-based approach, known as the Quadtree, which is used in many different fields, e.g. Web-Map Services, and allows for easy handling of the rather complicated multi-resolution grid structures.

The Quadtree is a hierarchical tree structure in which each node is allowed to have either zero or four child nodes, hence the name. The starting node or root Quadtree tile is a square representing the whole globe excluding the polar regions from 85.05° latitude north and south. In the first step, the root tile is divided into four square subtiles, the NE, NW, SW, and SE regions. The dividing lines are the prime meridian and the Equator. Each of these four tiles can be further divided into four square subtiles. This way, the entire globe can recursively be divided into as many tiles as desired (Samet, 1984). This indexing process can go on for any number of steps. The flexibility of the Quadtree approach allows the decomposition of tiles to be performed at different levels for different locations, resulting in a multi-resolution tile grid.

A sample test forecast model was created on various types of multi-resolution grids, which were designed based on the available input data for the forecast, the global earthquake catalog. For generating these seismic density-based multi-resolution grids, we define a threshold for the maximum number of earthquakes allowed per cell, N_{max}. If the earthquake count in a cell exceeds N_{max}, then that cell is further divided into four sub-cells. The resulting four sub-cells receive their share of earthquakes depending on the locations of the earthquakes within the cell. This cell division repeats until no cell contains more than N_{max} earthquakes.

Testing the sample forecast on these different multi-resolution grids shows that the grids with higher resolution perform worse that the lower-resolution grids, showing the need to adjust the precision of earthquake forecasts. Overly precise forecasts tend to overfitting of the forecast while too low precision leads to the loss of necessary spatial resolution. Our new experiment and the implementation of the Quadtree in pyCSEP allow modelers to optimize their forecast based on available data.

Publication outputs:

Asim, K. M., D. Schorlemmer, S. Hainzl, P. Iturrieta, and W. H. Savran (submitted). Multi-resolution grids in earthquake forecasting: the Quadtree approach, Bull. Seismol. Soc. Am.

8. Testing Non-Linear Site Effects

Nonlinear site effects mainly occur for large ground motion at soft soils where there are few measured observations. Most nonlinear site amplification models used in ground-motion models (GMMs) are either partly or fully based on numerical simulations. To test the prediction power of nonlinear site-amplification models, Loviknes et al. (2021) developed a testing framework using observed site-amplification from the KiK-net network in Japan. They tested the non-linear site-amplification models of Seyhan and Stewart (2014), Sandikkaya et al. (2013), Hashash et al. (2020) and the site-amplification model in the GMM of Abrahamson et al. (2014).

The testing framework of Loviknes et al. (2021) consist of three parts:

1. A simple linear ground-motion model is derived on the dataset of interest.
2. The residuals between the predicted linear ground motion and each observation are split into between-event, between-site random effect and record-to-record variability.
3. Site-amplification models are tested against the residuals of individual well-recorded stations and stations grouped into site proxy bins.

Out of all the soft-soil stations in the KiK-net network, 19 stations have recorded sufficient strong-motion records to be included in the test. For most of the selected stations, the linear site amplification model had the best score. Only 5 stations had a non-linear site amplification model score better than the linear amplification model.

Publication outputs:

Loviknes, K., S. R. Kotha, F. Cotton, and D. Schorlemmer (2021). Testing Nonlinear Amplification Factors of Ground-Motion Models, Bull. Seismol. Soc. Am. 111, 2121–2137, doi: 10.1785/0120200386.