

CSEP Workshop: Informing earthquake debates with CSEP results

Report for SCEC Award #18147
Submitted November 15, 2018

Investigators: Max Werner (Bristol), Tom Jordan (USC), Warner Marzocchi (INGV Rome), Andy Michael (USGS Menlo Park), David Rhoades (GNS Science), Hiroshi Tsuruoka (ERI Tokyo)

I. Project Overview	i
A. Abstract	i
B. SCEC Annual Science Highlights	i
C. Exemplary Figure	i
D. SCEC Science Priorities	i
E. Intellectual Merit	ii
F. Broader Impacts	ii
G. Project Publications	ii
II. Technical Report	1
A. Project Objectives	1
B. Methodology	1
C. Results	2
D. Significance	7
E. References	7

I. Project Overview

A. Abstract

In the box below, describe the project objectives, methodology, and results obtained and their significance. If this work is a continuation of a multi-year SCEC-funded project, please include major research findings for all previous years in the abstract. (Maximum 250 words.)

The Collaboratory for the Study of Earthquake Predictability (CSEP) aims to develop a global cyber-infrastructure for the independent evaluation of earthquake forecasting models and prediction algorithms, both prospectively and retrospectively. CSEP thereby contributes to an objective and independent assessment of the predictive power of scientific hypotheses about earthquake occurrences.

The 2018 CSEP workshop focused on the transition to CSEP2.0. After recruiting Bill Savran as CSEP software programmer and publication of a 2018 SRL special issue with 9 papers on the results of CSEP1.0, the CSEP community focused on planning the next phase for CSEP. The current CSEP infrastructure cannot accommodate the new scientific experiments we wish to conduct, from several points of view: 1) the current forecast specification is too restrictive, 2) the current CSEP work flow is too prescriptive, and 3) new, computationally costly models have become available (e.g. UCERF3ETAS).

After perspectives on CSEP2 priorities from SCEC, the USGS and the international CSEP community (US, Europe, Japan, Italy, New Zealand, China), participants presented and developed blueprints for plans for evaluating new types of forecast models, including the current USGS Operational Aftershock Forecast (OAF) models (Reasenber-Jones, FAST-ETAS) and California's UCERF3-ETAS.

B. SCEC Annual Science Highlights

Each year, the Science Planning Committee reviews and summarizes SCEC research accomplishments, and presents the results to the SCEC community and funding agencies. Rank (in order of preference) the sections in which you would like your project results to appear. Choose up to 3 working groups from below and re-order them according to your preference ranking.

Collaboratory for the Study of Earthquake Predictability (CSEP)
Earthquake Forecasting and Predictability (EFP)
Working Group on California Earthquake Probabilities (WGCEP)

C. Exemplary Figure

Select one figure from your project report that best exemplifies the significance of the results. The figure may be used in the SCEC Annual Science Highlights and chosen for the cover of the Annual Meeting Proceedings Volume. In the box below, enter the figure number from the project report, figure caption and figure credits.

Figure 1 (from Field et al 2017): Illustration of the new CSEP type of experiment, which will assess the predictive skill of sets of simulated catalogs rather than synoptic maps. Left: U3ETAS simulations of aftershocks after a M6 Parkfield earthquake. Right: No-faults version of the same model. CSEP questions include: Is u3etas more informative than a standard ETAS model? If so, when? How should simulation-based forecasts be evaluated? How can the process be automated within CSEP servers?

D. SCEC Science Priorities

In the box below, please list (in rank order) the SCEC priorities this project has achieved. See <https://www.scec.org/research/priorities> for list of SCEC research priorities. For example: 6a, 6b, 6c

5a, 5b, 1e

E. Intellectual Merit

How does the project contribute to the overall intellectual merit of SCEC? *For example: How does the research contribute to advancing knowledge and understanding in the field and, more specifically, SCEC research objectives? To what extent has the activity developed creative and original concepts?*

The results contribute to SCEC's goal of understanding the predictability of earthquakes. The workshop has led to new blueprints of forecast experiments that target the predictive skill of USGS and other OEF and OAF models. The developed tests for consistency between observations and forecasts are to a large extent new, original concepts.

F. Broader Impacts

How does the project contribute to the broader impacts of SCEC as a whole? *For example: How well has the activity promoted or supported teaching, training, and learning at your institution or across SCEC? If your project included a SCEC intern, what was his/her contribution? How has your project broadened the participation of underrepresented groups? To what extent has the project enhanced the infrastructure for research and education (e.g., facilities, instrumentation, networks, and partnerships)? What are some possible benefits of the activity to society?*

The predictability of earthquakes is of broad interest. Government agencies use seismic hazard models for building planning and other purposes, but the underlying hypotheses in source models remain debated. Our results contribute to this debate. SCEC-sponsored CSEP workshops are the global focal point for CSEP collaborations and progress.

G. Project Publications

All publications and presentations of the work funded must be entered in the SCEC Publications database. Log in at <http://www.scec.org/user/login> and select the Publications button to enter the SCEC Publications System. Please either (a) update a publication record you previously submitted or (b) add new publication record(s) as needed. If you have any problems, please email web@scec.org for assistance.

II. Technical Report

The technical report should describe the project objectives, methodology, and results obtained and their significance. If this work is a continuation of a multi-year SCEC-funded project, please include major research findings for all previous years in the report. (Maximum 5 pages, 1-3 figures with captions, references and publications do not count against limit.)

A. Project Objectives

The goal of this joint SCEC/USGS/CSEP workshop was to develop new experiments for testing (i) USGS and other OEF/OAF models (ii) testing u3etas, (iii) testing forecasts of finite ruptures on faults. We arranged for presentations and group work to develop blueprints of these experiments, building on CSEP's biweekly teleconference calls and a special CSEP-USGS call in August 2018 to begin developing the experiments.

This focused, by invitation-only workshop brought together members of the global CSEP community, SCEC scientists and IT personal and USGS representatives. The program emphasized the CSEP nodes in California, New Zealand, Italy, China and Japan, and concluded with a session on future directions.

B. Methodology

This one-day workshop included sessions on the following topics:

1. Overview, Motivation and Desired Outcomes
 - Perspectives on CSEP2 objectives from SCEC, USGS and European RISE proposal
 - Group Discussion
2. Planning new CSEP experiments I
 - Experiments for Reasenbergs-Jones and Fast-ETAS models (USGS OAF models)
 - Experiments for u3etas model (USGS OEF California)
 - Testing methods for forecasts specified as synthetic catalogs
 - Group Discussion
3. Planning new CSEP experiments II
 - Perspectives from New Zealand, Japan, China and Italy
 - Group discussion
4. Testing Forecasts of Finite Ruptures
 - Status of RSQSim Forecasting
 - Experiments for testing finite rupture forecasts
 - Group discussion
5. Group work: Prototyping new CSEP experiments
 - Group 1: Testing USGS OAF models
 - Group 2: Testing u3etas
 - Group 3: Testing finite ruptures
 - Group 4: Testing b-value variations, foreshocks, and other specific hypotheses
 - Group Feedback and discussion

C. Results

Session 1: Overview, Motivation and Desired Outcomes

Moderator: M. Werner Reporter: D. Rhoades

Max Werner introduced the workshop background and objectives: We want to define at this workshop the first new experiments for CSEP2.0, including tests of the USGS models. We need to produce draft documents on what those experiments will look like.

Bill Savran presented next. CSEP2 will uphold principles of CSEP1: transparency, reproducibility, a controlled software environment. CSEP1 experiments are no longer running automatically, but can be run "on-demand". Results will be disseminated as data archive and curated data set. The data archive is larger than 10TB. The 1-day forecasts for CA are mostly complete, except for a few models since 2017. New efforts will focus on developing CSEP2 software for forecasting experiments, rather than daily running. Experiments will be defined by the scientific community, can be retrospective or prospective, and will focus on answering particular scientific questions. Software principles: Run on Linux, retrieve catalogs etc from external sources, controlled environment, represented by DAGs.

John Vidale presented SCEC objectives for CSEP2. International datasets are likely to be the most effective to measure the power of forecasts. We need to progress the basic science. CSEP should have concrete demonstration of progress. How much progress has been made, and what are use cases? In response, Andy Michael stated it was difficult to quantify value of use cases because of their variety of uses for a complex field of users.

Andy Michael: USGS agrees that rigorous independent testing is useful and supports funding of Bill's position. CSEP has made impacts: Helmstetter adaptive smoothing has been included in UCERF3. S-test has been adopted to select models for OEF. Different models give different outputs. Sometimes testing is not a deciding factor, as other factors outweigh test results. Testing can be used to hone parameters and choices between generic and sequence-specific parameters of R-J model. Testing can also lead to societal acceptance. What does it mean to test a set of simulated catalogs? Can we fairly compare models? We need prospective, global and retrospective tests. We might want to run tests in-house at USGS using CSEP2 toolbox.

Discussion: We need to respect other nation's own authority for doing OEF in their own territory.

Max Werner (in lieu of Danijel Schorlemmer): The RISE proposal in Europe may be able to support a programmer to help with CSEP. The recent 2018 SRL paper on CSEP achievements and priorities was supported by a wide group. Now it's time to figure out the details of the future priorities. CSEP has difficulty maintaining 4 testing centers because of a lack of developer time. It is also a monolithic and sophisticated (object-oriented) software system and as such not easy to be used for model development. GFZ will take charge of testing centers in Europe and Japan. Danijel's proposal is to break up the existing system into a toolbox. We need to examine software requirements to run experiments. We suggest a workshop in Potsdam to thrash out design of software. Discussion: We want access to forecasts for further analysis after an "experiment" is run. Camilla: We should make sure that everyone can contribute to the toolbox software, not just have access to it (eg Github setup).

Session 2: Planning new CSEP experiments I:

Moderator: D. Rhoades Reporter: J. Gilchrist

Jeanne Hardebeck presented ideas for testing the USGS OAF model Reasenber & Jones (R&J). Forecasting should eventually switch from R&J to sequence based ETAS forecasts, and testing goals include using R&J as a baseline forecast for model testing but we need to decide whether the testing codes should be set up internally or externally to CSEP.

Nicholas van der Elst presented the USGS OAF model Fast ETAS. Stand-alone software has been developed for a Fast ETAS beta version that has several ‘tweaks’ for efficiency, but performance (misfit, over/under prediction, surprise rate, spatial variation...) still need to be tested, and other tests should be performed to determine the value of added computations (additional parameters, site corrections, Bayesian forecasts...).

Ned Field presented the USGS OEF model UCERF3 ETAS. Simulation based forecasts have been implemented for UCERF3-ETAS, and both retrospective and prospective (at least for small magnitudes) tests have been developed, however the most important test, perhaps, is to determine the usefulness of fault-based ETAS for the various uses/users. Figure 1 shows a sample set of simulations for a scenario earthquake, in comparison with a no-faults version of the model.

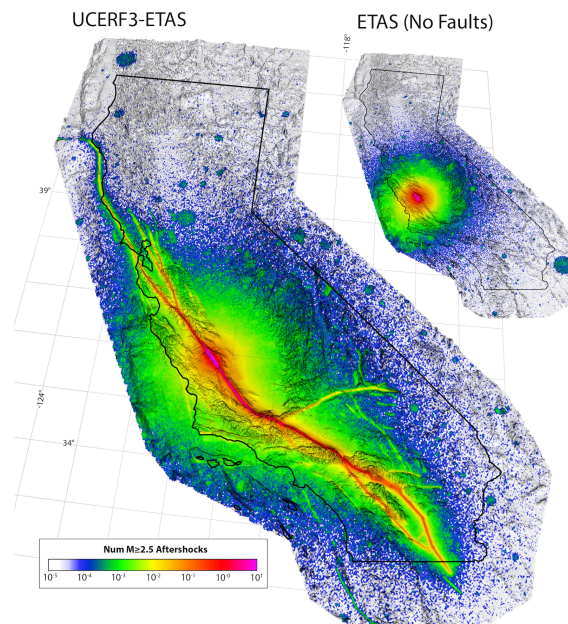


Figure 1: From Field et al. 2017. Illustration of forecasts specified as sets of synthetic catalogs, here using the u3etas and no-faults-u3etas models. u3etas is a high-priority target for CSEP-testing.

David Rhoades presented the current status of statistical tests for synthetic catalogs (rather than synoptic forecast maps). Testing with synthetic catalogs provides greater flexibility, however decisions need to be made with regards to generation and storage of forecasts (updating vs recomputing), as well as what factors (periods, magnitudes...) are more important (and worthwhile) for updating of forecasts.

Session 3: Planning new CSEP Experiments II: Perspectives from China, Japan, New Zealand and Italy

Moderator: A. Michael Reporter: M. Werner

Matt Gerstenberger (GNS Science) provided an overview of CSEP New Zealand and recommendations for CSEP2 priorities. He recommends separating forecasts from the evaluation process and to include software tools for hybridization (i.e. merging two or more forecast models for joint forecasts). He recommends making the evaluation and hybrid tools open access. Matt supported continuing the current CSEP1 grid-based testing and to improve it with non-Poisson testing. CSEP New Zealand also believes that testing in the hazard domain is important and should be planned for the next CSEP phase. In addition, he recommended broader involvement from the community, and using CSEP has a resources for building better forecast models, e.g. via community-contributed software, to turn CSEP into more of a resource than a monolithic software system.

Jiancang Zhuang provided a perspective from CSEP Japan. He provided an update on ETAS model developments, by including a finite source of the mainshock slip model and isotropic aftershock kernels that are strongest near the high-slip patches. The second extension involves a hypocentral ETAS model that can model 3D seismicity in depth. He recommended CSEP consider testing focal mechanism forecasts and 3D hypocenter.

Angie Zhuang provided CSEP China updates. The current system there is not live because of a lack of expertise, and so a China-centric system has been developed instead. Their recommendations involve making CSEP easier, and to strengthen the international collaboration and communication.

Warner Marzocchi added perspectives on how to develop new experiments, while keeping CSEP2 as closely aligned with CSEP1 as possible to maintain continuity. He suggested to keep the grid, while using non-Poissonian synthetic catalogs to evaluate skill.

Yosi Ogata provided an overview of progress on forecasting methods, including

Session 4: Testing Finite Ruptures

Moderator: M. Werner Reporter: M. Page

Tom Jordan reported on progress for time-dependent forecasting using RSQSim and u3etas. Focussing on RSQSim, he identified a need for learning how to assimilate deterministic simulation data into probabilistic forecasting models. They are seeking to use cross calibration and cross testing to understand differences between UCERF3 and RSQSim. Their approach uses a state-dependent forecasting step that uses a Markov renewal process as an underlying probabilistic model. Simple Bayesian methods can be used to assimilate deterministic simulation data into probabilistic models. This is feasible for fairly large numbers. Cross calibration and testing can help us understand differences between hazard predictions.

Dave Jackson presented an update on testing finite rupture forecasts. He focused on the epistemic uncertainty in conditional stopping probabilities (e.g. of segment boundaries). Discussion revolved around the importance from a hazard point of view. For hazard what matters is participation rather than nucleation. In addition, exact segmentation doesn't drive hazard – it's changes in slip rate and paleoseismic event rate that drives magnitude distribution and hazard. Another question revolved around the San Geronimo pass. UCERF3 modelers have doubts about the model's appropriateness there, and so would like to make changes going forward (to make the model more forgiving) – so would tests of segmentation reveal a deep model issue?

Session 5: Break-out group work

Groups developed documents summarizing approaches to forecast experiments, reproduced below:

Group 1: Testing USGS Aftershock Forecasts: Ideas from CSEP Workshop on Sept. 8, 2018

Jeanne Hardebeck, Andrew Michael, Negar Nazari David Rhoades, and Nicholas van der Elst

Testing is needed to hone existing forecast methods, provide a route for new methods to be accepted, verify forecast accuracy and build trust with users. Working with CSEP provides an independent authority to validate the testing methods and results, further building transparency and trust.

Testing methods must be applicable to non-Poissonian distributions and correlated forecasts from ETAS models, epistemic uncertainty incorporated into non-stationary Poisson forecasts such as Reasenberg and Jones, and correlations from overlapping forecast windows. This can be accomplished with simulation based forecasts and tests that project those forecasts onto spaces such as space-time-magnitude grids of numbers of events and inter-event times and distances to be sensitive to inter-event correlations.

There are a number of proposed approaches to carry out these tests including using empirical distributions to compute likelihoods, pseudo-likelihoods, and comparing K-S statistics between the forecast and observations to the same statistic from the forecast to simulated observations. We need to do synthetic tests to

better understand which approaches provide the greatest statistical power for various features of seismicity under realistic conditions for the amount of data.

In addition to selecting the best models, there are a number of topics that can be addressed with CSEP testing:

- 1) Useful forecast horizons at different points in an aftershock sequence. For individual models, such as ETAS, this will test our ability to estimate useful parameters. For ensemble models, such as used in NZ, this may determine the time periods at which various components are most useful. Note that in many engineering applications (tagging buildings) and policies, the first week is not considered because the aftershock rate is so high. Thus, forecasts after the first week are important and we need to determine at what point we can make those forecasts.
- 2) Whether to use Bayesian updating of parameter distributions or some other ensemble approach between the generic and sequence specific models. These ensembles may do a better job of maintaining the true variability of seismicity in the forecasts which is represented in the generic models but may be lost as we converge to sequence specific parameters.
- 3) Will allowing b-value to vary improve both our generic, sequence-specific, Bayesian and/or ensemble forecasts.

Some tests need to be done on the outputs of greatest interest to the users such as numbers of expected events or ground motion hazard. Even if these don't have the greatest statistical power they may be more effective in convincing users that the forecasts are valid.

Tests on ground motion may need to go beyond hazard curves to include expected numbers of times a specific ground motion will be exceeded. Otherwise, we may not be able to maintain non-Poissonian distributions and correlations in the ground motion tests.

A problem with tests based on ground motion is that a biased ground motion prediction equation (GMPE) could lead to an opposite bias in the earthquake rupture forecast models. This bias could be avoided by applying the same GMPE to observations and forecast simulations. However, then the GMPE is simply a forecast smoothing function and we are not truly testing based on observed ground motions.

We need to account for catalog incompleteness in the tests, just as we do in estimating parameters and making forecasts. We want to forecast true earthquake behavior because that is what users experience. However, the tests will be done on incomplete catalogs and we don't want the tests to be biased by comparing forecasts of completed earthquake behavior to incomplete catalogs.

Some tests need to be done on the actual forecasts released by the USGS, including manual intervention by a seismologist. Others need to be done on fully automated forecasts to accurately test the methods. These tests will need to be done both in retrospective and prospective modes. Whether to carry out the fully automated tests by installing the operational systems in CSEP or submitting the forecasts as externally needs further discussion. The actual released forecasts will need to be tested in an external mode. So that mode will need to be developed and could be used for both purposes, if that allows for sufficient transparency and confidence in the tests.

Group 2: Testing u3etas

Notes by M. Werner

U3etas generates synthetic catalogs, thus requiring a new CSEP experiment blueprint and new tests. The model falls into a broad group of simulation-based forecast models that also include standard ETAS models. As such, it can be a useful model for starting a broader class of CSEP experiments.

Testing u3etas will require substantial computing resources. We discussed where these would be available, e.g. via the USC cluster, or via the cloud. Initial tests should establish the required computation resources. A second question revolves around how to automate this process from the CSEP platform.

Group participants agreed that retrospective tests were the ideal starting place, e.g. the aftershock sequences of the largest earthquakes in California over the last 2-3 decades, for which there are good instrumental catalogs. Simulations should start with the mainshock and generate forecasts of various horizons (1 week, 1 month, 1 year) and these should be updated whenever there is a significant event (e.g. M3.5) or after 3 days or so.

Testing methods remain to be assessed, but candidates include the tests that David Rhoades and others developed for simulation-based catalogs. Initial tests will be visual: checking that the aftershock productivity matches observations, that spatial forecasts match observations etc. These would lead to formal definitions of tests, e.g. the N-test.

We discussed calculating pseudo-likelihood scores, by re-introducing the grid, calculating mean rates in each grid cell, and using a pseudo-likelihood function (i.e. an approximation to the actual LL function, which is not analytical) to compute pseudo-LL scores. We also discussed the question of power in existing CSEP consistency tests, and ways to identify what drives the CSEP test results.

The plan is for Bill Savran to establish computational resources and to develop retrospective simulations of aftershock sequences, starting with the 1993 Landers earthquake.

Group 3: Testing finite ruptures *Notes by M. Page*

We propose a “similarity index” methodology that compares synthetics produced by model finite ruptures to observables following real earthquakes.

Two types of similarity indices could be used as separate tests within CSEP. The first is aftershock similarity. There would be a module within CSEP that produces a synthetic aftershock distribution for any finite fault rupture. This module would be applied to the finite ruptures of a given earthquake probability model under test within CSEP. The synthetic aftershock locations for a given model rupture would then be compared to the observed aftershock locations for the real earthquake.

A second type of “similarity” index would be based on ground motions. A ground motion module within CSEP would apply a GMPE (for example, PGA) to finite fault ruptures from the models under test. This would be compared to the Shakemap for real earthquakes (Shakemaps are routinely produced for large global earthquakes by the USGS).

Both similarity indices would vary from 0 to 1; with 1 meaning perfect agreement. Each model in CSEP would be scored on each observed earthquake; this score would sum over all model earthquakes and be proportional to model earthquake rate multiplied by the similarity index. In this way, models with high rates of earthquakes with high similarities to what just happened would be scored high. In practice, most model earthquakes would likely not need to be included in the similarity calculation because they would have a negligible contribution to the sum (due to not overlapping with the real earthquake area or having a very different magnitude). There would also be a second term (much as likelihood has) that penalizes for the overall earthquake rate (so that a model doesn’t “win” the test by just having high rates of all ruptures).

Development of the ground motion similarity index would also naturally allow for the development of a hazard test within CSEP, which would just involve summing the ground motions from all observed earthquakes and comparing that to summed model ground motions. It would be perfectly possible for a model to do very poorly on rupture by rupture basis but nevertheless predict total hazard quite well; for this reason, the hazard test would be a separate test from the individual earthquake ground motion similarity test.

Group 4: Testing b-values, long-term rates, foreshock algorithms and other CSEP experiments. *Notes by M. Werner on the basis of J. Zhuang’s presentation*

Participants noted the potential of induced seismicity as a complementary point of view on the predictability of seismicity. This would require operational parameters (e.g. injection data) as a model input. Could CSEP provide this?

The issue of complete v incomplete data was discussed. Catalogs are always incomplete, while some models, e.g. ETAS, is a model for complete data. But in reality the data is incomplete, so we need a model for the mechanism of missing data to take this into account.

Understanding the physics behind foreshocks, quiescence, fluid intrusion was discussed, and the possible implications for b-value variations and their relation to stress. Forecast models were discussed that build on ETAS as the backbone and include other components to assess predictive skill of precursory information.

The group noted the existence of lots of diagnostic tools and metrics.

Finally, the group questioned the current utility of AI methods for improving forecasts.

D. Significance

The workshop developed blueprints for future CSEP experiments, opening the path towards independent evaluations of government agency operational earthquake forecasts, including testing R&J, Fast-ETAS and u3etas. Participants also developed a new method for testing forecasts of finite ruptures. As experiments and tests are developed for model classes, rather than the specific models, the developed CSEP tools will be useful to a much range of models.

E. References