CSEP Workshop: Informing earthquake debates with CSEP results

Report for SCEC Award #19235
Submitted November 13, 2019

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

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 2019 CSEP workshop focused on two major themes: Operational Earthquake Forecasting (OEF) during the Ridgecrest earthquake sequence and progress towards CSEP2.0 priorities. The workshop featured four sessions. The first two sessions focused on the forecasting of the Ridgecrest sequence. Speakers presented several forecast models, including the official USGS OAFs, simulations from the UCERF3-ETAS (u3etas) model and Coulomb-stress based models. Bill Savran showed new CSEP2 tests of u3etas Ridgecrest forecasts comprised of simulated earthquake catalogs. The third session focused on CSEP updates from around the globe (Italy, New Zealand, Japan, China and global). The fourth session featured a stimulating discussion around the relative paucity of earthquakes in California over the last hundred odd years compared with earthquake rates implied by the paleo-record.

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.

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. Continued improvements in Coulomb rate-state models were presented that are beginning to challenge benchmark ETAS models. USGS public aftershock forecasts were evaluated with a focus on which features could be improved for societal purpose.

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 Pubications 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 objectives of this joint SCEC/USGS/CSEP workshop were (1) to derive lessons from the operational forecasting during the Ridgecrest sequence, (2) assess progress towards CSEP2 objectives, (3) discuss updates from the global CSEP community and (4) present developments in the paleo-hiatus problem.

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.

B. Methodology

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

1. Ridgecrest Sequence I
    - Welcome and Overview
    - USGS Operational Aftershock Forecasting
    - Coulomb stress rate/state forecasting
    - U3etas forecasts
    - Group discussion
        1. Effect of near real-time data products on forecast quality
        2. Predictive skill of the Coulomb hypothesis
2. Ridgecrest Sequence II
    - U3etas evaluations and development of new CSEP2 tests
    - Future UCERF models
    - Group Discussion
        1. Faults, fault models and the earthquakes in-between
        2. Lessons from the Ridgecrest sequence for OEF
3. Global Perspectives
    - Perspectives from Italy, New Zealand, Japan, and China
    - Global earthquake forecasts
    - Status of CSEP2 @ SCEC
    - Group discussion
        1. The global collaboration
        2. Progress towards CSEP2 goals
4. The Paleo-Hiatus in California
    - Hiatus probabilities – two approaches

- Overdue?
- Explaining the paleo-hiatus
- Group discussion

5. Group Feedback and discussion

C. Results

Session 1: Ridgecrest I
Moderator: M. Werner Reporter: D. Rhoades

Jeanne Hardebeck - USGS Operational Aftershock Forecasts (OAF).

Manual forecasting went live in August 2018 and automated forecasting in September 2019. The system has issued OAFs following most earthquakes of magnitude M≥5 in the U.S. (13 events so far), including the M7.1 earthquakes in Anchorage and Ridgecrest.

The forecasts use the Reasenberg & Jones (RJ) model, initially with generic global or California parameters. The RJ model is approved by the National Earthquake Prediction Evaluation Council (NEPEC). It is a simple and stable model. The aim is for transparency, and to provide a baseline for more complex models, such as ETAS. Challenges include the use of sequence-based parameter estimates and the non-Poissonian nature of aftershock occurrence.

Forecasts give a 95% confidence region for aftershock occurrence. They are evaluated by comparing the forecast and observed number of events with M≥3, 4, 5 etc. This evaluation ignores non-independence and overlapping periods. Short duration (1-day) forecasts are usually accurate. Longer term forecasts can over- or under-predict. Performance improves with sequence-specific parameters. The generic p-value in the RJ model appears to be too low.

During the Ridgecrest sequence, the first forecast was issued 35 minutes after M6.4 event. A new start was made after the M7.1 event. Using sequence-specific p and c values led to under- and over-prediction at different times. This sequence is too complex for the simple RJ model. The magnitude of completeness versus time is fitted well by the Helmstetter et al (BSSA, 2006) model.

Simone Mancini - Coulomb stress/Rate-State (CRS) forecasts

Five CRS models of increasing complexity and an ETAS model have been pseudo-prospectively tested on the Ridgecrest sequence, with a learning phase of 1/1/1981- 3/7/2019, a testing phase of 4/7/2019-4/8/2019 and a 3-rupture-length test region. The target events are M≥2.5. The forecasts start after the Mw 4.0 foreshock and are updated daily. The simpler CRS models use only the focal mechanism (FM). The more complex CRS models progressively incorporate finite-length slip models (FLM), secondary triggering and spatially variable receiver faults.

Over the whole test period, the ETAS model shows the best spatial consistency with the observations, but not in the first 24 hours after the M7.1 mainshock. Immediately after the mainshock the CRS models do better. The spatial consistency of the CRS models improves with added complexity. The combination of stress inversion and small-scale structural heterogeneity represented by past focal mechanisms corresponds to the best performing CRS model.

Preliminary CRS models systematically underestimate daily seismicity rates. CRS models with triggering perform well after the M6.4 event. All non-preliminary models overpredict the number of aftershocks after the M7.1 event. The L-shaped seismicity region presents a problem for the CRS models.

The comparative evaluation of ETAS and CRS models suggests that statistical forecasts should incorporate fault information (ETAS-fault), while stress-based models should include FLMs and consider structural heterogeneities to increase their information gain.

Kevin Milner- OEF U3ETAS forecasts during Ridgecrest sequence

U3ETAS merges point-process spatiotemporal clustering with the UCERF3 finite fault model. Its output is in the form of synthetic catalogs. It requires high performance computing. It is not operationalised; it is run on demand. During the Ridgecrest sequence, the first forecasts were produced about 1 hour after the M6.4 event. The Ridgecrest sequence occurred in an “off-fault” region. There is an estimated probability of 2.7% of another M6.4+ in the next week and 0.24% of an M7.1+. These estimates are lower than in a generic model, because of anti-characteristic behaviour in off-fault regions in U3ETAS. After the M7.1, the probability of another M7.1+ still lower than in the generic model because of the anti-characteristic magnitude distribution.

We need a finite-fault source (FFS) to better forecast the spatial distribution of aftershocks. Adding a FFS increased probability on the Garlock fault (1 month) from 0.62% to 1.71%, then to > 4% with another FFS. There is high sensitivity to the FFS. Shakemap was used as a more authoritative FFS. Tele-seismic inversion sources were far too long. Even with the Shakemap source, some sensitivities still exist.

Should a parameter be added to U3ETAS to force off fault area magnitudes to be Gutenberg-Richter? Could some sort of source uncertainty parameter be included to make results less sensitive to poorly constrained FFS? There were wide ranging results for triggering on other faults among 100,000 simulations but, with its default parameter stings, the model overpredicted the number of aftershocks in the Ridgecrest sequence.

Discussion

Yosi Ogata asked where did the 5% probability of something larger, reported to the media, come from? He calculated 20% after the M6.4 and M7.1 events using ComCat data. Andy Michael stated that the 5% came from an early generic model of Lucy Jones. USGS currently has 10% probability as a generic model estimate.

What causes the p-value to vary in the RJ forecasts? Clustering out of the ETAS model or something else? The p-value changes reflect multiple processes evolving on multiple timescales.
Max Werner asked: Is Ridgecrest a candidate for a CSEP retrospective forecasting experiment (like Canterbury)? Also, what lessons do modellers have for data-product providers? Finite fault model is one. What others? Andy commented that the catalog has improved markedly over time, but the location uncertainties are important in some regions, and reducing the magnitude of completeness is helpful.

Session 2: Ridgecrest 2
Moderator: D. Jackson Reporter: M. Werner

Bill Savran: Simulation-based test and examples from u3tas forecasts of the Ridgecrest Sequence

Bill presented new tests designed to overcome Poisson and independence assumptions. These were applied to weekly (disjoint) u3tas forecasts for the Ridgecrest sequence. The new tests included an N-test built from the empirical number distribution from the 100,000 forecast simulations. The new Magnitude test employs the concept of a union catalog, which combines all the simulated catalogs into one. The joint magnitude distribution is scaled to the number of observed events. Then, a discrepancy metric between the union catalog and each simulated catalog is introduced, and the metrics from the simulated catalogs are compared with the observed value. Bill also introduced two pseudo-likelihood tests, of which one isolates the spatial dimension.

Bill next proposed to evaluate multiple time periods by assessing the distribution of p-values against a uniform distribution. A perfect model would generate p-values against data according to a uniform distribution, so deviations can tell us about potential discrepancies. Bill found that the u3tas model

Morgan Page asked whether forecasts were penalised unnecessarily for all the small events triggered by exceedingly rare large aftershocks. Bill replied that he's made progress on this question with conditional likelihood tests that condition the ensemble of forecasts by the number of observed events.

Ned Field: Future UCERFs

Ned presented challenges and needs for future UCERFs. These included:

1. Improved representation of epistemic uncertainties
2. better understanding of what faults actually present
3. inferring MFD characterists near faults
4. the need for elastic rebound
5. details of large event triggering
6. Valuations
7. u3tas requires time-dependent background rates because of missing ancient identified parents; and spatial variability
8. better presentation of uncertainties in MFDs. e.g. Poisson vs ETAS uncertainties - is there a bulge?

A question was whether the empirical MFD during the Ridgecrest sequence was non-GR? Data shows that the M3.5 range shows deviations from GR. Is square distance metric appropriate? Bruce Shaw proposed CSEP take into account uncertainties in magnitude determinations, as catalog operators might actually change their magnitude estimation algorithms.

Ogata: Ridgecrest

Ogata-sensei showed aftershock forecasts after Searles Valley and Ridgecrest. After the M6.4, the probability of an event larger than M6.4 was about 10% over the next day or so. Ogata also showed variability of the ETAS parameters as a function of time after both main shocks, showing a decrease of p-value after M6.4, while it was stable after the M7.1 (apparently in contrast to the USGS RJ model findings). His foreshock model shows that the (Coso) region exhibits generally a high foreshock probability, as well as high background rates.

Discussion

How different is the Ridgecrest earthquake sequence from the other M7 sequences in California? Quite different in terms of finite source model e.g. from El Mayor. How do we anticipate Ridgecrest earthquake in advance? When do you introduce new information or extra degrees of freedom?

Ned Field proposed to define use metrics that measure 'value' of forecasts. For that we need to understand uses.
Yosi Ogata: We should use a conditional magnitude test, rather than an unconditional magnitude test, e.g. looking at mainshock magnitudes conditional on foreshock magnitudes. This would be important to evaluate foreshock models with time-dependence magnitude distributions. Non-GR magnitude distributions that vary in space (e.g. UCERF) also require conditional magnitude tests.

Andy Michael: Regarding valuation, Powell centre solicited user feedback, and FEMA provided some information. A lot of use outside re/insurance will not be very quantitative, and response/value may be subjective and emotion-based. So perhaps defining value to users will be difficult and may not reflect how users actually make decisions. This chimes with behavioural research that shows most people

Faults and earthquakes in between:

How do mesh the two scales? UCERF showed 50% of subsequent event being on and off fault. So Ridgecrest consistent with that. The mapped faults will never be complete, so these surprises will continue to happen. But are too many future earthquakes not on faults, e.g. like in New Zealand.
How do we test the hypothesis that the large events on faults are different from those off fault. Morgan Page has a paper from last year in which she tested for variations of b-values before and after definitions of faults have been made. No obvious data to suggest that faults can be identified by b-values before large events happen that define the faults.
Field: To test, we may need to stack faults around the world?
Stirling: Also need to consider time-dependence of seismicity, and its effect on MFD, e.g. characteristic vs GR.
Albanna: We should use physics based models.
Page: Seismicity very diffuse versus moment release (which is geodetic and on faults). Zeng will talk about diffussiveness and rupture strain might be time-dependent.
Max: Ridgecrest would not have happened in RSQSim-land. But it would have generated Coulomb stress changes on the model fault network and so therefore off-fault events influence the model - do we understand the implications to RSQSIm answers to questions we want to ask of it for future UCERF?
Page: If faults unknown, we need to fill in stochastic faults.

Conditional s-test discussion
Morgan, Bruce, Ned made suggestions for conditional S-tests in CSEP2.

Session 3: Global CSEP Perspectives
Moderator: D. Rhoades Reporter: A. Michael

Giuseppe Falcone explained how CSEP Italy is an integral part of their OEF program which will be moving from being available only to government agencies to being fully public in 2020. The public forecast will be the probability of M≥4 earthquakes. The Italian forecasts rely on an ensemble of ETAS models and STEP (an ETES model had a bug in it and so is not being used). All individual models and the ensembles are under testing. Currently, the SMA weighting scheme, which tends toward giving some weight to all of the models, is being used to produce the ensemble. However, ensembles using the GMA and gSMA schemes are outperforming the SMA ensemble and the official forecast model will move to one of those two schemes. Those schemes tend toward giving all weight to the best performing model and so this move will have the effect of choosing an ETAS model, which has performed best on recent sequences, as the basis for the forecasts. The use of an ensemble allows the system to change this choice or give more equal weights to the models based on future seismicity. Other future plans include migrating the software to Python and including focal mechanism information in the forecast models.

David Rhoades updated us on CSEP New Zealand which is not currently operational due to inadequate staffing. After the extreme efforts needed to respond to the Canterbury and Kaikoura earthquake sequences they are focusing on model development. The unique features of models being developed in New Zealand are the use of EEPAS and the focus on medium-term and long-term ensemble models for hazards assessment. For the hazard assessment models, they are considering how to best estimate hazards in low seismicity rate areas such as Canterbury and Aukland, the largest city in New Zealand. For the EEPAS model, current work includes a global model and automatic treatment of parameter uncertainties and tradeoffs. The latter is being tested using results from the RSQSim numerical earthquake simulator. David Rhoades showed some interesting results suggesting that the EEPAS model could have provided useful forecasts of the Ridgecrest earthquakes in California.

Yosi Ogata showed that CSEP Japan is currently testing 160 models many of which are variants of the ETAS model. The extremely large number of earthquakes triggered by the Tohoku-Oki earthquake made passing the N-test difficult but the models are recovering to better performance as time passes. The Information Gain and Probability Gain metrics are proving particularly useful in their work. Under the next phase they will be adding many new models and some new tests. The new models include real-time forecasts, automatic swarm detection, maximum intensity forecasts, and a 3-D spatial forecast model in the Kanto region. The results from CSEP Japan are used to advise government panels such as Earthquake Research

Committee, which meets monthly, and the Coordinating Committee for Earthquake Prediction Research, which meets quarterly. They will be holding a CSEP Japan workshop in November, 2019.

Yongxian Zhang talked about the very wide range of models that CSEP China is testing. They are working on the most unusual seismicity-based models which move beyond the standard clustering approaches such as ETAS and STEP. They are testing 11 models that include RI, PI, LURR, and hybrids such as MMEP (Multi-Method Earthquake Prediction). To focus on areas with the best data, including geodetic observations, they are establishing two Seismic Experiment Sites which will focus on a systems approach including engineering. The results of the tests are used to inform their Annual Consultations on earthquake probabilities. The funding for these two sites begins in 2020. With the testing center for these sites all data and results will be openly available.

Jose A. Bayona Viveros presented the current efforts of the GEAR model group to produce a new global seismicity and geodetic hybrid model. The GEAR models primarily depend on the geodetic component at plate boundaries and smoothed seismicity in between. GEAR1 has a focus on improving performance in subduction zones using a new regionalization scheme, a new global strain rate map, and velocity dependent seismogenic depth limits. It also uses regionalized parameters and moment balancing between geodesy and earthquake rates. The GEAR models performed well in short-term CSEP tests and these tests are also being applied to the new model and are being used in its development.

Considering the activities in all of the CSEP testing centers, one issue is that each center is focused on testing a unique range of models and many are only being tested in one place. Faster progress may be possible if more models are tested in a variety of regions. Perhaps the CSEP2 software will make that easier to achieve.

Session 4: Earthquake Rupture Simulators & the Paleo-hiatus
Max Werner (moderator) and Kevin Milner (reporter)

Hiatus probability - two approaches - Glenn Biasi

Glenn reported that the basic observation of the hiatus problem is fairly simple: 31 paleo sites in UCERF3, none with a ground rupturing earthquake in the past 30 years. One can get a rough estimate of probability from “independent” conditional probabilities, and a better estimate through removing correlations using paleoseismic data. Either way the probability of a 100 year hiatus is very low.
Here, he focused on 6 well studied sites. In approach 1, he estimated correlations between nearby paleo sites to reduce problem to a net Poisson event rate (removing correlations). In approach 2, he sampled event chronology from event time PDFs, to test for hiatus. The take home message is: Rate that either event chronology has 0 events in 100 years is very low, even with just 6 sites. This virtually never happens when more sites are included.
Question from Max/Ned - what’s your interpretation? We may be in very strange times, “we don’t know what these paleoseismic data mean”. Events at sites were regular in the past. One interpretation is that fault system is locked, activity is currently off the main system. Even if multiple events happen at these sites in the next 10 years, our variability models are wrong (not enough variability).

California: Underdue - Due - Overdue? - Dave Jackson

Dave suggested that we talk about “overdue,” but that implies that at times we would be “underdue,” yet we never talk about that. Common assumptions include: Important faults are known and mapped, and big earthquakes occur primarily on them; events are quasi-periodic. Dave contends that we don’t really know much about the record in the last 1,000 years unless we trust the paleoseismic record. Hiatus corresponds with the instrumental era. NZ had a similar hiatus in the last century. Data selection is difficult without bias - need to define terms and rules before looking at data (but not always possible). Dave looked at data from 2 intervals: before and after 1920. Data before tends to fall right on mapped faults, which is a bit suspicious and probably a result of analysts assigning locations onto faults by default. Catalogs looks complete down to ~1800 (below that around 1850). Rate of M7’s is pretty constant over 200 years, 19 in total. There is a

66% chance of having 9 or more M7's in the past 100 years (estimated from Poisson's distribution with actual event count). Onderdonk et al (2013) found displacement in paleo site during the instrumental era that couldn't be attributed to an earthquake at that location. Maybe other paleo events from the pre-instrumental record aren't earthquakes? Note overdue for a M7, and recurrence statistics on individual faults are meaningless because they depend on questionable data. Paleo sites can't all be overdue, while the state as a whole is not.

Comment from Mark Stirling - some of these structures are clearly very important for accomodationg plate movement. And some paleo sites have quite clear evidence of violent shaking at the site. Dave responded that he was not saying that none are earthquakes, some clearly are. Glenn Biasi noted that M7 statewide rate is a different question than point observations on faults. Glenn suggested to come up with specific examples in the record that might be questionable and discuss with paleo geologists. Dave responded that a lack of contrary evidence doesn't mean it was an earthquake. Mike Blanpied asked how many false positives would we need to have in order for the data to agree. Mark Stirling noted that 2 nearby trenches in NZ examined by different teams, one found 3 and the other found 6.

Explaining the paleo-hiatus - Morgan Page

Morgan presented u3etas simulations. Even at century scale, ETAS models have large variability of 100-year rates. Dead Sea Transform has huge variability: some 5,000 years have no events, some have 8. ETAS can match most of this (but not the longest period earliest in the catalog). UCERF3 paleo hiatus probabilities (accounting for probability of detection in trench) are: time-independent 0.5%; time-dependent 1.7% (higher due to elastic rebound); ETAS 3.5%, higher yet due to aftershock clustering. In comparison, RSQSim probability is 3.7%, but there are some differences in processing from u3etas results: statistics were not conditioned on post-1919, instead they represent generic 100 year period. Also the model doesn't fit the paleoseismic data, and it has lower rates than UCERF3 at paleo sites. Chance of 100-year hiatus sometime in a generic 1000-year period is much higher (28%).

Session 5: Discussion
Moderator: Max Werner Reporter: Bill Savran

The workshop ended with a broader discussion about the philosophy of the new CSEP2 tests. USGS released forecasts to the public creates need for accountability and transparency toward the public. Tests are qualitative in nature. CSEP tests seem limited for helping model development and maybe they should be model specific. The spatial test is the most diagnostic for a model, and we should focus our efforts on these types of diagnostic tests.

Regarding the framework and code developments toward CSEP2, toolkit should be able to easily export to other tools languages. The Python toolkit should clarify exactly what each function is doing. It will need good documentation. Can we use information about the models, b-values, MFDs, to help us limiting the computational cost? Italian OEF has moved from Matlab to Python, and should develop a database to store the results and then automatically create figures from there. This is consistent with what SCEC has been considering

Concerning lessons learned from Ridgecrest, participants discussed on and off-fault earthquakes: how can you score earthquakes that occur in this (off-fault) region. How can we say whether an earthquake is a surprise or not a surprise? We can use strain-rates to understand regions where to expect earthquakes, off-fault versus within a straining (localized?) area.

D. Significance

The workshop brought together the SCEC, USGS and global CSEP community to discuss a number of important topics in earthquake forecasting and model evaluation, including (1) the current state of operational earthquake forecasting and forecast evaluation on the basis of the Ridgecrest sequence experience

(2) global progress on forecasting and model developments and (3) progress on the paleo-hiatus problem. The CSEP workshop at SCEC forms the focal point of the global CSEP collaboration and is a crucial venue for exchanging ideas and making progress towards the CSEP2 goals.

E. References