CSEP Workshop: Informing earthquake debates with CSEP results

Report for SCEC Award #17175
Submitted November 15, 2017

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

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 cyberinfrastructure for the prospective and blind evaluation of earthquake forecasting models and prediction algorithms. CSEP thereby contributes to the objective assessment of the predictive power of scientific hypotheses about earthquake occurrences. This workshop had the following objectives: 1) to retrieve and interpret new CSEP results from around the globe, 2) to develop a publication plan, 3) to plan next steps for retrieving short-term model results, and 4) to wrap phase 1 of CSEP and begin planning a phase 2 that is aligned with current community needs. A volume with extended abstracts is available on request. Speakers presented CSEP results from California, New Zealand, Italy, Japan and the global experiment. Participants agreed to publish new CSEP results in a focus section of the Seismological Research Letters (submission deadline 1 February 2018 and publication date Jul/Aug 2018). Wrapping up phase 1 of CSEP will require reprocessing in California and New Zealand to obtain complete evaluations of short-term models because of earthquake catalog issues (in New Zealand) and incompletely processed results. Recommendations were collected for CSEP2.0, ranging from new experiments and tests (e.g. to test UCERF3-ETAS) to recommended model developments to accessible outreach tools.

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: Spatial visualization of the likelihood of earthquakes M4.95+ between January 2011 and August 2017 given the probabilistic forecast “Neokinema” by Bird & Liu (2007). Size of squares (earthquakes) indicates magnitudes; in-fill colour of squares denotes difference in log-likelihood between Neokinema and a spatially uniform benchmark. The Brawley earthquakes of 2012 are well forecast by Neokinema; the Hawthorne earthquakes in 2016 are not, and two more earthquakes are surprising (circles).

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. Workshop results suggest that strain-rate based forecasts are competitive alternatives to smoothed seismicity forecasts; Coulomb-based forecasts can compete with purely statistical models; and new ways for assessing and visualizing predictive skills have been developed.

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. Early career researchers (including two women) were invited to present their work. SCEC-sponsored CSEP workshops remain 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 goal of this joint SCEC/USGS/CSEP workshop was to inform scientific debates about the predictability of earthquakes through CSEP results. Our objectives were to present available global CSEP results and to analyze their bearing on ongoing contentious debates in the seismological community. Targeted debates include: 1) How do magnitude distributions differ on-fault and off-fault? 2) Does the b-value vary in space and/or time? 3) How do strain rates map into earthquake rates? 4) Does the spatial distribution of small earthquakes help forecast large earthquakes? 5) What is the predictive skill of the Coulomb stress hypothesis? 6) What are maximum magnitudes on fault segments? 7) How does elastic rebound manifest itself in earthquake clustering? 8) How should ensemble models be constructed to provide optimal forecasts in an operational setting?
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 and Japan, and concluded with a session on future directions.

B. Methodology

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

1. CSEP Results I: Evaluations of Long-Term Models
    - Results from California, Italy and Japan
2. CSEP Results II: Evaluations of Long-Term Models
    - Results from New Zealand and China
    - Evaluations of global high-resolution forecasts
    - Group discussion
3. CSEP Results III:
    - Results from the retrospective Canterbury experiment
    - Results from California, Japan and Italy
    - 3D forecasting experiments underneath Kanto, Japan
    - Group discussion
4. Wrapping Up CSEP 1.0
    - CSEP achievements, further results analysis
    - Publication plan
5. The next phase: CSEP 2.0
    - Pop-up presentations on future directions
    - Break-out sessions

C. Results

Session 1: CSEP Results I: Evaluations of Long-Term Models

Moderator: D. Rhoades Reporter: J. Zhuang

Anne Strader reported the evaluation results on the long-term (five earthquake) forecasting models for California, including RELM models, UCERF2 and NSHMP, using the CSEP testing procedure. The main results show USGS seismicity models passed all tests during five-year period; however, there is evidence of long-term underprediction from 40-year experiment results. UCERF3 addresses UCERF2's underestimated seismicity rates through implementation of elastic-rebound models and multi-fault ruptures, which increase the probability of large earthquakes along major known faults. Forecasts rejected in favor of UCERF2 during 40-year period only included southern California: are incorrect USGS model seismicity rates confined to central/northern California?

Using Bootstrapping methods, Gerestenberger tested the consistence of AKJ, ALM and SUP models by N-, L-, S- and M-tests. The testing results show that consistence does imply better performance in forecasting. He concludes that each test-period (likely) contains some non-random sample. Therefore, CSEP tests are informative of how a model performs in that one period and may hint at how a model will perform in other periods. However, extrapolate from any one period to any other is not reliable.

Max Werner reports the progress on hybrid RELM models for California and their evaluation results. Different from typical methods for ensemble modelling, which includes Bayesian model averaging, parameters involved in combination method (e.g. weights) of hybrid modeling are determined over some training period and are then fixed. This study considers multiplicative hybrids. The results show that HKJ-hybrid models gain better performance in the CSEP testing. Figure 1 shows a spatial visualisation of the likelihood of earthquakes given the Neokinema (RELM) forecast by Bird & Liu (2007) compared against a uniform benchmark.

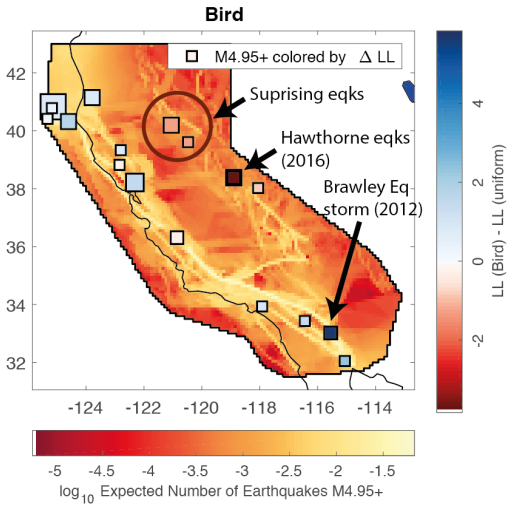

Figure 1: Spatial visualization of the likelihood of earthquakes M4.95+ between January 2011 and August 2017 given the probabilistic forecast "Neokinema" by Bird & Liu (2007). Size of squares (earthquakes) indicates magnitudes; in-fill colour of squares denotes difference in log-likelihood between Neokinema and a spatially uniform benchmark. The Brawley earthquakes of 2012 are well forecast by Neokinema; the Hawthorne earthquakes in 2016 are not, and two more earthquakes are surprising (circles).

Taroni reports the evaluations results of 5 and 10-year CSEP Italy models and ensembles. The results imply that he best model use both seismicity from catalog and fault information and that under-smoothed (granular) models have poor performance.

Tsuruoka reports the results on the 3-mont and 1-year CSEP testing classes, using L-, CL, N-, M, S-, and W/T tests. The results show that it is difficult to forecast Mainland testing region rather than all Japan testing region. It is effective to evaluate models by Information Gain for multi-round of testing. Also, after the 3.11 Tohkuo EQ, for Kanto-testing region, Omori-Utsu formula is powerful to forecast earthquake num-

bers. He mentioned that extra care need to be taken of resolution of the test region when using S-Test.

In summary, better models can be yielded through including more information ingredient from multiple types of data, such as seismicity, faults, and other geophysical data. Technologically, ensemble or hybrid modelling provides better forecast results based on available individual model. The results also reveals some problems of related to the CSEP tests: (1) The resolution problem in S-tests has been reported by several reports.; (2) for long-term, our test period is not long enough; (2) data quality control of the target earthquake catalogs requires pre-process of the raw catalog; (3) after the occurrence of huge earthquakes, the seismicity rate changes in a large scale, causing poor performance for long-term models, such as the time periods after the Tohoku earthquake in Japan and the Emilia earthquake in Italy.

Session 2: CSEP Results I: Evaluations of Long-Term Models (continued)
Moderator: D. Jackson Reporter: M. Stirling

Part 2 of this session provided overviews of CSEP activities in New Zealand and China, as well as those of globally-based studies. David Rhoades described the development and activities of the New Zealand Testing Centre. The Centre was established in 2008, but was plagued by many issues associated with the catalogue in the early years. Delays in finalising the locations and magnitudes of earthquakes, magnitude determinations, and the transition to SC3 automation were issues that delayed the commencement and progress of CSEP efforts. Consequently, only earthquakes occurring post 2011 have been considered in CSEP, but considerable progress has been made since that time. Many earthquake sequences have happened in the country since 2011, and all of the models installed in CSEP under-predicted these large amounts of seismicity. However, a very useful result to emerge is that hybrid models have been found to out-perform the other models.

The progress of CSEP in China was briefly discussed by Yongxian Zhang. China is a large CSEP region that has experienced many earthquakes since CSEP was initiated there in 2010. The talk provided some interesting insights into the process and challenges of regularly communicating CSEP 5 year, 1 year, and 1 days forecast results to the wider seismological, seismic hazard, and disaster risk reduction communities. A “Wenchuan earthquake 10 years later” symposium will be held in-country next year, and CSEP progress will be showcased at the symposium.

Two globally-based CSEP efforts were described in the latter part of the session. Peter Bird provided an overview of his Global Earthquake Activity Rate (GEAR) model of earthquake forecasts, based on seismicity (non-declustered) and geodetic data. The results of this study show that the GEAR1 hybrid model outperformed the other global models tested, which is similar to conclusions reached in a similar analysis a few years ago. In general, the CSEP forecasts performed best in the high seismicity interplate areas, but poorly in areas of intraplate earthquakes. Oklahoma earthquakes were highlighted as examples of the latter, which could technically be discounted due to being induced. However the occurrence of natural intraplate earthquakes (e.g. Australia) shows that CSEP has a lot of work to do in intraplate areas, and will involve a lot of thinking beyond the normal input datasets and methods. In the interplate areas, testing over successive testing periods is needed in order to verify the success of the model in these areas. Max Werner continued the global CSEP overview by showing the sensitivity of test results to such parameters as seismicity grid resolution and depth of seismicity. The GEAR1 model was shown to provide the best information gain of all the models tested.

Session 3: CSEP Results II: Evaluations of Short-Term Models
Moderator: M. Gerstenberger Reporter: N. van der Elst

Camilla Cattania presented an evaluation of physical, statistical and hybrid models during the 2010-2012 Canterbury earthquake sequence. The experiment uses the Canterbury sequence as racetrack for competing forecasts in a similar manner to the retrospective 1992 Landers earthquake sequence experiment, in which statistical models (STEP and ETAS versions) competed against physics-based models (Coulomb/rate-state (CRS). In the Landers experiment, physics-based models performed poorly. The Canterbury experiment pitches smoothing kernel models, ETAS and STEP models against improved physics-based models. Smoothing models don’t work well; ETAS and CRS models perform better, with CRS

models close to or better than ETAS models when measured by an information gain per earthquake. Hybrid models perform well, and CRS models do better than benchmark ETAS models – maybe because they always contain some fault information. To make physical models competitive, receiver uncertainty must be included, and refined grid needed. These appear to be important ingredients in explaining the improvement of the CRS models since the Landers study.
Bruce Shaw recommended isolating the Coulomb component by masking out the near-field (where Coulomb stress is poorly resolved) so we're confident that the Coulomb component is providing the predictive skill. Using optimally oriented planes or uncertainty in near-field is misleading because prediction success does not reflect the physical model itself. Rate-state friction theory also predicts that rates are multiplicative, so instantaneous intensity penalizes the model in places with low background rate. Ned Field questioned how much complexity is 'useful' for OEF purposes.

Max Werner presented a status update of 1-day model results in California. The inventory of results revealed issues: some test results were missing, some test results were wrong. CSEP therefore needs to verify, reprocess and reproduce the results database. Available reliable results show that updated smoothed seismicity models (e.g. PPE) do as well as ETAS and STEP when no big aftershock sequences happen. Over the long term, however, clustering models strongly outperform time-independent smoothed seismicity models. Non-parametric smoothing models currently perform very well: they forecast persistence (namely that tomorrow is similar to today) but can adapt to earthquake sequences (unless a large quake happens right before midnight. A Bayesian ensemble model places significant weight only on K3 and K3+ETAS after very little time (months), as the other 1-day models are downweighted because of surprising earthquakes. Bayesian model averaging, however, is not a conservative weighting strategy (more of a model selector rather than blender), so maybe other weightings (optimized for other forecast metrics, say) could do better. Downweighting correlated models does not produce greater skill (in this case) than straight Bayesian weighting. Camilla Catania commented that we should all be optimizing to the same format/geometry/likelihood function to compare the specific ingredients of the models. Jiancang Zhuang asked how missing aftershocks are dealt with in these short-term forecasts; for simplicity, incompleteness is not dealt with (but at M3.95+, their effect is unlikely to be dominant). Yosi Ogata commented that we keep saying Coulomb Rate State, but we are actually only talking about coseismic static stress change rather than complete Coulomb stress (dynamics, fluids, postseismic slip, triggered aseismic slip…). One could use slow slip areas, and induced seismicity areas to test the physical ingredients in a new apparatus, and maybe learn a lot.

Matteo Taroni presented 1-day forecast results from CSEP Italy. He shows that the ETAS model outperforms STEP in Italy (STEP performs better than the very smooth ETAS v1 by Zhuang in California, but STEP has less predictive skill than more granular ETAS models). The daily number forecast works well at predicting 1 event per day, but "fails" any time there is more than 1 event. This is probably related to the Poisson assumption of distribution of numbers. The ETES model needs to be recalibrated, it clearly contains a bug because it overpredicts substantially. Matteo also showed ensemble models, in which he combined models by the information gain exp(LL/N) rather than the probability gain exp(LL) to suppress some of the variability. The ensemble performs less well than the ETAS model. Lessons learned include that the Poisson hypothesis is too naïve for 1-day tests (there is too much variability); that ETAS is a good model; that STEP may be too granular (spatially compact) when dealing with the Emiglia-Romana earthquake sequence; that an ensemble is a good choice a priori (and not much worse than ETAS a posteriori). Matt Gerstenberger suggested we evaluate the observed catalog with the actual likelihood function produced by the ETAS model, rather than evaluating Poisson distance between observed and expected number. Warner Marzocchi commented that we need to come up with a weighting scheme that works well both for a group of roughly equal models, and also when one model is clearly best. Straight averaging is good for equal models, Bayesian weighting is more severe - more of a model "selector" - when one model is much better than the others.

Hiroshi Tsuruoka presented 1-day model results from Japan. All models fail on active days (similar to Italy results). The Poisson assumption contributes to these failures because its range is too narrow compared to both observations and most model forecasts. ETAS/ETAS/HISTETAS models perform best. A 3D experiment uses a Beta function for the earthquake depth distribution.

Ogata presented forecasting developments in Japan: 2d and 3d spatial models and models with non-uniform b-values. He presented a 3D model needed beneath Kanto (where 3 plates converge). The Bayesian HISTETAS with spatially variable parameters was presented and its parameters interpreted: spatially variable b values help for the region as a whole, but outliers cause significant failures. Ogata suggested evidence for b-value differences between mainshocks and aftershocks. Warner Marzocchi commented that defining aftershocks can bias the b-value of that population. Max Werner commented that models with spatiotemporally varying b-values are quite different from tests of models that are magnitude indifferent (ie all use the GR distribution). Non-GR distributions in UCERF3-ETAS, for example, suggest very different triggering potential depending on ripeness/characteristic-ness. These models need to tested urgently.

During discussion, Max Werner suggested CSEP assess how the 1-day forecast performances vary with tectonic region. Are there any significant/meaningful differences? What should we use as a benchmark/reference model? Warner Marzocchi suggested to use the global Kagan/Jackson 1-day model because it covers the globe consistently. However, the global model runs on the CMT catalog, while regional models use local magnitude scales. Also, the global models forecast above M5.95 and are very smooth, but at least they are not spatially uniform. Yan Kagan's PDE-based global forecast uses a lower threshold of M4.8, but the model is not in CSEP yet. TripleS is another option as it is easy to implement and a transparent code available. Dave Jackson suggested all assumptions that go into each model should be specified, so we can compare success by assumption, rather than model. Conclusions of the discussion**:** 1) Replace Poisson distribution in likelihood tests. Use a more sophisticated reference model than uniform Poisson. 2) Investigate optimal weighting strategies for ensemble models. Bayesian might not be the most useful. 3) Test the physical ingredients, not the statistical distributions. Find other CSEP playgrounds (slow slip, intrusion, injection) to test physical ingredients in places where there are fewer competing triggering agents.

Session 4: Wrapping up CSEP 1.0
Moderator: P. Maechling Reporter: F. Silva

Participants identified a priority question: How do 1-day models perform across different regions? A good starting point for a comparison across tectonic regions is a 1-day global model. Issues include that global models run on the gCMT catalog and at high magnitude threshold M5.8+ and thus produce a very smooth model. Yan Kagan has produced time independent model based on ANSS catalogs. The answer may depend on the specific question. Another option is TripleS.

Participants discussed further opportunities with the available data within CSEP1.0. Data is available since Aug 2007 (10TB). Accomplishments of CSEP need to be identified and disseminated, as well as scientific advances. Some of the major achievements mentioned include: CSEP results influenced how hazards models are constructed, namely how seismicity source models are constructed in several countries. California and Italy are examples of this. In addition, testing methods of CSEP have been adopted by researchers outside CSEP. The Helmstetter model indicates the utility of small quakes for anticipating-larger quakes. Surface strain rate models are competitive predictors of seismicity rates to smoothed seismicity models.

Where/how should CSEP1.0 results be published? BSSA or SRL were identified for special issues. Further options include 1 overview paper with short/long term results combined; a white paper with the CSEP philosophy; an EOS project update/priorities that could be written by the group; a summary paper in a higher impact journal; a special issue; other venues: Natural Hazards, EQ Spectra (engineering seismology, different audience).

How can we make CSEP data available within and outside group? There are 10TB of data in California testing center. Should these data be accessible in raw form or should/need they be curated/quality-controlled first? Data sharing/usage would be more useful if we knew more about input models and assumptions. This is relatively easy for RELM experiment. Would it be a good idea if someone looked at how many earthquakes to have in a test to make it useful? But, this may not be something we can gener-

alize; perhaps predicting number is more important than EQ/km2. Models we have share several features, perhaps we should extract important characteristics from these models to answer questions.

What are the opportunities right in front of us? Ideas discussed include template-based catalogs; whether globally Mw 6 are good at predicting Mw 7, 8+ (e.g. use GEAR1 and perform self-consistency tests); investigate consistency across model regions; invite all to test their models in China, where lots of data are available and present an opportunity to test models.

Session 5: The Next Phase: CSEP 2.0
Moderator: W. Marzocchi Reporter: A. Strader

The "CSEP 2.0" component of the SCEC CSEP workshop focused on how the CSEP testing center's capabilities and priorities will evolve past the current state. A series of presentations outlined future forecasting experiments that could be implemented into CSEP, as well as methods to incorporate new types of input data and utilize currently available data and experiment results.

Philip Maechling presented suggestions for how the testing center system and software can adapt to increasing amounts of data and forecasting experiments in the next years. Over the first ten years of CSEP operations, the CSEP testing center has accumulated 10 TB of data, including forecast experiment results and software source input earthquake catalogs. With the expansion of testing regions, testing classes and types of forecast models, the SCEC CSEP operational testing center is reaching its computational and file-management limits. Possible solutions include distributing operational CSEP processing onto multiple servers (single server to multi-server to workflow phase), maintaining a current list of all earthquake catalogs, forecasts, and results produced by the system, and improving automated processing of error detection, retry and recovery. Complete CSEP data archives should also be preserved and curated as an approved research dataset, including clearly identified and retrospectively updated earthquake forecasts and evaluation results.

Yosihiko Ogata presented a method for computing the probability that the first event in an earthquake cluster is a foreshock. In previous prediction models, the magnitude distribution, defined by one b-value throughout an entire region, has been assumed separable from the space-time component of the ETAS model. Ogata showed that the b-value is dependent on the earthquake location and seismic history at a given location. These dependencies can be utilized to forecast future earthquake magnitudes based on seismicity characteristics such as sequential b-value changes, space-time clustering intensity, precursory swarms, and seismicity quiescence or activation. By classifying historical seismicity in Japan as isolated events or members of earthquake cluster, it is possible to forecast the probability that the magnitude of the next earthquake will be greater than the preceding earthquakes by a certain threshold.

Shunichi Nomura presented a modified version of the aforementioned foreshock discrimination model, where foreshocks and earthquake swarms were not classified by magnitude differences and mainshock magnitudes as well as locations are forecasted for a forecasting period of 30 days from the last earthquake. Using single-link clustering, seismicity clusters are then identified, and features such as increasing seismicity over time and seismicity propagation can be extracted. Such features are incorporated into a logistic regression model to calculate the probability of a mainshock occurring during the next 30 days.

Anne Strader presented (on behalf of Danijel Schorlemmer) some suggestions for expanding the variety of forecasting experiments conducted through the CSEP testing center. The CSEP testing centers, in collaboration with the Global Earthquake Model (GEM) testing center have the capability to evaluate seismicity models and ground-motion models, including GMPEs and intensity prediction equations. Current forecasting experiments are based on a grid of seismicity rates or earthquake probabilities, resulting in difficulty utilizing new authoritative data, non-standard forecast types, or conducting experiments on non-gridded testing areas. In the future, CSEP should expand to conduct tailored forecasting experiments; for example, analyzing the forecasting power of b-value anomalies and discriminating foreshocks (testing the method previously introduced by Y. Ogata). The GFZ forecasting group also introduces the Dynamic Risk Quantification (DRQ) project, which aims to develop and refine data-driven global hazard and risk models using CSEP testing results at the seismicity model development stage. Additionally, fully testable GMPEs can be built and coupled with dynamic exposure models based on OpenBuildingMap.

David Rhoades presented testing of simulation-based seismicity models as a solution to avoiding the Poisson assumption in evaluating clustering models. In particular, forecasts sometimes fail the N-test due to underestimated uncertainty from the Poisson assumption. The factorial term in the discrete Poisson likelihood function also causes discrepancies in forecasts' log-likelihood scores based on the study region's gridding scheme, which is arbitrarily selected and lacks a physical basis. Consistency tests of simulation-based models allow for a wide range of tests, with the ability to evaluate earthquake-clustering behavior. Using a general approach, a statistic or distribution of statistics is calculated from the real earthquake catalog and compared with the statistic's distribution from multiple model simulations. It is possible to combine results from multiple test periods while omitting the Poisson assumption by observing whether p-values over multiple test periods are uniformly distributed on the interval (0, 1). Deviations from the uniform distribution indicate how models are inconsistent with observations (for example: over/underprediction and over/under-dispersion).

Nicholas van der Elst introduced Turing-style tests to evaluate the consistency of synthetic catalogs produced from the UCERF3 seismicity model with observed seismicity. Initial results show that the observed California earthquake catalog has a higher b-value for small magnitudes than UCERF3-ETAS. Furthermore, UCERF3-ETAS generates synthetic earthquake catalogs that are more spatially diffuse than observed seismicity. UCERF3-ETAS also produces less inter-sequence aftershock variability than observed, due to using one set of direct Omori parameters.

Dave Jackson presented prospective test results for the 1995 WGCEP earthquake forecast. The model categorized earthquakes as either characteristic "cascade" events, or by their epicentral location in one of 65 seismotectonic regions. Two models were evaluated: one (the "preferred" model) assuming quasi-periodic events on major faults, and the other (the "alternate" model) assuming Poissonian seismicity. Both models failed the N-test by overpredicting the number of Mw>=6 earthquakes from 1995-2017. No characteristic earthquakes occurred during the observation period, causing the preferred model to fail at the 95% confidence level, while the alternate model barely passed. However, the spatial and magnitude distributions for both models were consistent with observed seismicity (although there were only three observed events).

D. Significance

E. References