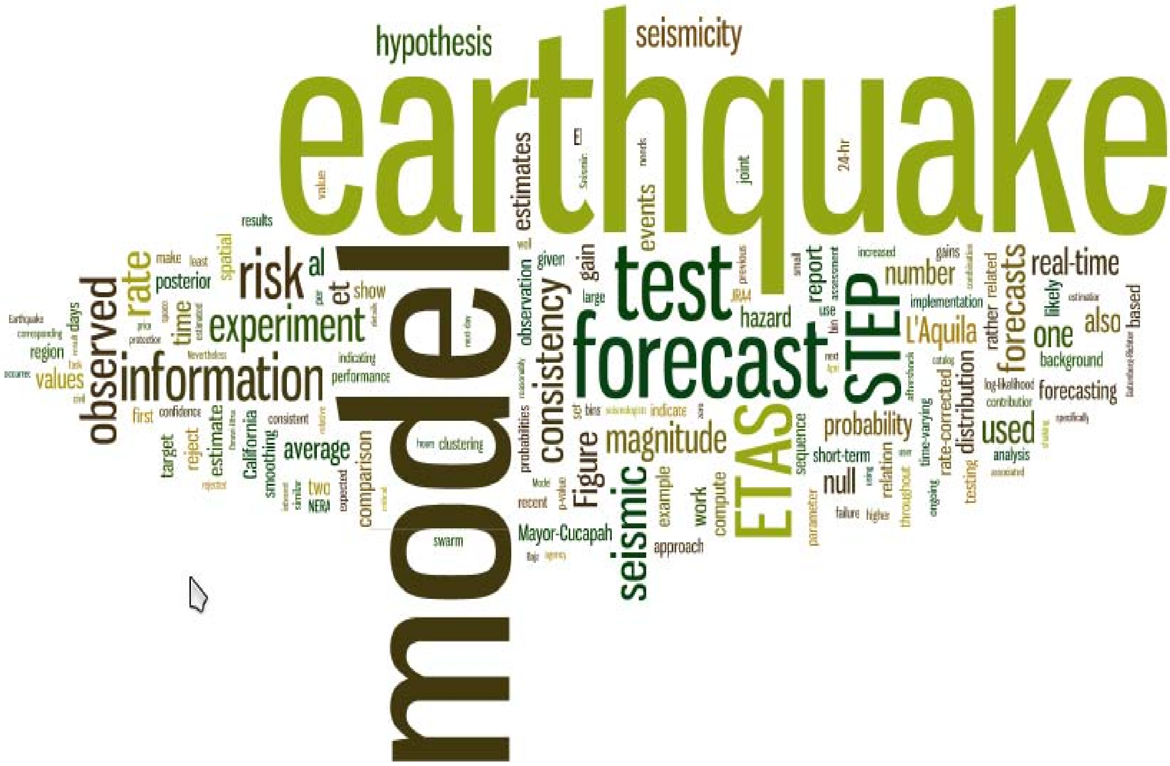
Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation

Report

State of the art and user needs in real-time risk assessment and decision support

Activity:	*Real-time seismic risk assessment and decision support*
Activity number:	*WP14, Task 14.1*
Deliverable:	*State of the art and user needs in real-time risk assessment and decision support*
Deliverable number:	*D14.1, Version 12.4*
Responsible activity leader:	*J. Douglas Zechar*
Responsible participant:	*ETH Zurich*
Author:	*J. Douglas Zechar*

Seventh Framework Programme
EC project number: 262330


earthquake
model
forecast
test
STEP
ETAS
information
experiment
observed
risk
rate
hypothesis
seismicity
seismic
magnitude
consistency
forecasts
real-time
number
L'Aquila
used
one
also
probability
null
Figure
estimates
events
hazard
average
comparison
California
Mayor-Cucapah
forecasting
background
distribution
rate-corrected
time-varying
short-term
values
time
posterior
estimate
reject
target
smoothing
clustering
swarm
approach
work
compute
example
sequence
parameter
report
gain
given
observation
probabilities
recent
show
performance
indicating
confidence
first
region
days
make
spatial
results
al
et
two
NERA
joint
24-hr
likely
based
rather
related
analysis
testing
throughout
relation
failure
higher
implementation
log-likelihood
contribution
gains
increased

ABSTRACT

Seismic hazard is the natural threat posed by earthquakes, and to assess seismic hazard one must estimate the probability of rupture and the corresponding anticipated ground motions. When seismic hazard is coupled to the built environment and information about the spatial distribution of people, one can estimate seismic risk. Given the large uncertainties associated with each of these estimates, and the fact that such estimates are amplified when trying to perform a real-time analysis, there is little to say about the state of the art of real-time seismic risk assessment: at present, there are no end-to-end systems for estimating real-time and time-varying seismic risk. Therefore, in this report I detail recent advances related to some of the important ingredients in the real-time seismic risk recipe. In particular: I summarize a retrospective time-varying evaluation of seismic risk related to the L'Aquila earthquake sequence; I report results from an ongoing next-day earthquake forecast experiment in California involving two models that will likely form the basis for JRA4 development; I describe some existing related efforts and clarify their relation to NERA JRA4; and I make a few comments regarding user needs.

(This report is intended to be a living document, at least throughout the NERA project cycle, and therefore its version is denoted by the year and month of its most recent update.)

Keywords: NERA JRA4, seismic risk, seismic hazard, earthquake forecasting, ShakeMaps, SAFER, ETAS, STEP, user needs

1. INTRODUCTION

Seismic risk may be defined as the potential for earthquakes to cause injuries, deaths, and damage to man-made structures (e.g., buildings, bridges, dams, etc.). From this perspective, seismic risk varies in space and time and is a function of the natural hazard posed by earthquakes coupled with anthropogenic considerations such as population density, building classification and the distribution of structures and infrastructure, and construction quality. Seismic hazard in turn depends on the likelihood of earthquake nucleation and factors that include the distance from the earthquake, the direction in which the earthquake rupture propagated, the depth of the earthquake, and soil/rock type.

Even without a fundamental physical understanding of earthquake nucleation, there are certain elements of earthquake risk that are reasonably well-known—for instance, that earthquakes cluster in space and time, that large earthquakes occur less frequently than small earthquakes, and that large earthquakes are followed by "aftershocks," smaller events that occur near and soon after the "mainshock." Moreover, we have some understanding of how epicentral distance, hypocentral depth, soil type, directivity, and other factors influence shaking caused by an earthquake. And from engineering analyses we have an idea of how different building types react to shaking with different frequency spectra. Therefore, one can estimate seismic risk. And given recent advances in short-term earthquake forecasting, one can even consider the problem of real-time seismic risk.

Nevertheless, as of this writing, it appears that no country, agency, or institution has implemented a system for real-time seismic risk assessment. Therefore, in this report, I will describe the status and recent progress related to some of the components that would contribute to real-time seismic risk assessment and highlight some of the relevant difficulties and opportunities associated with this work. I summarize one proposed approach to real-time risk that was developed after the 2009 L'Aquila earthquake; I report results from an ongoing next-day earthquake forecast experiment in California; and I discuss the relation of some existing products to the problem of real-time earthquake risk assessment.

2. SHORT-TERM EARTHQUAKE RISK

Following the Mw 6.3 L'Aquila earthquake on 6 April 2009, van Stiphout *et al*. (2010) pondered the question "Are short-term evacuations warranted?" Using the sequence of earthquakes in the months preceding the L'Aquila earthquake, they applied a time-varying model (Reasenberg & Jones 1989) to estimate the rate of impending earthquakes. (Although this model was developed using earthquakes in California, Lolli & Gasperini (2003) estimated parameter values that could be used to analyze earthquake sequences in Italy). The output of the rate model was used as the input to a ground motion model (Michelini *et al*. 2008), which yielded time-varying estimates of seismic hazard. These seismic hazard estimates were then converted to a probabilistic loss curve using the QLARM approach (Trendafiloski *et al*. 2011). The loss curve describes the estimated risk at a given place and

time, and van Stiphout *et al*. analyzed how the estimated risk at L'Aquila changed throughout the sequence of events (see Figure 1).

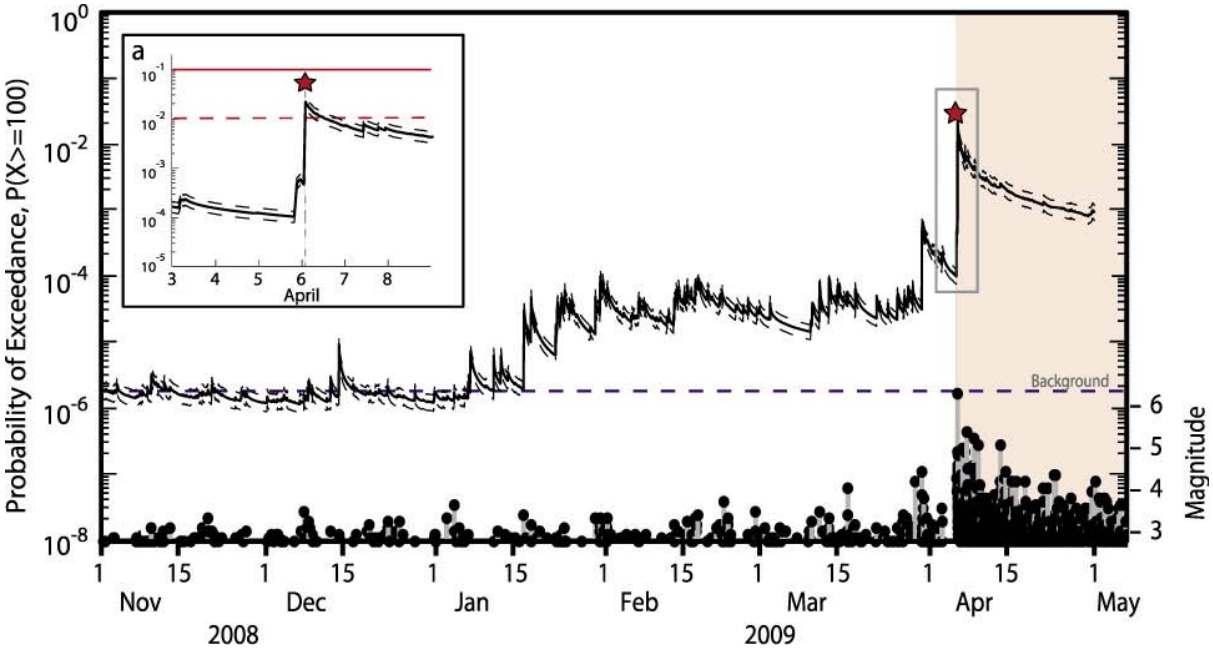


Figure 1. Time-varying loss estimates in the L'Aquila region. Probability of exceeding 100 fatalities in the next 24 hours. Earthquakes are shown with their magnitudes. Modified from van Stiphout et al. (2010)

As of this writing, the 2010 study of van Stiphout *et al*. seems to be the only attempt to work from the earthquake catalog all the way through to time-varying forecasts of seismic risk. Even in this relatively sophisticated study, several simplifying assumptions had to be made, and one could question the decisions made at nearly every step of the risk estimation process. Nevertheless, this study illustrated the procedure well, and it was useful in answering a specific question of interest: for the conditions of L'Aquila, a full evacuation was never worthwhile.

In the wake of the L'Aquila earthquake and related to the EU-sponsored SAFER (Seismic eArly warning For EuRope) project, scientists at the INGV (Istituto Nazionale di Geofisica e Vulcanologia) in Rome began producing “aftershock” forecasts for internal use; these forecasts were presented as maps of the expected distribution of earthquakes in the next 24 hours. Figure 2 is an example forecast for 12 April 2009, a few days after the “mainshock.” This forecast was based on the ETAS (Epidemic-Type Aftershock Sequence, Ogata 1988) model, and the details for this implementation were reported by Marzocchi & Lombardi (2009). ETAS is a statistical model that incorporates the first-order empirical observations of seismicity: the Gutenberg-Richter magnitude relation and the Omori-Utsu decay of aftershock productivity. This model represents seismicity as a combination of background seismicity (conceptualized as those events thought to be directly attributable to tectonic loading) and triggered earthquakes. Each earthquake above a minimum magnitude can stochastically trigger other earthquakes; the number of events that each earthquake can trigger is a very important model parameter. Owing to its conceptual elegance, the existence of many software tools for its estimation and simulation, and its demonstrated ability to fit

many data sets well, ETAS has become a standard model for short-term earthquake forecasting.

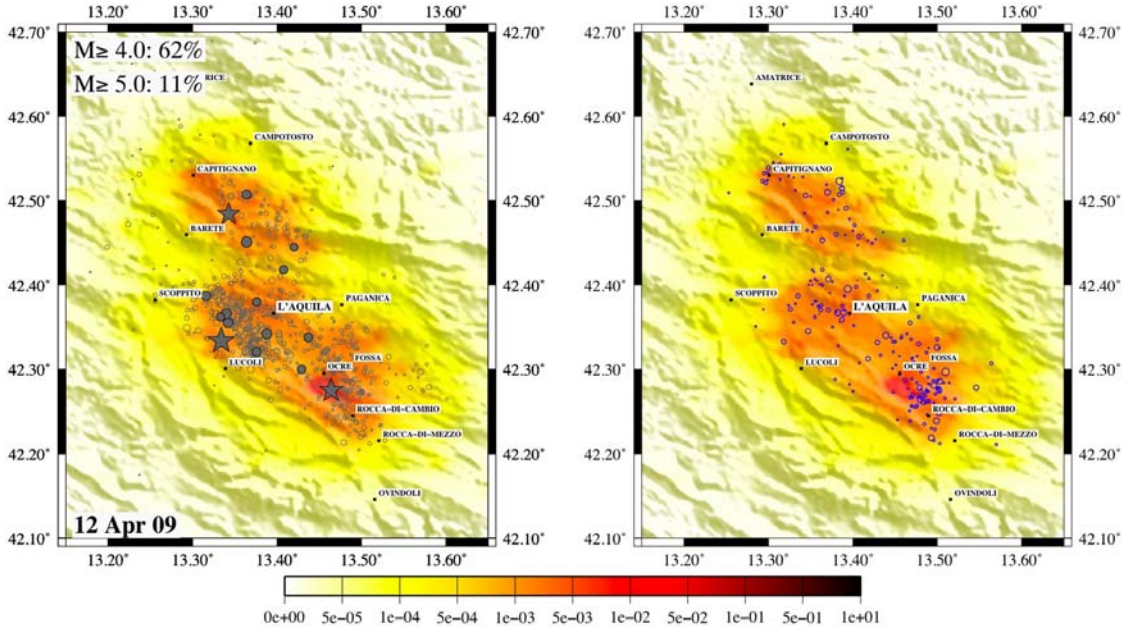


Figure 2: Expected number of M4+ earthquakes per square kilometer for 12 April 2009. Map on left shows events prior to making the forecast, map on right shows events during the forecast period. From Marzocchi & Lombardi (2009).

Despite its relatively long history, the ETAS model has rarely been applied in prospective settings, and the L'Aquila sequence was perhaps the first time it was used during a crisis. Nevertheless, as Marzocchi & Lombardi (2009) report, the forecasts were consistent with the observations, indicating a degree of usefulness. Moreover, the Italian civil protection agency used the forecasts, an act that is likely also a first for the ETAS model.

Inspired by the L'Aquila event and its aftermath, there has been increased interest in real-time hazard and risk estimation in Italy, and the L'Aquila earthquake was also the primary impetus for an international commission to convene and provide an overview and a road map for what is now being called “operational earthquake forecasting.” The international commission constructed a lengthy report (Jordan *et al*. 2011) that will likely serve as a guide for much of the research within Work Package 14. Some seismologists are of the strong opinion that, during a seismic crisis, seismological agencies and their workers should not refuse to make information available solely because the information is uncertain, because associated probabilities are quite small, or because seismologists fear that what they say will be proved wrong by future observations. Rather, seismologists should be working in tandem with decision-makers, providing them as much information as is available, and allowing the decision-makers to choose the appropriate actions.

It should be noted that other groups at INGV are also developing next-day earthquake forecasts, at least one of which combines the ETAS model and concepts from rate-and-state friction (Falcone *et al*. 2010). Moreover, the civil protection agency has expressed interest in continuing to work with researchers on operational earthquake forecasting, specifically to use forecast information internally (rather than broadcasting this information to the public). This

approach is conceptually similar to what is done in China, where seismologists convene for an annual consultation to discuss what they believe to be the regions in China most likely to experience damaging earthquakes in the coming year; this information is used to decide resource allocations, to prompt preventative mitigation actions, etc.

3. SHORT-TERM EARTHQUAKE FORECASTING

Although, as mentioned in the preceding section, there have been few efforts to systematically estimate real-time seismic risk, short-term earthquake forecasting has recently received increased attention. In this section, I describe an ongoing experiment in California to evaluate two short-term earthquake forecast models. It is likely that our Work Package will either use one of these models, both of these models, or some combination or variant of these models. The remainder of this section has been modified from a manuscript that I am preparing with co-workers (thus the use of the first-person plural rather than the first person singular).

Since September 2007, we have been conducting an experiment in California to test and evaluate two models that yield next-day seismicity forecasts: the Epidemic-Type Aftershock Sequence model (ETAS) and the Short Term Earthquake Probability model (STEP). ETAS and STEP are statistical models that incorporate well-studied empirical relations such as the Gutenberg-Richter magnitude distribution and the Omori-Utsu relation for the temporal decay of “aftershock” activity. The objective of the experiment is for each model to generate a forecast that specifies the expected number of earthquakes for the next 24 hours in bins specified by ranges of latitude, longitude, and magnitude. The spatial cells are 0.1 degrees by 0.1 degrees, and the models target all earthquakes with magnitude greater than or equal to 3.95 in bins of 0.1 units. We have examined the first three years of results; during this time, 270 target earthquakes occurred in the testing region, including a swarm in Baja California throughout February 2008 and the April 2010 M 7.2 El Mayor-Cucapah earthquake.

Through the testing component of this experiment, we discovered a flaw in the STEP implementation; it was corrected, and we evaluated forecasts based on the revised implementation. To evaluate the models, we conducted investigations of consistency (each model relative to the observation) and comparison (each model relative to the other). Based on joint log-likelihood measures, we found that both models are reasonably consistent with the the observed number of earthquakes and the magnitude distribution; ETAS is also reasonably consistent with the spatial distribution and the joint space-rate-magnitude distribution, while STEP is not. To compare the models, we computed the rate-corrected average information gain per earthquake; this is the natural logarithm of the probability gain. STEP had an average information gain of 0.43 relative to ETAS; in other words, STEP had, on average, a 54% higher forecast rate than ETAS in bins where target earthquakes occurred, indicating that, despite its inconsistencies, STEP was superior in this experiment.

3.1. Testing region and notation

In Figure 3 we show the testing region for this experiment; it is the same used in the Regional Earthquake Likelihood Models (RELM) experiment (Schorlemmer & Gerstenberger 2007, Schorlemmer *et al*. 2010). Indeed, this experiment is identical to RELM with two important exceptions: we consider M≥3.95 (rather than 4.95) earthquakes, and the forecast horizon is 24 hours (rather than 5 years).

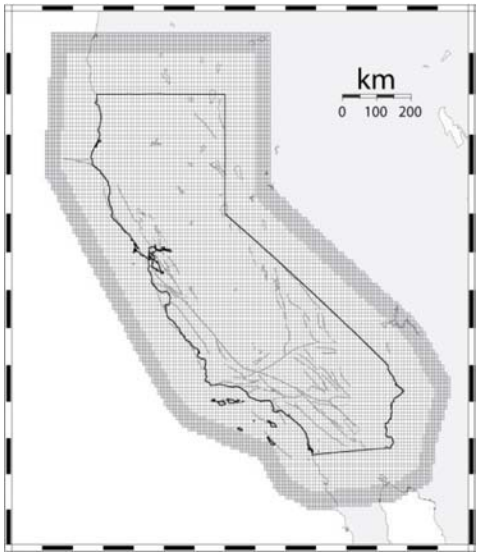


Figure 3: Testing region used for the 24-hr earthquake forecast experiment. From Schorlemmer & Gerstenberger (2010).

Each 24-hr forecast is specified as a 3D vector, $\boldsymbol{\Lambda}$, comprised of 391,782 bins defined by ranges of latitude-longitude-magnitude. In each bin, the model forecasts the number of earthquakes expected in the next 24 hours; in the *i*-*j*-*k*th bin, we denote this expectation as λ_{ijk} and each bin's forecast is characterized by Poisson uncertainty. Likewise, each 24-hr observation is denoted as 3D vector $\boldsymbol{\Omega}$, where w_{ijk} is the number of earthquakes observed in the *i*-*j*-*k*th bin. See Figure 4 for cartoon representations of the forecast and the catalog for each 24-hr period.

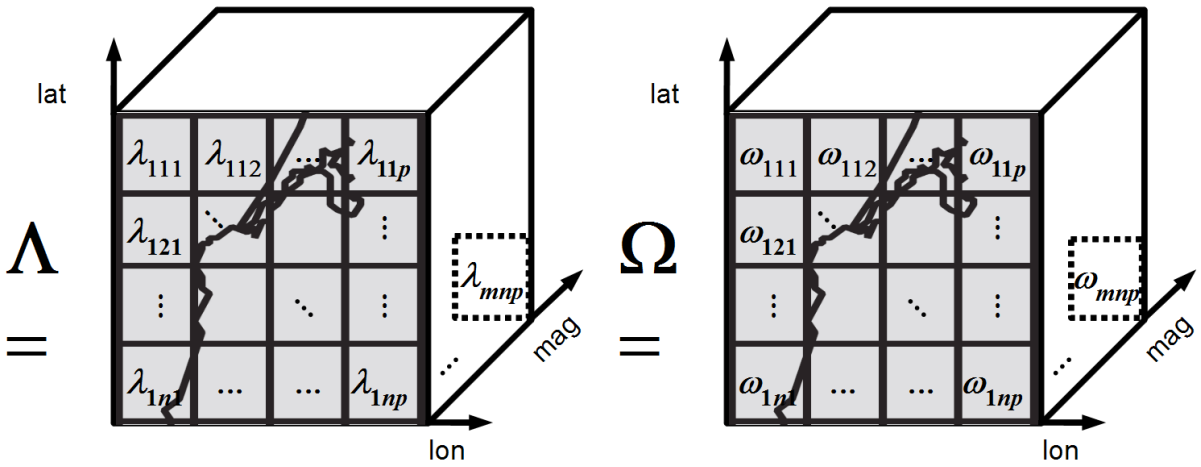


Figure 4: Cartoon representations of forecast and binned observations

3.2. Models

The ETAS implementation used in this experiment has been described in detail by Zhuang (2011). Under this model, the expected number of earthquakes at time/lon/lat *t*/*x*/*y* depends on contributions from previous earthquakes:

$$\lambda(t,x,y)=\mu(x,y)+\sum_{i:t_i<t}\xi(t,x,y;t_i,x_i,y_i,m_i)$$

where $m(x,y)$ is the background seismicity rate at point (x,y) and $\xi(t,x,y;t_i;x_i,y_i,m_i)$ is the contribution to point (t,x,y) from the previous earthquake at (t_i, x_i, y_i) with magnitude m_i. This contribution is taken to be separable into space, time, and magnitude components; the spatial contribution is isotropic and magnitude-dependent, the temporal contribution follows an Omori-Utsu decay relation, and the magnitude contribution follows the Gutenberg-Richter relation. For full details of this implementation, including computational issues related to the difficulty in estimating model parameter values, we refer the reader to the article by Zhuang (2011).

The STEP model follows the description given by Gerstenberger *et al*. (2005). It too is based on a background seismicity model, but rather than estimating this from previous seismicity, the background model is the source model of the National Seismic Hazard Mapping Project. When a new earthquake of sufficient size occurs, clustering models of various complexity are invoked. Under the simplest clustering model, the rate of "aftershocks" with magnitude m at time t is given by

$$\lambda(t)=\frac{10^{a+b(M_m-m)}}{(t+c)^p}$$

where M_m is the "mainshock" magnitude and all other values are constants; again, this is a mixture of the Omori-Utsu relation and the Gutenberg-Richter relation. Once an aftershock sequence exceeds a certain productivity, a sequence-specific clustering model is used, where the constants above are re-estimated for the individual sequence. At the same time, the model attempts to estimate sequence-specific spatially-varying parameter values. The resulting forecast is a weighted combination of these three clustering models and the background model; details were described by Gerstenberger *et al*. (2005). When the summed contributions from the clustering models are lower than the background model, the background model is used.

The STEP and ETAS models are conceptually similar and based on the same first-order empirical seismicity relations. On the other hand, STEP forecasts include somewhat more complex clustering models, and ETAS is characterized by computational complexity because it tries to estimate so many parameter values. One can also think of ETAS as simpler because all earthquakes and all points in the forecast are treated the same; there are no attempts to make sequence-specific parameter estimates.

3.3. Methods

We applied two classes of forecast evaluation: one for consistency of the forecast with the observation, and one that compares the performance of each model.

3.3.1. Consistency

We applied four consistency tests based on the joint log-likelihood of the observation given the forecast. For each model, for each 24-hr period, we compute the Poisson joint log-likelihood by summing over all bins:

$$L(\Omega|\Lambda)=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}\left(-\lambda_{ijk}+\omega_{ijk}\log\left(\lambda_{ijk}\right)-\log\left(\omega_{ijk}!\right)\right)$$

We want to test the **forecast consistency null hypothesis** that the observed catalog is consistent with the forecast or, more specifically, that the observed joint log-likelihood is a result of the forecast being the correct model for seismicity. To test this hypothesis, we simulate 10,000 catalogs from the forecast and compute the joint log-likelihood of each simulated catalog, yielding a set of $\hat{L}$, which is an estimate of the distribution of joint log-likelihoods under the forecast consistency null hypothesis. We then compute the fraction of simulated likelihoods that are smaller than the observed likelihood:

$$\gamma=\frac{|\{\hat{L}|\hat{L}<L\}|}{|\{\hat{L}\}|}$$

This is akin to a *p*-value; if γ is very small, we can reject the forecast consistency null hypothesis with great confidence; in other words, we infer that the observation is inconsistent with the 24-hr forecast. This procedure is known as the L test and evaluates the consistency of the space-rate-magnitude forecast with the space-rate-magnitude observation; we also apply equivalent tests that isolate the rate (N test), space (S test), and magnitude (M test) dimensions. See the article by Zechar *et al*. (2010) for details of each of these tests.

For each day of the experiment, we performed the L test and the N test. On days when no target earthquakes were observed, there is no spatial or magnitude observation, so we only perform the S test and M test on days where at least one target earthquake was observed. The performance of each 24-hr forecast is characterized by the respective quantile scores, and by choosing a critical value, we can reduce the result to a binary decision: to reject or not reject the forecast consistency null hypothesis. Following most other CSEP experiments, we chose a critical value of 0.05, corresponding to 95% confidence. We can then characterize the overall performance of the model by considering the failure rate, or the fraction of days upon which the test was performed (*N*) and the null hypothesis was rejected (*n*). With this failure rate, *n*/*N*, we can test the **model consistency null hypothesis** that the observed seismicity distribution is consistent with the model, or more specifically, that the observed failure rate is a result of the model being representative of seismicity. Given a critical value of 5%, we compute the probability of obtaining a failure rate at least as big as the observed:

$$p=1-F(n;N;0.05)\ ,$$

where $F(x;X;p_1)$ is the cumulative binomial probability of observing x successes in X trials where each trial has a success probability of p_1; n is the number of observed failures, and N is the number of days upon which the test was performed. If p is very small, the model consistency null hypothesis is rejected, indicating that the model is inconsistent with the observed seismicity.

3.3.2. Comparison

To directly compare the performance of the two models, we compute the rate-corrected average information gain per earthquake (Rhoades *et al*. 2011):

$$I_N(\Lambda^A, \Lambda^B) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\lambda_i^A}{\lambda_i^B} + \frac{N_B - N_A}{N}$$

where N is the number of observed target earthquakes, $\mathbf{\Lambda}^A$ and $\mathbf{\Lambda}^B$ are the forecasts to compare, λ_i^A is the rate of forecast $\mathbf{\Lambda}^A$ in the bin of the i^{th} earthquake, and N_A is the total rate forecast of model $\mathbf{\Lambda}^A$.

Here, a rate-corrected average information gain greater than zero favors Model A, zero indicates no preference, and a negative value favors Model B.

This calculation is based on the set of information gains (logarithm of the probability gains) for each observed earthquake; we subject this set of information gains to a Student's t-test and a Wilcoxon signed-rank test. Each of these test the **model comparison null hypothesis** that the model performances are similar, or more specifically that the rate-corrected average information gain is zero. Student's t-test assumes that the set of information gains is normally distributed and the Wilcoxon signed-rank test rests on the weaker assumption that the distribution is symmetric. The results of these T and W tests are the rate-corrected average information gain and a *p*-value for each test, which can be used to reject or not reject the model comparison null hypothesis.

Data

During the 3-yr period considered, 270 target earthquakes were observed. These are shown in Figure 5. The observed earthquakes include a swarm of approximately one dozen small events in Baja California and the El Mayor-Cucapah M 7.2 earthquake and its presumed aftershock sequence, comprising approximately 120 earthquakes. We note that the El Mayor-Cucapah earthquake occurred in the same region as the swarm.

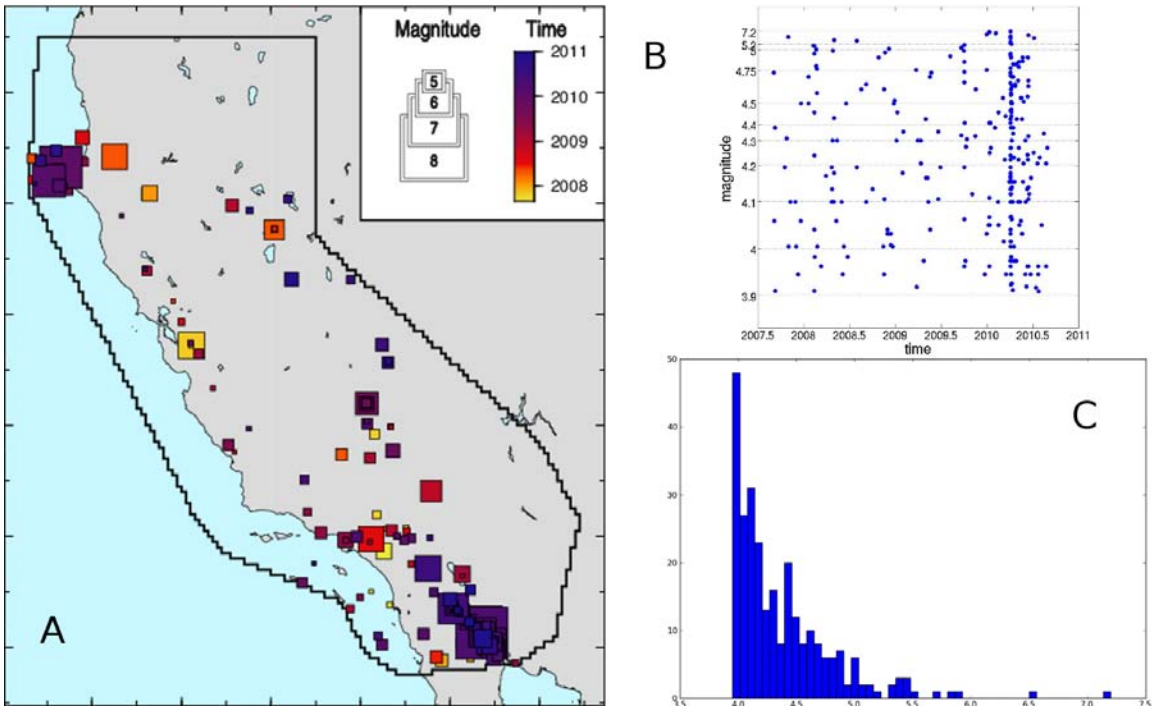


Figure 5: A) map of observed target earthquakes. B) Time and magnitude distribution of target earthquakes. C) Magnitude distribution of observed target earthquakes.

3.4. Results

When reviewing preliminary results of this experiment (during the testing phase), we uncovered an apparent error in the STEP implementation. During the Baja California swarm, STEP forecast rates were increased in the region of the swarm, but not where the events were occurring; rather, the rate increased were systematically offset. The panels in Figure 6 illustrate this error and its correction. Such an error is not uncommon when developing relatively complex earthquake forecast models and requesting a very specific forecast format for a particular region. This simply highlights the importance of the testing component of earthquake predictability experiments (see also the work described by Werner *et al*. (2010), where similar mistakes in CSEP-Italy models were discovered).

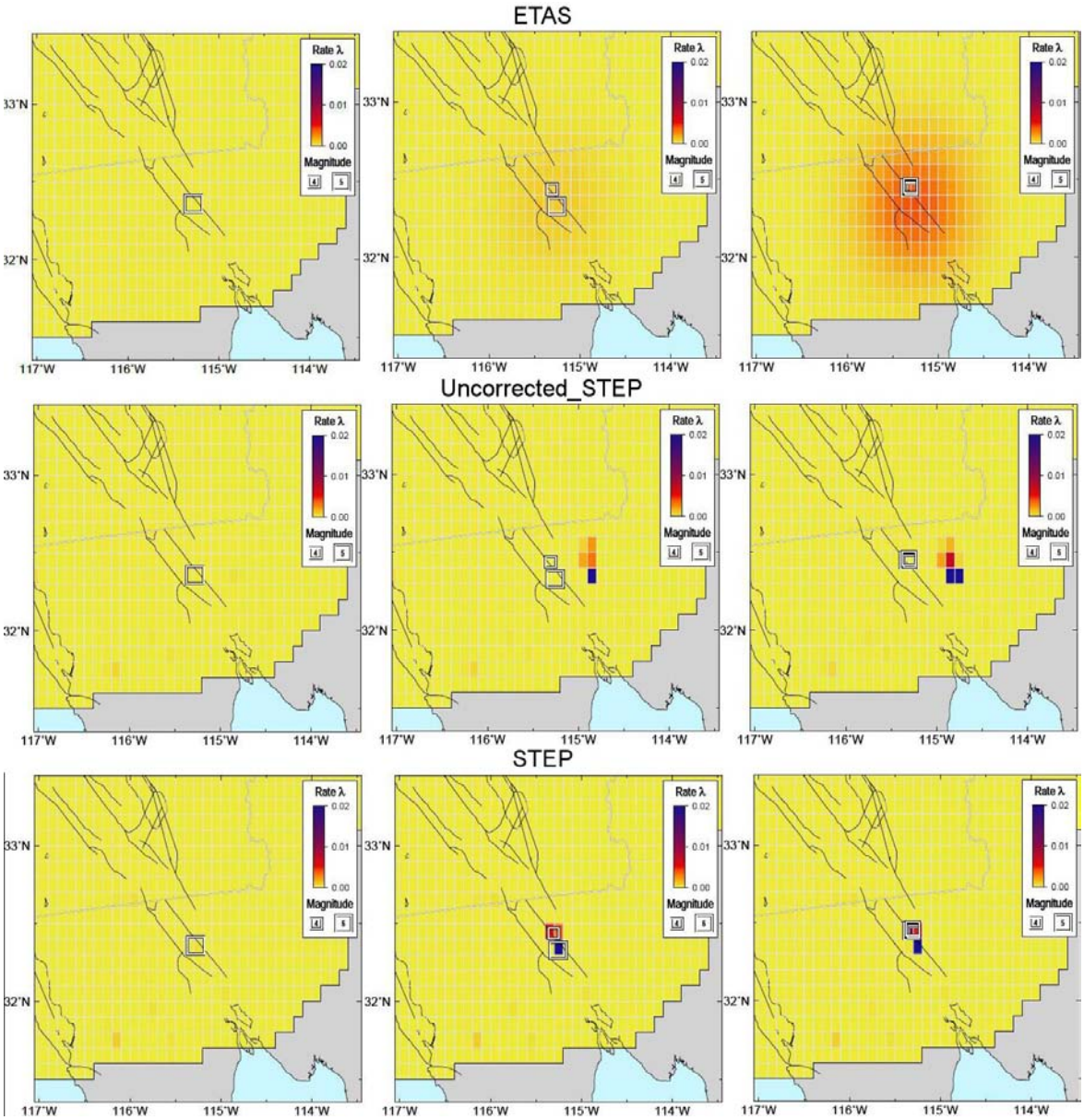


Figure 6: ETAS (top row), STEP (middle), and corrected STEP (bottom) forecasts during the Feb 2008 swarm (9, 11, and 12 February forecasts are shown). The middle row demonstrates the registration problem of the uncorrected STEP implementation, and the bottom row shows the correction.

In Table 1, we report the summary of consistency evaluations. Each cell shows the failure rate and the corresponding p-value (in parentheses). For example, the values in the first cell indicate that the ETAS model failed the N test on 18 days out of the 1096 days considered. The probability of failing at least this often is nearly unity, indicating that the model consistency null hypothesis cannot be rejected. The red cells indicate tests for which we can reject the model consistency null hypothesis with great confidence: the STEP model failed the S test and the L test far more likely than expected, indicating problems with the spatial component of the STEP model. We are currently conducting further analysis to understand the source of these failures. We note that we have also compared each model with a simple time-invariant smoothed seismicity model (TripleS; Zechar & Jordan 2010) and we can say with greater than 95% confidence that both STEP and ETAS are superior.

Evaluation metric	Model	
	ETAS	**STEP**
Rate (N test)	18/1096 (**1.000**)	19/1096 (**1.000**)
Magnitude (M test)	5/161 (**0.909**)	2/161 (**0.998**)
Space (S test)	7/161 (**0.699**)	21/161 (**0.000**)
Space-rate-magnitude (L test)	47/1096 (**0.877**)	77/1096 (**0.002**)

Table 1: Failure rates for each test for both models with corresponding p-value in parentheses.

In Figure 7 we show the cumulative number of earthquakes observed as a function of time, as well as the cumulative number of earthquakes forecast by each model as a function of time. The effect of the El Mayor-Cucapah earthquake is evidenced by the spike in the first half of 2010. Both models match the trend of the observation reasonably well, although the overall rate of STEP is consistently lower than that of ETAS.

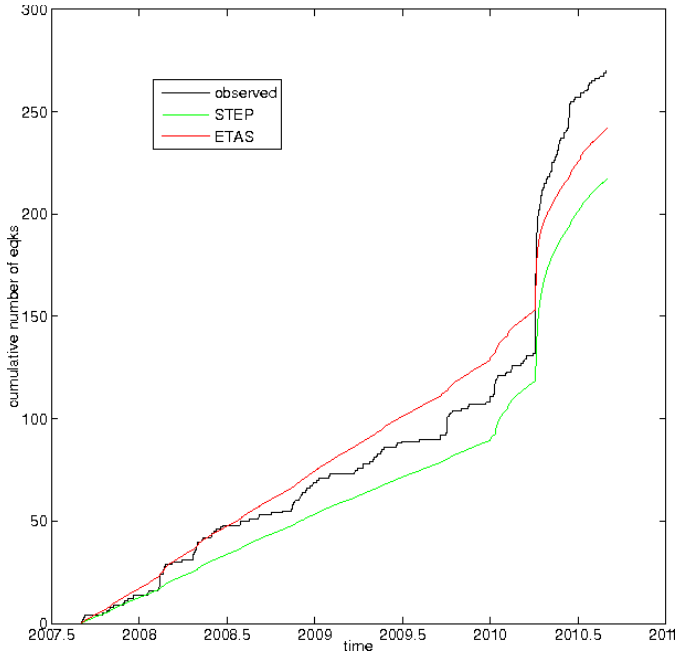


Figure 7: Cumulative number of observed target earthquakes (black) and STEP (green) and ETAS (red) cumulative rate forecasts.

In Figure 8 we show the individual information gains (without any rate correction) for each observed earthquake. The rate-corrected average information gain per earthquake is denoted by the horizontal dashed line. Points falling in the green region ($I > 0$) indicate a higher rate for STEP, and points in the red indicate that ETAS had a higher forecast rate in the bin where this earthquake occurred. The two lowest values correspond to earthquakes in Nevada, likely unrelated to the El Mayor-Cucapah earthquake. The next four lowest values correspond to earthquakes occurring ~65 km NW of the El Mayor-Cucapah earthquake. We are currently conducting further analysis to determine where, when, and for what magnitude earthquakes STEP has an advantage over ETAS.

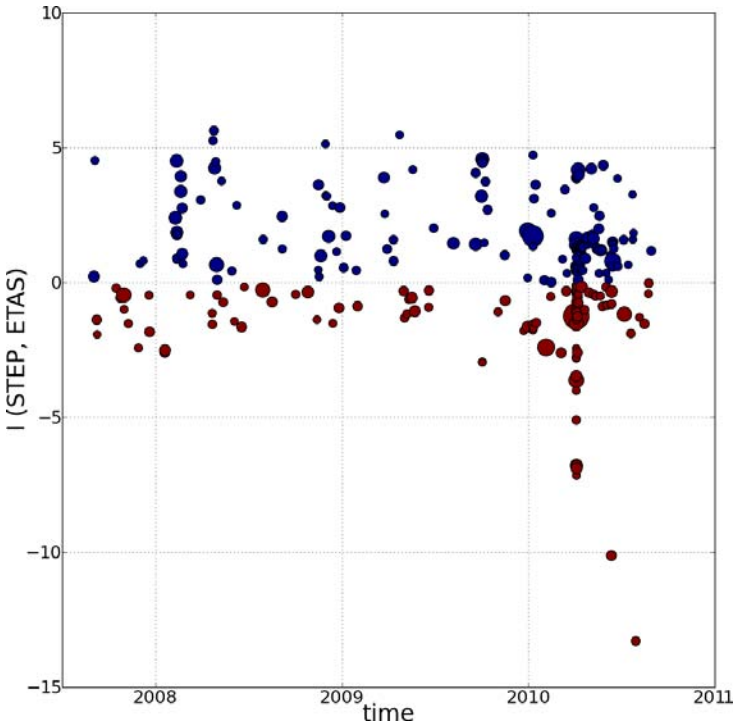


Figure 8: Information gain for each observed earthquake (blue favors STEP, red favors ETAS)

In Figure 9 we show the temporal evolution of the rate-corrected average information gain per earthquake, updated after each 24-hr experiment. Positive values indicate a preference for STEP and negative values indicate a preference for ETAS. The effect of the El Mayor-Cucapah earthquake and the events that followed in its wake are evidenced by a sharp drop in I_N, followed by an epoch of relative stability. Under both the T test and the W test, the final value of I_N(STEP, ETAS) = 0.43 can reject the model comparison null hypothesis with great confidence. We are currently conducting further analysis to determine where, when, and for what magnitude earthquakes STEP has an advantage over ETAS.

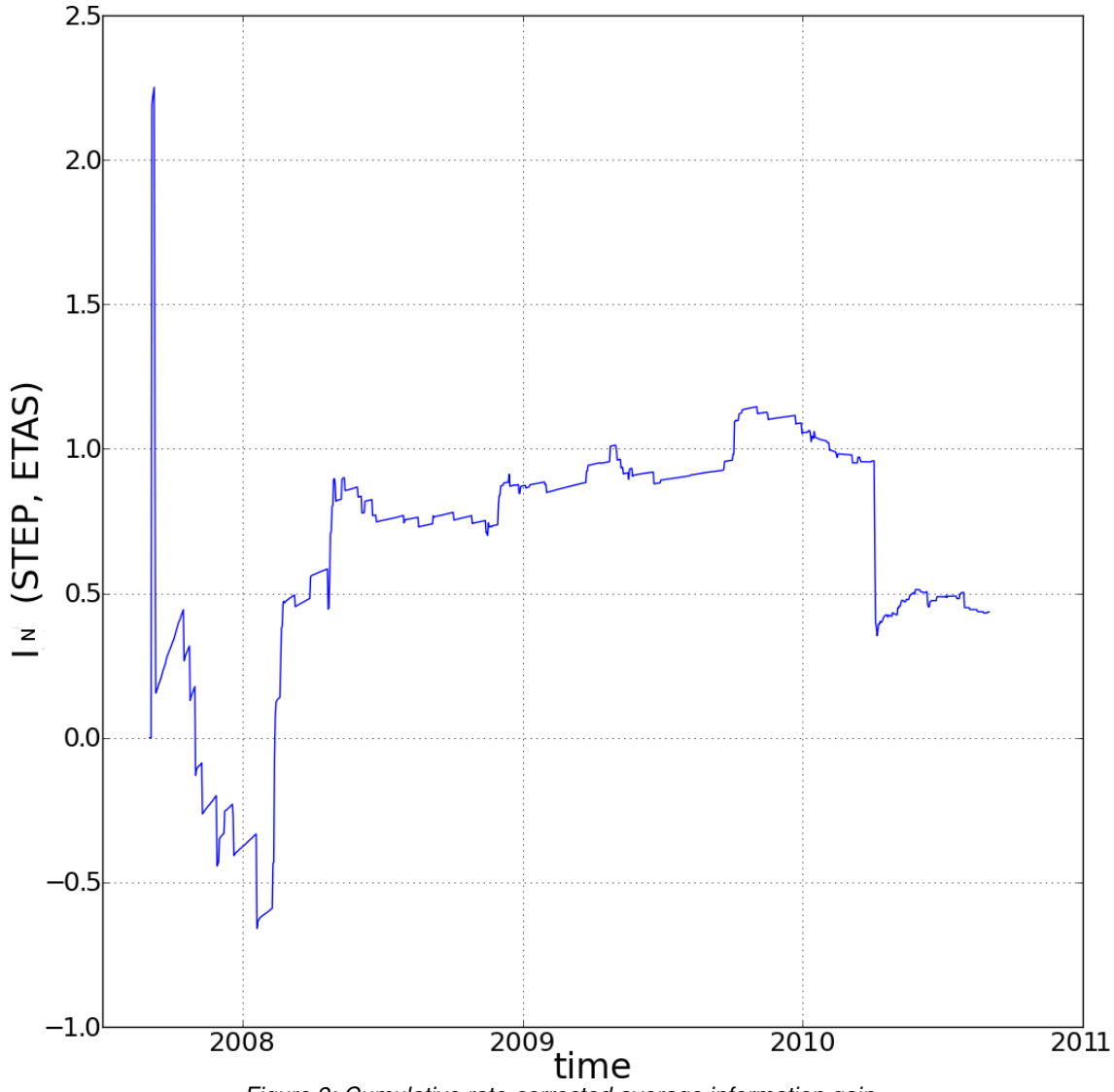


Figure 9: Cumulative rate-corrected average information gain per earthquake

In Table 2, we show some of the individual information gains used to compute the rate-corrected average information gain per earthquake. We do not list all 270 earthquakes, but we highlight the events of the 2008 Baja swarm (yellow), 2 foreshocks (green) to the El Mayor-Cucapah earthquake, the M 7.2 itself (red), and a subset of events that it presumably triggered (cyan).

date	time	lon	lat	mag	STEP forecast	ETAS forecast	probability gain	information gain
02.09.07	17:29	-117.48	33.73	4.73	1.05E-06	1.13E-06	0.93	-0.071724546
04.09.07	14:47	-117.34	32.77	4.06	1.40E-04	2.03E-06	69.04	4.2346984483
…	…	…	…	…	…	…	…	…
09.02.08	07:12	-115.28	32.36	5.10	6.78E-06	9.07E-07	7.47	2.0110110192
11.02.08	18:29	-115.26	32.33	5.10	1.14E-03	1.57E-05	72.72	4.2866594511
11.02.08	23:06	-115.31	32.44	3.95	9.31E-04	1.84E-04	5.07	1.6240736213
12.02.08	04:32	-115.32	32.45	4.97	2.49E-04	7.45E-05	3.35	1.2077487898
12.02.08	04:40	-115.29	32.44	4.04	9.12E-04	7.50E-04	1.22	0.1958118924
12.02.08	09:20	-115.30	32.44	4.25	4.86E-04	3.78E-04	1.29	0.252883976
12.02.08	09:27	-115.31	32.46	4.45	7.11E-04	2.34E-04	3.04	1.1126289041
13.02.08	07:14	-115.29	32.42	3.99	2.67E-03	8.16E-04	3.28	1.1874859327
19.02.08	22:41	-115.31	32.43	5.01	3.84E-04	1.65E-05	23.29	3.1480555074
20.02.08	01:28	-115.31	32.43	4.83	3.91E-03	2.03E-04	19.24	2.9569847568
22.02.08	19:31	-115.29	32.42	4.78	9.46E-05	5.39E-05	1.76	0.5627661031
22.02.08	19:33	-115.27	32.39	4.58	8.06E-04	8.39E-05	9.60	2.2622109008
…	…	…	…	…	…	…	…	…
29.03.08	22:01	-115.23	32.38	4.18	4.10E-04	2.67E-05	15.37	2.7322357401
21.04.08	14:16	-115.49	32.13	4.10	4.75E-06	2.18E-05	0.22	-1.5232283836
…	…	…	…	…	…	…	…	…
17.03.10	17:01	-115.31	32.32	3.96	3.98E-05	3.77E-05	1.06	0.0554972145
31.03.10	09:20	-115.23	32.35	4.16	8.66E-05	2.19E-05	3.95	1.3744955801
04.04.10	22:40	-115.29	32.26	7.20	5.80E-08	3.61E-07	0.16	-1.829321232
04.04.10	22:50	-115.05	32.10	5.51	1.95E-07	1.31E-05	0.01	-4.203575381
04.04.10	22:54	-115.50	32.44	4.03	3.19E-06	3.14E-04	0.01	-4.5893344668
04.04.10	23:09	-115.33	32.20	4.79	8.00E-05	7.90E-05	1.01	0.0129749005
04.04.10	23:15	-115.26	32.30	5.43	5.13E-05	1.81E-05	2.83	1.0387253504
04.04.10	23:19	-115.75	32.59	3.98	4.50E-06	1.34E-04	0.03	-3.3935525958
04.04.10	23:19	-115.74	32.62	3.97	4.32E-06	8.93E-05	0.05	-3.0285560688
…	…	…	…	…	…	…	…	…
24.08.10	21:43	-115.41	32.36	4.22	7.32E-05	7.00E-05	1.05	0.0452032114
29.08.10	15:53	-115.29	32.18	4.28	2.21E-04	6.38E-05	3.46	1.2411211502

Table 2 Selected observed earthquakes and corresponding probability/information gains. Rows in yellow indicate swarm events; the row in red indicates the El Mayor-Cucapah earthquake; rows in green are foreshocks (identified a posteriori); rows in cyan are likely aftershocks.

3.5. Possible explanation

The consistency and comparison methods do not explicitly address why a model does well or poorly, or why one model outperforms another. Based upon visual comparison of the

STEP and ETAS forecasts, we hypothesized that the method used for spatial smoothing was causing the difference between the models. To investigate this, we designed a method to reduce the two-dimensional forecasts to a scalar describing the smoothness of the forecast. To compute this smoothing index, we first compute the two-dimensional Fourier transform of the gridded forecast. The smoothing index is defined as the difference between the maximum and the average. In Figure 10, we show the smoothing index for the STEP and ETAS forecasts throughout the experiment. This plot confirms what is hinted at in Figure 6: ETAS is much smoother than ETAS.

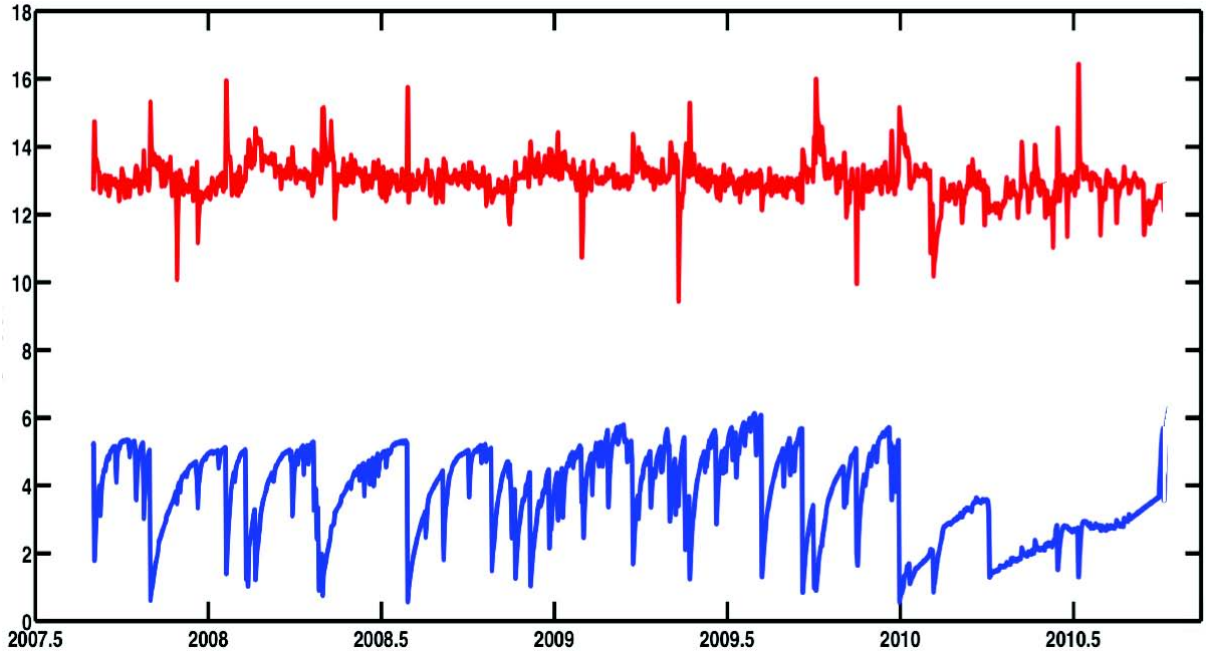


Figure 10: Smoothness index for ETAS (red) and STEP (blue) forecasts throughout the experiment

In fact, spatial smoothing differences can explain the overall results from the consistency tests and the comparison test. ETAS uses a very broad smoothing kernel and STEP uses a very narrow kernel. In the comparison test, STEP benefits because it essentially has a "higher resolution" than ETAS; ETAS is smoother than the observed seismicity. At the same time, in the consistency tests STEP is penalized for not smoothing enough. In this sense, STEP is overly precise—when an earthquake happens that is somewhat distant from the recent events, STEP is judged to be inconsistent.

3.6. An alternative approach

The evaluation methods considered above emphasize hypothesis testing (in the consistency tests) and model selection (in the comparison tests). In either case, one potential result is to "reject" a forecast model if a *p*-value is smaller than some arbitrarily chosen critical value. An alternative approach that we (Marzocchi, Zechar, and Jordan) are developing as part of the EU-sponsored REAKT (strategies and tools for Real Time EArthquake RisK ReducTion) project is to use Bayes' Theorem to calculate posterior probabilities for each model in a set of candidate models. This is also based on likelihood but

can additionally incorporate prior knowledge, including information about model correlation and performance in previous experiments. We write Bayes' Theorem:

$$P(\Lambda^j|\Omega) = \frac{P(\Lambda^j)\exp(L(\Omega|\Lambda^j))}{P(\Omega)}$$

where $P(\Lambda^j)$ is the prior probability of the j^{th} model, $P(\Omega)$ is a normalizing constant, and the term on the left hand side is the posterior probability, the probability that the j^{th} model is the best from the set of candidate models. Using this type of analysis, no model is rejected; rather its posterior probability will approach zero if it provides a very poor fit to the observed data. We also consider how the posterior probabilities evolve throughout an experiment, updating after each new target earthquake and allowing events to delineate testing phases: the prior for the current phase is the posterior from the previous phase. Rather than assuming that all models are equally likely in the beginning of the experiment (i.e., using a uniform prior), we explicitly account for model correlation in the initial priors. Models that are highly correlated will have reduced priors (except when we have only two models). One can then use the posterior probabilities to build a weighted-average ensemble forecast; if the weights are themselves just the posterior probabilities, this is called Bayesian Model Averaging (BMA).

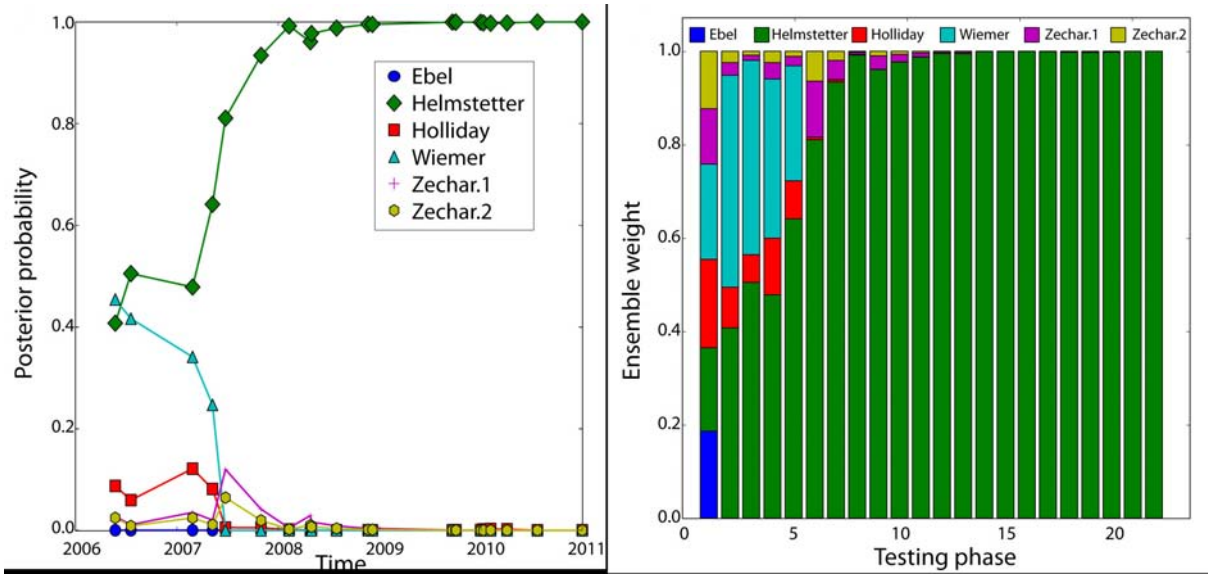


Figure 11: Posterior probabilities (left) and BMA weights (right) for each of the candidate models throughout the experiment

In Figure 11, we illustrate this method using a subset of the RELM forecasts. The panel on the left shows how posterior probabilities evolve throughout the experiment, and the panel on the right shows how the BMA ensemble forecast is constructed.

A similar approach might yield increased predictive skill for short-term earthquake forecasting. Considering the example of STEP and ETAS, blending these models might alleviate the inconsistencies of STEP and allow a higher predictive performance. Earthquake forecast model averaging, or more generally, model combination, is an ongoing research topic that can drive developments of JRA4.

4. ASSOCIATED DEVELOPMENTS

There are several ongoing developments that are related to the objectives of JRA4. ShakeMaps are maps of estimated shaking intensity and ground motion that are constructed by the USGS immediately after large earthquakes; these maps an be used to estimate regions of high damage and thereby allocate emergency resources. In Figure 12, I show an example ShakeMap for a recent earthquake east of Tokyo. In this example, the shaking was rather light and therefore no rescues were required.

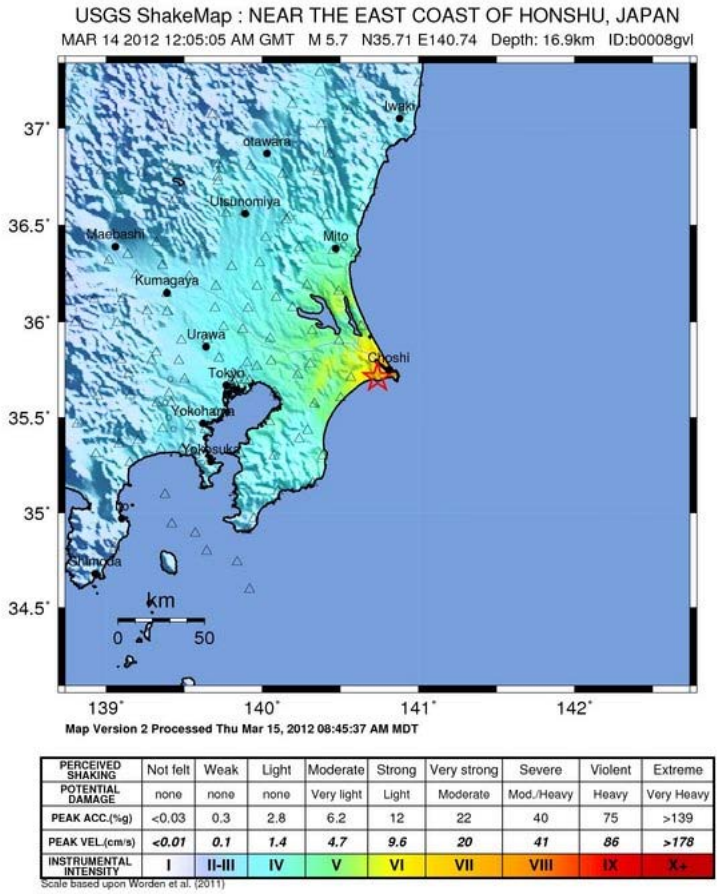


Figure 12: Example USGS ShakeMap for M5.7 event east of Tokyo

ShakeMaps are produced after an earthquake based on instrument measurements and are produced in "nearly real-time" and could potentially be used in a real-time risk assessment system. In particular, ShakeMap information could be used to drive the updating of building fragility functions (i.e., NERA Task 14.2).

The USGS also operates the PAGER (Prompt Assessment of Global Earthquakes for Response) system, which estimates the number of deaths and economic losses in the wake of large earthquakes. This is an alternative to the WAPMERR (World Agency of Planetary Monitoring and Earthquake Risk Reduction) estimates of the same quantities; WAPMERR also estimates injuries. Again, both of these systems are conditional on large earthquakes happening and only make assessments of losses after the earthquake. This is different from the work to be done in NERA Task 14.1, but the PAGER and WAPMERR estimates might

be comparable with the output of that task. In Figure 14 I show example PAGER analysis of the disastrous Haiti M7 earthquake of 2010.

USGS science for a changing world

Earthquake Shaking — Red Alert

USAID FROM THE AMERICAN PEOPLE

GSN ANSS

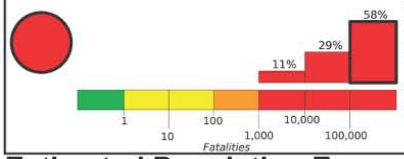
M 7.0, HAITI REGION

Origin Time: Tue 2010-01-12 21:53:10 UTC (16:53:10 local)
Location: 18.45°N 72.57°W Depth: 13 km

PAGER Version 1

Created: 36 weeks, 2 days after earthquake

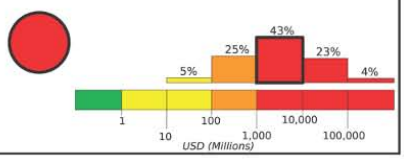
Estimated Fatalities



Red alert for shaking-related fatalities and economic losses. High casualties and extensive damage are probable and the disaster is likely widespread. Past red alerts have required a national or international response.

Estimated economic losses are 20-100% GDP of Haiti.

Estimated Economic Losses



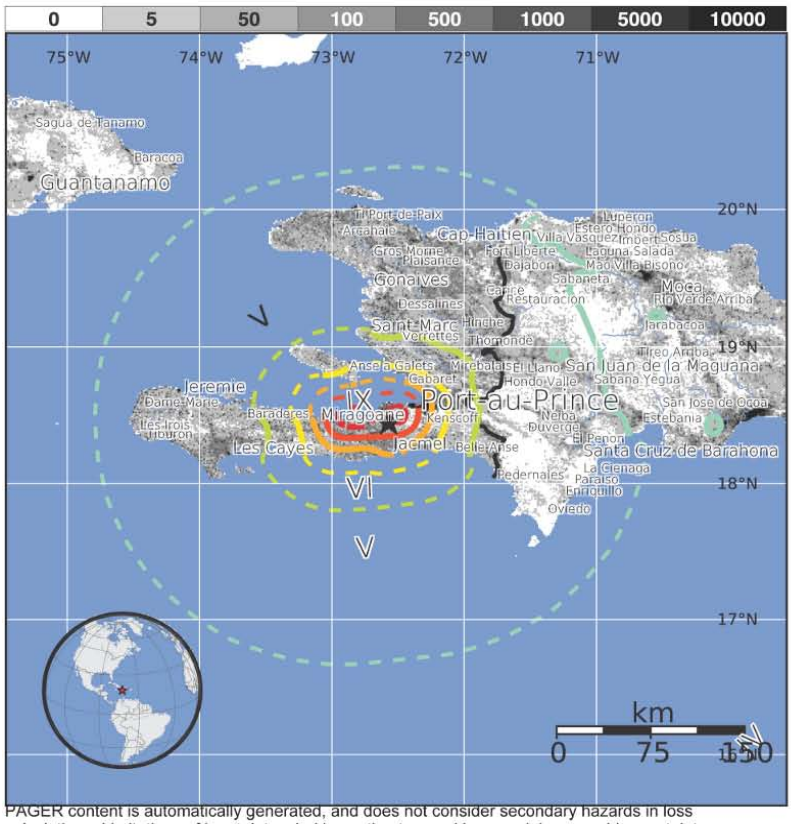
Estimated Population Exposed to Earthquake Shaking

ESTIMATED POPULATION EXPOSURE (k = x1000)		- -*	50k*	7,468k*	6,361k	926k	598k	2,030k	908k	118k
ESTIMATED MODIFIED MERCALLI INTENSITY		I	II-III	IV	V	VI	VII	VIII	IX	X+
PERCEIVED SHAKING		Not felt	Weak	Light	Moderate	Strong	Very Strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	Resistant Structures	none	none	none	V. Light	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy
	Vulnerable Structures	none	none	none	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy	V. Heavy

*Estimated exposure only includes population within the map area.

Population Exposure

population per ~1 sq. km from Landscan



Structures:
Overall, the population in this region resides in structures that are highly vulnerable to earthquake shaking, though some resistant structures exist. The predominant vulnerable building types are concrete/cinder block masonry and mud wall construction.

Historical Earthquakes (with MMI levels):

Date	Dist. (km)	Mag.	Max MMI(#)	Shaking Deaths
2003-09-22	246	6.4	IX(132k)	1
1984-06-24	355	5.2	V(440k)	5
1984-06-24	342	6.7	VII(326k)	5

Recent earthquakes in this area have caused secondary hazards such as landslides that might have contributed to losses.

Selected City Exposure

from GeoNames.org

MMI	City	Population
X	Petit Goave	118k
X	Gressier	26k
X	Grand Goave	49k
IX	Leogane	134k
VIII	Carrefour	442k
VIII	**Miragoane**	**89k**
VIII	**Port-au-Prince**	**1,235k**
VIII	Delmas 73	383k
V	**Verrettes**	**49k**
IV	Santo Domingo	2,202k

bold cities appear on map (k = x1000)

PAGER content is automatically generated, and does not consider secondary hazards in loss calculations. Limitations of input data, shaking estimates, and loss models may add uncertainty.
http://earthquake.usgs.gov/pager

Event ID: us2010rja6

Figure 13 Example PAGER output, shown here for the 2010 Haiti M7 earthquake.

5. USER NEEDS

Identifying user needs is a goal associated with NERA Task 14.4, but it is worth making a few relevant comments in this report.

User needs will undoubtedly vary from nation to nation. A striking example is the difference between Italy's reaction to the 6 April 2009 M6 L'Aquila earthquake and Japan's reaction to the 11 March 2011 M9 Tohoku earthquake. As mentioned earlier in this report, the Italian civil protection agency has expressed its intent to develop operational earthquake forecasting infrastructure and an interest in working closely with researchers on short-term earthquake forecasting. On the other hand, the majority of efforts in Japan are aimed at continued improvement of physics-based long-term seismic hazard modeling, with little work targeting risk estimation. It is therefore difficult to make general statements about user needs in this context.

Nevertheless, in the case of the L'Aquila earthquake, the international commission made some recommendations to Italian civil protection, and given that our work package emphasizes crisis situations for which L'Aquila is the primary example, these recommendations can guide JRA4 work. Specifically, we will employ rigorous quantitative methods with an aim to generate probabilistic risk estimates for decision-makers, and we aim to make these time-varying estimates as quickly as possible given inevitable delays in data. The precise form of this output will be guided by the findings of Task 14.4; on this point our work package participants are quite flexible. Given the recent developments described in this report, we are poised to make genuine progress on the problem of real-time earthquake risk assessment.

6. REFERENCES

Falcone, G.; Console, R. & Murru, M. (2010). *Short-term and long-term earthquake occurrence models for Italy: ETES, ERS and LTST*, Annals of Geophysics 53 : 41-50.

Gerstenberger, M. C.; Wiemer, S.; Jones, L. M. & Reasenberg, P. A. (2005). *Real-time forecasts of tomorrow's earthquakes in California*, Nature 435 : 328-331.

Jordan, T. H.; Chen, Y.-T.; Gasparini, P.; Madariaga, R.; Main, I.; Marzocchi, W.; Papadopoulos, G.; Sobolev, G.; Yamaoka, K. & Zschau, J. (2011). *Operational Earthquake Forecasting: State of Knowledge and Guidelines for Utilization*, Annals of Geophysics 54 : 315-391.

Lolli, B. & Gasperini, P. (2003). *Aftershocks hazard in Italy Part I: Estimation of time-magnitude distribution model parameters and computation of probabilities of occurrence*, Journal of Seismology 7 : 235-257.

Marzocchi, W. & Lombardi, A. M. (2009). *Real-time forecasting following a damaging earthquake*, Geophysical Research Letters 36.

Michelini, A.; Faenza, L.; Lauciani, V. & Malagnini, L. (2008). *ShakeMap implementation in Italy*, Seismological Research Letters 79 : 688-697.

Ogata, Y. (1988). *Statistical models for earthquake occurences and residual analysis for point processes*, Journal of the American Statistical Association 83 : 9-27.

Reasenberg, P. A. & Jones, L. M. (1989). *Earthquake hazard after a mainshock in California*, Science 243 : 1173-1176.

Rhoades, D.; Schorlemmer, D.; Gerstenberger, M.; Christophersen, A.; Zechar, J. & Imoto, M. (2011). *Efficient testing of earthquake forecasting models*, Acta Geophysica 59 : 728-747.

Schorlemmer, D. & Gerstenberger, M. C. (2007). *RELM testing center*, Seismological Research Letters 78 : 30-36.

Schorlemmer, D.; Zechar, J.; Werner, M.; Field, E.; Jackson, D. & Jordan, T. (2010). *First Results of the Regional Earthquake Likelihood Models Experiment*, Pure and Applied Geophysics 167 : 859-876.

Trendafiloski, G.; Wyss, M. & Rosset, P. (2011). *Loss estimation module in the second generation software QLARM*, in Human Casualties in Natural Disasters: Progress in Modeling and Mitigation, ed. R. Spence, E.So, C. Scawthornp, Springer: Dordrecht, Heidelberg, London, New York, 95-104.

van Stiphout, T.; Wiemer, S. & Marzocchi, W. (2010). *Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake*, Geophysical Research Letters 37.

Werner, M.; Zechar, J.; Marzocchi, W. & Wiemer, S. (2010). *Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts*, Annals of Geophysics 53 : 11-30.

Zechar, J.; Gerstenberger, M. & Rhoades, D. (2010). *Likelihood-Based Tests for Evaluating Space-Rate-Magnitude Earthquake Forecasts*, Bulletin of The Seismological Society of America 100 : 1184-1195.

Zechar, J. & Jordan, T. (2010). *Simple smoothed seismicity earthquake forecasts for Italy*, Annals of Geophysics 53 : 99-105.

Zhuang, J. (2011). *Next-day earthquake forecasts for the Japan region generated by the ETAS model*, Earth Planets and Space 63 : 207-216.