



**HAL**  
open science

## Comparative evaluation of physics-based and statistical forecasts in Northern California,

M. Segou, T. Parsons, W. Ellsworth

► **To cite this version:**

M. Segou, T. Parsons, W. Ellsworth. Comparative evaluation of physics-based and statistical forecasts in Northern California,. *Journal of Geophysical Research: Solid Earth*, 2013, pp.6219-6240. 10.1002/2013JB010313 . hal-01183185

**HAL Id: hal-01183185**

**<https://hal.science/hal-01183185v1>**

Submitted on 1 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

## Comparative evaluation of physics-based and statistical forecasts in Northern California

M. Segou,<sup>1,2</sup> T. Parsons,<sup>1</sup> and W. Ellsworth<sup>1</sup>

Received 23 April 2013; revised 15 November 2013; accepted 21 November 2013; published 18 December 2013.

[1] We perform a retrospective forecast test using Northern California seismicity for the period between 1980 and 2009. We compare 7 realizations of the short-term clustering epidemic-type aftershock sequence (ETAS) model, and 21 models combining Coulomb stress change calculations and Rate/State theory (CRS) to forecast seismicity rates in 10 day time intervals. We employ a common learning phase (1974–1980) for CRS models to ensure consistency, and we evaluate the forecasts with log likelihood statistics to detect any spatial inconsistencies and compare the total numbers of forecasts versus observed events. We find that: (1) ETAS models are better forecasters of the spatial evolution in seismicity in the near-source region, (2) CRS models can compete with ETAS models away from the mainshock rupture, and for short periods after mainshocks, (3) CRS models with optimally oriented receiver fault planes perform better in the first few days after mainshocks, whereas mapped fault planes should be implemented for longer-term forecasting, and (4) CRS models based on shear stress change calculations have comparable performance with Coulomb stress change models, with the benefit of lesser parameters involved in stress calculations. We conclude that physics-based and statistical forecast models are complimentary to each other and that future forecasts should be based on statistical models for near-source regions, and physical models for longer periods and distances. However, the realization of the CRS models involves a number of critical parameters (reference seismicity rates, regional stress field, and loading rates), which should be retrospectively tested to improve the predictive power of physics-based models.

**Citation:** Segou, M., T. Parsons, and W. Ellsworth (2013), Comparative evaluation of physics-based and statistical forecasts in Northern California, *J. Geophys. Res. Solid Earth*, 118, 6219–6240, doi:10.1002/2013JB010313.

### 1. Introduction

[2] During the last decade, a wide range of statistical and physics-based models was developed to forecast earthquake occurrence following moderate and great earthquakes. Empirical statistics and cascading effects are the basis for epidemic-type aftershock sequences (ETASs) [Ogata, 1988], whereas rate/state friction laws [Dieterich, 1994] combined with the idea of static stress triggering [Stein *et al.*, 1997; Harris and Simpson, 1996; Toda *et al.*, 2005] are the cornerstone of physics-based-models. The consideration of spatiotemporal clustering effects through empirical-statistical models is important for short-term earthquake forecasting [Field *et al.*, 2009]. Recently, Steacy *et al.* [2013] supports that including spatial constraints from Coulomb stress changes can increase forecasting power of the statistical models. However, we ultimately want to understand the physics underlying

aftershock occurrence, that the empirical methods only describe, and the theoretical bases of stress transfer and rate/state friction offer explanations. Thus, their direct comparison is important not only for operational forecasting [Jordan *et al.*, 2011] but also for understanding the physics of earthquake triggering.

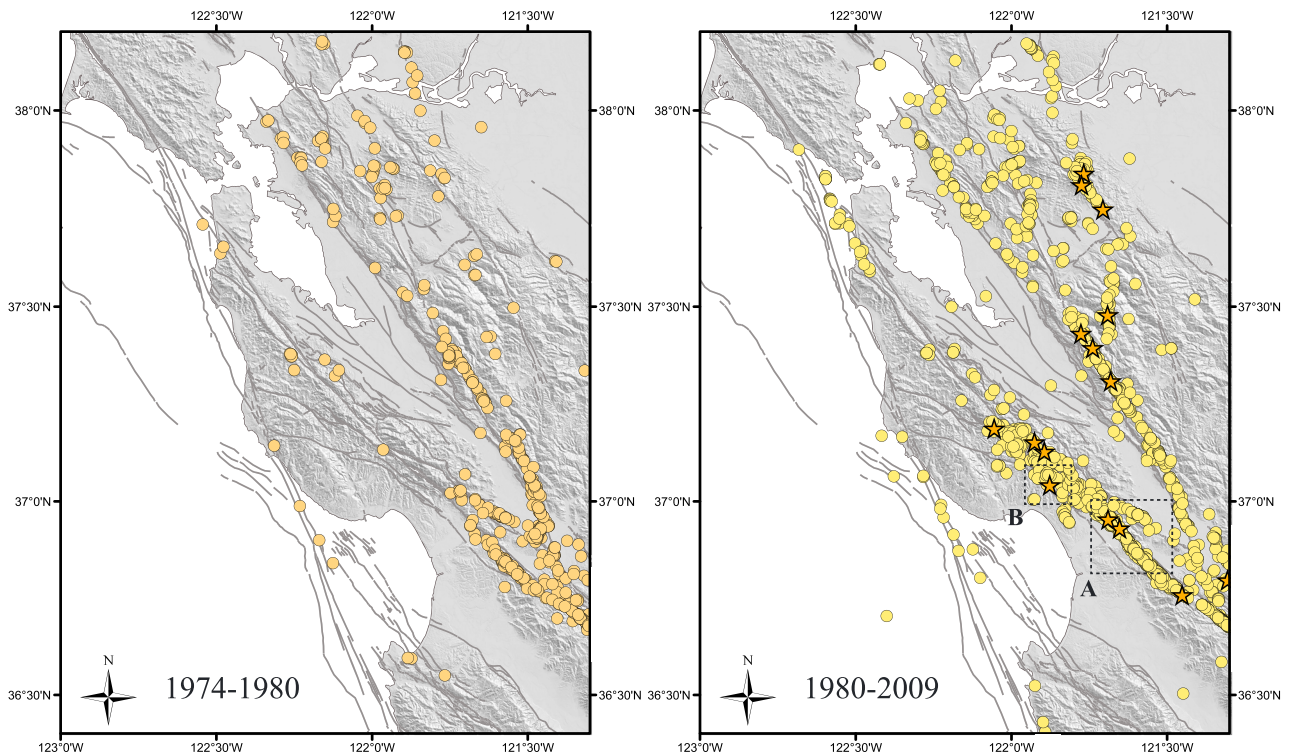
[3] We develop a wide range of physics-based (CRS) and short-term clustering statistical models (ETAS) to retrospectively forecast Northern California seismicity between 1980 and 2009 with  $M_L \geq 3.0$ . Other statistical forecasting methods have been proposed lately, such as the EEPAS (every earthquake a precursor), PPE (Proximity to Past Earthquakes), and double branching process models by Rhoades [2007] and Marzocchi and Lombardi [2008], respectively. Although these individual models address intermediate to long-term clustering, recent research [Rhoades, 2013] reveals that a mixture of ETAS, EEPAS, and PPE models may also improve short-term forecasting. We implement twenty-one CRS models considering various sources of uncertainty such as different mainshock slip models, receiver plane geometry, stressing rates, receiver fault friction coefficient, and reference seismicity rate models. We also compare seven ETAS realizations; those are based on two basic principles of ETAS modeling that (1) the update time should be reasonably short (24 h) and (2) all available earthquakes should be used for cascading earthquake sequences, which is found to improve statistical forecasting methods [Helmstetter, 2003]. It is clear that the rules for forecast model

<sup>1</sup>United States Geological Survey, Menlo Park, California, USA.

<sup>2</sup>Now at Observatoire de la Cote d'Azur, Valbonne, France.

Corresponding author: M. Segou, Observatoire de la Cote d'Azur, 250 Rue Einstein, Sophia-Antipolis, FR-06560 Valbonne, France. (segou@geoazur.unice.fr)

©2013. American Geophysical Union. All Rights Reserved. 2169-9313/13/10.1002/2013JB010313



**Figure 1.** Seismicity with magnitude  $M_L \geq 3.0$  within our study area ( $-123.0^\circ\text{W}$ ,  $36.4^\circ\text{N}$  to  $-121.3^\circ\text{W}$ ,  $38.2^\circ\text{N}$ ). (left) The 177 earthquakes from the available ANSS catalog inside the learning phase of our forecast experiment between 1974 and 1980 and (right) 15 major events (stars), for which we have calculated the stress changes imparted (listed in Table 1) together with 1459 earthquakes (1980–2009), taken from the high-precision relocated catalog of *Waldhauser and Schaff* [2008] for the time period 1984–2009 and the available ANSS catalog between 1980 and 1984. Also noted subregions A and B that have been used for performance evaluation of the forecast models.

implementation are not the same for physics-based and ETAS models, but in this study we focus on the comparison between the theoretically advantageous ones from each family, e.g., the forecast time window is set to 10 days for CRS models, but we adopt a daily updated ETAS model, since *Nanjo et al.* [2012] has demonstrated the link between poor performance and extended update intervals for statistical models.

[4] We evaluate the relative performance of the forecast models with log likelihood statistics [*Schorlemmer et al.*, 2007; *Zechar et al.*, 2010; *Rhoades et al.*, 2011]; our goal is not necessarily to determine the best performing individual model, but instead through comparative performance, to identify which components of the physical and statistical models succeed in different spatial, temporal, and geological parameterizations.

[5] We implement the suggested log likelihood statistics to different subregions along San Andreas Fault for one year following the 1989 Loma Prieta ( $M_w = 6.9$ ) rupture. During our evaluation period the spatiotemporal evolution of seismicity within these subregions is influenced by at least four mainshocks with  $M_L \geq 5.0$ , included in our modeling. We identify a time and space-dependent character of the predictive power of the forecast models developed in this study. Furthermore, we support that forecast models based on rate/state friction and static stress changes constitute a feasible solution for prospective experiments within the frame of operational earthquake forecasting.

## 2. Data

[6] We use the earthquake catalog and the focal mechanisms of the Advanced National Seismic System (ANSS) for shallow events (depth,  $z < 30$  km) inside the learning phase of 1974–1980 (Figure 1a), and the latest version of the high-precision relocated catalog of *Waldhauser and Schaff* [2008] during our testing phase between 1980 and 2009 (Figure 1b).

[7] Our models are implemented in a  $2.5 \text{ km} \times 2.5 \text{ km}$  grid that encompasses the San Francisco Bay region of Northern California between  $123^\circ\text{W}$ – $121.3^\circ\text{W}$  longitude and  $36.4^\circ\text{N}$ – $38.2^\circ\text{N}$  latitude, for forecasting earthquakes with  $M_L \geq 3.0$ , the latter corresponding to the threshold of completeness for the modern instrumental California catalog [*Felzer*, 2008; *Hardebeck*, 2013]. This grid is used for calculating expected aftershock rates in ETAS models and the coseismic static stress changes from 15  $M_L \geq 5.0$  earthquakes that occurred during our testing period (1980–2009), listed in Table 1. The consideration of the most recent earthquakes with  $M_L \geq 5.0$  (e.g., the 2007 Alum Rock event) aims in providing updated estimates of forecast parameters (e.g., the state variable  $\gamma$  in the rate-and-state equations, see section A1), which can be further employed for prospective forecasting in case a major earthquake occurs within our study area.

[8] Mainshock sources are represented by boxcar uniform and variable slip models, the latter available from the finite-fault

**Table 1.** Seismic Parameters of 15  $M_L \geq 5.0$  Mainshocks in Our Study Area Between 1980 and 2009 Taken From the *Waldhauser and Schaff* [2008] High-Accuracy Relocated Catalog for Northern California<sup>a</sup>

No	Year	Month	Day	Hr	Min	$M_L$	Depth (km)	Earthquake Name
1	1980	1	24	19	0	5.8	14.79	Livermore I
2	1980	1	24	19	1	5.1	6.9	Livermore II
3	1980	1	27	2	33	5.4	14.43	Livermore III
4	1984	4	24	21	15	6.2	7.968	Morgan Hill ( <i>Beroza and Spudich</i> [1988])
5	1986	3	31	11	55	5.7	8.394	Mt. Lewis
6	1988	2	20	8	39	5.1	8.222	Alum Rock 1988
7	1988	6	13	1	45	5.3	8.869	Hollister
8	1988	6	27	18	43	5.3	11.544	Elsman Lake I
9	1989	8	8	8	13	5.4	12.586	Elsman Lake II
10	1989	10	18	0	4	7.0	16.412	Loma Prieta mainshock ( <i>Beroza</i> [1991])
11	1989	10	18	0	41	5.1	13.864	Loma Prieta aftershock-St Cruz Mts
12	1990	4	18	13	53	5.4	4.611	Watsonville I
13	1990	4	18	15	46	5.1	6.338	Watsonville II
14	1998	8	12	14	10	5.1	7.746	SJ Bautista
15	2007	10	31	3	4	5.4	7.486	Alum Rock 2007 ( <i>Murray-Moraleda and Simpson</i> [2009])

<sup>a</sup>We cite in parentheses in the last column the main slip distributions used to describe the three larger mainshocks in our study.

source model database. We have considered alternative slip distributions for the 1984  $M_w=6.4$  Morgan Hill, 1989  $M_w=6.9$  Loma Prieta, and the 2007  $M_w=5.4$  Alum Rock earthquakes and we employed as mainshock sources the models developed by *Beroza and Spudich* [1988], *Beroza* [1991], and *Murray-Moraleda and Simpson* [2009]. For deriving uniform-slip source models, we estimated the amount of slip using the moment relation of *Hanks and Kanamori* [1979], assuming that the moment magnitude equals the local magnitude, while the source dimensions are based on the scaling relation of *Wells and Coppersmith* [1994] for the specific faulting style.

### 3. Description of Forecast Models

[9] The predictive power of a CRS model is based on the idea that stress is transferred through the elastic crust following the distortion caused by slip during an earthquake. Representing this slip as a series of dislocations allows crustal volumes, subject to calculated stress increases and decreases to be identified, with the stress-increased volumes expected to be at sites where most aftershocks will occur [e.g., *Stein*, 1999]. The Coulomb failure criterion was implemented as follows:

$$\Delta\tau = |\Delta\bar{\tau}_f| + \mu(\Delta\sigma_n + \Delta p) \quad (1)$$

where  $|\Delta\bar{\tau}_f|$  is the shear stress change parallel to the fault's rake,  $\mu$  the friction coefficient,  $\Delta\sigma_n$  the stress change normal to the fault plane, and  $\Delta p$  the pore pressure change. Our final implementation uses the effective coefficient of friction, which also models Skempton coefficient  $B_k$ , to incorporate pore fluid effects so that friction becomes  $\mu = \mu(1 + B_k)$  and the Coulomb criterion, following *Rice's* [1992] formulation, is given by

$$\Delta\tau = |\Delta\bar{\tau}_f| + \mu(\Delta\sigma_n) \quad (2)$$

[10] Sources of calculation uncertainty include the friction parameter and its relationship with pore fluid pressure and the adequate representations of coseismic slip distributions. Additional sources of uncertainty with regard to the simplified

static stress hypothesis could be dynamic, secondary, and postseismic relaxation effects, including deep fault afterslip.

[11] Short-term clustering models (ETAS) are based on the Omori and the Gutenberg-Richter law and the idea that large magnitude earthquakes trigger even more smaller magnitude aftershocks [*Ogata*, 1988]. Although, operational forecasting is mostly based in short-term statistical clustering models [*Jordan et al.*, 2011], the knowledge of the physical system driving earthquake occurrence can be approached only through extensive testing of physics-based models, allowing us to understand the significance behind a number of critical parameters and features such as the representation of the stress field, the completeness of earthquake catalogs and network detectability, loading rates on active faults, and aftershock decay. Following *Jordan and Jones'* [2010] definition, "the goal of operational earthquake forecasting is to provide the public with authoritative information on the time dependence of regional seismic hazards" which from the scientific perspective is based on the research and development of earthquake clustering models.

#### 3.1. Rate-and-State Friction Forecast Models

[12] In this study, physics-based forecasts are represented by combining Coulomb stress changes and rate/state equations for expected seismicity rates following a stress perturbation [*Dieterich*, 1994]. The cornerstone of physics-based forecasting is the assumption that static Coulomb stress changes (DCFF) caused by a dislocation in an elastic medium are responsible for triggering aftershocks. We employ a common learning phase (1974–1980) for the CRS models to ensure consistency. In order to model the evolution of seismicity we calculate the DCFF, at each node of the 2.5 km  $\times$  2.5 km grid, imparted by the 15  $M_L \geq 5.0$  events between 1980 and 2009 and we update the forecast after each major event listed in Table 1. We note that we compare between physics-based and statistical models for the 300 day period following the  $M_w=6.9$  Loma Prieta 1989 earthquake. We develop three primary classes of physics-based models depending on how the stress-change receiver planes are handled: CRS-1, CRS-2, and CRS-3. We adopt a friction coefficient  $\mu=0.2$  for our CRS-1 and CRS-2 models, assuming that major active faults in Northern California are



weak based on seismological, borehole, and heat flow data [Lachenbruch and Sass, 1980; Townend, 2006; Mount and Suppe, 1992; Parsons et al., 1999; Geist and Andrews, 2000]. Our assumption is further supported by Reasenber and Simpson [1992] who studied the response of regional seismicity to the static stress changes following the Loma Prieta earthquake and found that, “the agreement is best for models with low values of the coefficient of friction ( $0.1 \leq \mu \leq 0.3$ ) on Bay Area faults.” The CRS-3 model tests the predictive power of forecast models based on shear stress changes ( $\mu = 0$ ). A summary of the main features and parameters for each CRS model is shown in Table A1.

[13] From equation (A1) it is obvious that CRS models depend on three critical parameters: (1) the reference seismicity rate (or background seismicity rate), (2) the constant secular shear stressing rate or equivalent, such as the fault loading rate, and (3) the term  $A\sigma$  (assumed to be 0.05 MPa in this study) [Toda et al., 2005]. We note that there is a trade off between the two parameters,  $A\sigma$  and  $\dot{\tau}$ , and our implementation here chooses to fix  $A\sigma$  and vary the stressing rate  $\dot{\tau}$  based on previous researches for major faults in Northern California. We estimate the expected seismicity in each node at ( $dt=$ )10 day time intervals inside our testing phase (1980–2009), except in the case where two or more  $M_L \geq 5.0$  events occur less than 10 days apart, in which case we allow time intervals to vary accordingly.

[14] We have implemented an extensive set of background seismicity models and fault loading rates to analyze the sensitivity of our forecasting results to the above CRS components. Analytically, our background rate formulations include a declustered model following Reasenber [1985], a gridded, and a smoothed-gridded model using a Gaussian operator with correlation distance 5 km. We have considered: (1) uniform tectonic stressing rates over our study area equal to 0.065 bar/yr, inversely proportional to the aftershock sequence duration of the Loma Prieta earthquake and (2) fault-specific loading rates by Smith and Sandwell [2003], Smith-Konter and Sandwell [2009], and Parsons [2002]. For the last two loading rate representations, we have extended the fault-specific loading rates to 10 km distance from the surface traces of the faults.

[15] As previously noted, the difference between models CRS-1 and CRS-2 (bold denotes the generation) is the receiver fault plane approach. One approach for mapping stress change patterns is to calculate values for optimally oriented faults [King et al., 1994], assuming these are the most likely to host aftershocks. This approach tends to work well in explaining observations [e.g., Hainzl et al., 2009], but it also involves a number of free parameters because it requires estimation of the magnitudes and directions of the regional stress field, and it potentially resolves stresses on fault planes that at one hand may not necessarily exist, and on the other hand if they exist, we have identified them as “optimal” based on the assumption that we have an adequate representation of the remain shock stress field. A second approach is to define receiver fault plane orientations and mechanics based on geological knowledge, a process that is also subject to uncertainty. It is unclear at this time which method is better [e.g., Steacy et al., 2005; Cocco et al., 2010], so we make calculations on optimally oriented fault planes (CRS-1) and on a dense grid of 3-D fault planes representing the predominant fault structures in Northern California (CRS-2) (Figure A1). We have set the 3-D fault plane grid (Figure A1) to comply with

the following concepts: (1) we only use seismological data (Figure A2) to influence our definitions of faulting style from the same period of the forecast learning period (1974–1980) and (2) we describe primary geological features that are consistent with the long-term deformation history (major faults and recent topography). For the development of the predominant geology grid, we refer the reader to see section A3. We also provide Coulomb stress change calculations along San Andreas Fault, following Loma Prieta earthquake, implemented in CRS-1 and CRS-2 models in Figure A3. For the CRS-1 models, the regional stress field representation was taken from Hardebeck and Michael [2004], with the maximum compressive stress set to N19°E, while the regional stress  $\sigma_1 - \sigma_3 = 10$  MPa [Toda et al., 2005]. This gives the optimally oriented fault plane models (CRS-1) an advantage in the sense that a better-defined stress direction is available than would have been guessed during the learning period (1974–1980).

[16] We note that our CRS implementation in this study is based on stress perturbations from major events; however, the results of a complimentary study of the first two authors suggest that the consideration of small magnitude could improve the predictive power of physics-based models under specific circumstances (M. Segou and T. Parsons, The stress shadow problem in physics-based aftershock forecasting: Does incorporation of secondary stress changes help?, *Geophysical Research Letters*, 2013, in review), depending whether these events can play an important role in elastic stress redistribution [Marsan, 2005; Helmstetter et al., 2005].

### 3.2. Statistical Forecast Models

[17] Epidemic-type aftershock (ETAS) models have been extensively used the last few years in forecasting to predict aftershock occurrence both in space and time after a mainshock (Werner et al. [2011] and Woessner et al. [2011] among others). The models have evolved from the formulations by Ogata [1988, 1998], Ogata and Zhuang [2006], and Zhuang et al. [2002], and express the seismicity rate  $\lambda(x,y,t)$  for  $M > M_{th}$  at a spatial bin  $(x,y)$  and time  $t$  given by

$$\lambda(x,y,t) = \mu(x,y) + \sum_{i:t_i < t} \frac{K e^{\alpha(M_i - M_{th})}}{(t - t_i + c)^p} f_i(x - x_i, y - y_i; M_i) \quad (3)$$

where  $M_{th} \geq 2.5$  corresponds to the minimum triggering magnitude,  $K \left( = A c^{(p-1)^2} \right)$  in  $day^{(p-1)}$  unit is related to the short-term productivity, directly expressed by  $A$  that corresponds to the average number of offspring events by an earthquake at the threshold magnitude, while  $\alpha$  is the efficiency of an event to generate aftershocks relative to its magnitude. The spatial PDF  $f(x,y,M)$  represents the distribution of distance to offspring events relative to the magnitude of ancestors, and is given by the analytical form

$$f(x,y;M) = \frac{q-1}{\pi D(M)} \left( 1 + \frac{x^2 + y^2}{D(M)} \right)^{-q} \quad (4)$$

where  $D(M) = d e^{\gamma(M - M_{min})}$ ,  $d$  and  $q$  are the spatial fitting parameters, while  $\gamma$  is proportional to  $\alpha$  of equation (3).

[18] For the seven ETAS models, we use all available earthquakes with  $M_L \geq 2.5$  to forecast earthquakes with  $M_L \geq 3.0$ . ETAS-1 to ETAS-6 models correspond to daily updated forecasts, whereas the ETAS-7 model is updated

**Table 2.** Estimates of Space-Time ETAS Parameters<sup>a</sup>

	ETAS Number							Parameter
	1	2	3	4	5	6	7	
$K$	0.061	0.211	0.154	0.154	0.154	0.105	0.061	Related to productivity
$\alpha$	1.0	0.7	0.8	0.8	0.8	0.9	0.8	$\alpha$ value
$c$ (days)	0.16	0.16	0.16	0.16	0.16	0.16	0.16	$c$ value
$p$	1.27	1.27	1.27	1.27	1.37	1.27	1.27	Decay parameter
$b$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	$b$ value
$\gamma$	1.0	0.7	0.8	0.5	0.8	0.9	0.8	$\gamma$ value
$d$ (km)	0.79	0.79	0.79	0.79	0.79	0.79	0.79	
$q$	1.96	1.96	1.96	1.96	1.96	1.96	1.96	
$M_{\min}$	2.5	2.5	2.5	2.5	2.5	2.5	2.5	Minimum magnitude for triggering
$dt$ (days)	1	1	1	1	1	1	10	Update interval

<sup>a</sup>In this table  $K$ ,  $c$ , and  $p$  are Omori law values [Utsu, 1961] governing the decay rate of aftershocks,  $\alpha$  and  $\gamma$  estimates the magnitude efficiency of an earthquake in generating its offspring, and  $d$  and  $q$ , are spatial fitting parameters. The parameter  $dt$  corresponds to the time interval for the model update and  $M_{\min}$  is the minimum magnitude for triggering.

every 10 days, which eventually leads to a poor performance depending on the number of earthquakes between model updates [Nanjo *et al.*, 2012]. We distinguish between two groups of ETAS models: ETAS-1 and ETAS-7, and ETAS-2 to ETAS-6 based on the modification of parameters provided in Hardebeck [2013] and Werner *et al.* [2011], respectively. Hardebeck [2013] focuses in deriving and validating ETAS parameters for California within the framework of Unified California Earthquake Rupture Forecast, whereas Werner *et al.* [2011] parameter estimation has been found to optimize daily forecasts, including secondary triggering effects. Here, we set the  $c$  value to 0.1667 days to reflect the network detection limitations for the first 4 h [Enescu *et al.*, 2009] following the Loma Prieta mainshock [Dietz and Ellsworth, 1990; Dietz and Ellsworth, 1997] discussed in the following paragraph. We also consider fixed spatial fitting parameters  $q$ ,  $d$  equal to 1.96 and 0.79 km. For determining the  $\gamma$  parameter, Ogata and Zhuang [2006] support two approaches; either estimate by fitting the data or fix  $\gamma$  value equals  $\alpha$  value ( $\alpha = \gamma$ ). With the exception of the ETAS-4 model, where  $\alpha$  ( $= 0.84$ ) and  $\gamma$  ( $= 0.49$ ) are taken by fitting the complete instrumental catalog for the entire testing period, ETAS models consider  $\alpha = \gamma$ .

[19] According to the productivity law, each mainshock with magnitude  $M$  triggers  $\sim 10^{\alpha M}$  number of events and the relative abundance of  $\sim 10^{-bM}$  events is given by the Gutenberg-Richter law, which leads to a contribution of magnitude  $M$  events  $10^{-(b-\alpha)M}$ . In ETAS-1 model, we consider  $\alpha = b = 1.0$ , therefore the contribution of each magnitude class is equal whereas in ETAS-2 model  $\alpha < b = 1.0$ , which increases the importance behind the contribution of smaller magnitude events. According to Werner *et al.* [2011], this has, as a consequence, that small, undetected earthquakes have a significant, time-dependent impact on the observed seismicity budget and any failure to model their effect causes parameter bias [Sornette and Werner, 2005a, 2005b; Saichev and Sornette, 2005, 2006]. In this study for ETAS modeling we consider  $M_{\min} = 2.5$  and  $m_d = 3.0$ , which corresponds to the minimum magnitude for triggering and detection limit of the network, and the threshold magnitude for our forecast, respectively. This consideration is more critical for the statistical forecast immediately after the Loma Prieta mainshock rather than any other modern event, since there is a larger discrepancy between the minimum magnitude of the modern instrumental catalog and the

detection magnitude since the network experienced power cuts and seismic station malfunctions within the first 24 h [Dietz and Ellsworth, 1990], which is expected to influence the performance of the daily forecast for the first day.

[20] We carefully examine this underreporting in 4 h windows using frequency-magnitude density distributions to estimate completeness; hours 1 to 4 appear complete down to  $M3.0$  with  $b$  value  $0.83 \pm 0.09$  for events  $M \geq 3.0$  (Figure A4a), while for the remaining first day we reach a  $b$  value  $0.81 \pm 0.05$  for events  $M \geq 1.5$  (Figure A4b). We present the variability of magnitude of completeness with time after the mainshock for the first 10 days in Figure A4c. We observe that approximately after the first 4 h, the magnitude of completeness is well below the minimum magnitude for triggering ( $M_{\min} = 2.5$ ). We therefore estimate that approximately three events between  $M2.5$  and  $M3.0$  are not reported within the first 4 h in subregion A, which does not jeopardize the ETAS' forecasting effort. We note that the problem of short-term aftershock incompleteness is still a standing issue, which becomes increasingly important for forecast models demanding real-time high-accuracy hypocenters for small magnitude events.

[21] We consider the branching ratio  $n = N_{\text{aft}}/N$  equal to 0.67 [Hardebeck, 2013], where  $N_{\text{aft}}$  is the number of events triggered by mainshocks [Helmstetter and Sornette, 2002]. We determine the productivity parameter  $K$  in relation with the true branching ratio  $n$ , which we estimate from the apparent branching ratio  $n_a$ , using the formalism for  $M_{\min} \leq m_d$  distinguishing between the minimum triggering magnitude  $M_{\min}$  and the detection magnitude  $m_d$ . [Werner, 2008]. We provide in the section A2 of the examples of parameter estimation for ETAS modeling for varying  $a$  values,  $b = 1.0$ ,  $c = 0.1667$  days for  $M_{\max} = 6.9$ ,  $M_{\min} = 2.5$ , and  $m_d = 3.0$ . The estimated parameters and implementation details for ETAS models are summarized in Table 2.

#### 4. Performance Evaluation of Forecasting Models

[22] In seismology, the basic idea behind using the average likelihood per event as a metric for the predictive power of the model was defined in the early work of Kagan and Knopoff [1977] and discussed for its theoretical basis by Harte and Vere-Jones [2005]. In recent years, worldwide initiatives, such as the Collaboratory for the Study of Earthquake Predictability and Regional Earthquake

Likelihood Models project, have proposed statistical tests regarding the efficient testing of earthquake forecasting models [Schorlemmer et al., 2007; Werner et al., 2010; Zechar et al., 2010; Rhoades et al., 2011; Marzocchi et al., 2012].

[23] We implement the suggested log likelihood statistics to different subregions along San Andreas Fault for approximately one year following the 1989 Loma Prieta rupture (see Figure 1b). During our evaluation period, we anticipate the influence of at least four mainshocks ( $M_L \geq 5.0$ ) within these subregions. Analytically, the southern San Andreas Fault in Santa Cruz Mountains (subregion A) represents the prolongation of the main aftershock zone of Loma Prieta but also the epicentral area of the Watsonville events, the  $M_L = 5.4$  Watsonville I (180 days after) and  $M_L = 5.1$  Watsonville II (112 min after Watsonville I). Studying this subregion is important since: (1) it holds ~40% of the Loma Prieta aftershock sequence but lies outside the main aftershock zone [Oppenheimer, 1990; Kilb et al., 1997; Beroza and Zoback, 1993], (2) the aftershock triggering mechanism is not clear; it may partly be a sympathetic slip on the Sargent Fault following the 1989 mainshock or/and postseismic deformation [Mueller and Johnston, 2000], and (3) the Watsonville sequence within the subregion presents the opportunity to study the nucleation of major aftershocks through a rate/state framework. The importance behind the near-source region of the Loma Prieta event (subregion B) is self-evident; the authors note that the effects of the  $M5.1$  aftershock at the northern part of the rupture, 37 min after the mainshock, are also considered.

[24] We use four statistical tests: (1) the modified  $N$  (number) test, comparing the number of forecast and observed events, (2) the  $S$  (space) test, for measuring the relative performance and the spatial consistency of each forecast, (3) the  $T$  test representing the information gain per earthquake and its variability within each model, and (4) the  $R$  (ratio) test, to determine the rejection percentage of a model when compared with a reference forecast. Possible biases of these tests have been discussed by Lombardi and Marzocchi [2010], Werner and Sornette [2008], and Woessner et al. [2011]. Rhoades et al. [2011] discuss how the  $R$  test can be potentially difficult to interpret in cases where we observe a mutual rejection on behalf of the models [Schorlemmer et al., 2007] and that it relies to computational time-consuming simulations.

[25] The metrics  $\delta_1$  and  $\delta_2$  of the modified  $N$  test (equations (A9) and (A10)) answer two questions assuming that the forecast is correct: first, What is the probability observing at least  $N_{\text{obs}}$  events?, and second, What is the probability of observing at most  $N_{\text{obs}}$  earthquakes? If the observations fall into the tails of the distributions, then we reject the forecasts as inconsistent with the observations at 95% significance level. We expect that  $\delta_1 \approx 1 - \delta_2$  because they share a complimentary role, and we reject a forecast model if either  $\delta_1(t) < \alpha_{\text{eff}}$  or  $\delta_2(t) < \alpha_{\text{eff}}$ , where  $\alpha_{\text{eff}} = 0.025$  corresponds to the effective significance value [Zechar et al., 2010]. For example, in the case that we have  $N_{\text{obs}} = 16$  earthquakes, we do not reject the forecast at the 0.05 significance level if it predicts between  $N_1 = 9$  and  $N_2 = 26$  events. We perform two types of the  $N$  test, an incremental one within 10 day intervals, and a cumulative one that includes the total number of events at time  $n \cdot 10$  days ( $n = 1, 2, \dots, 30$ ).

[26] We measure the spatial consistency of the forecast ( $S$  test) by comparing the distribution between observed ( $\omega$ )

and forecast ( $\lambda$ ) events using equations (A11) and (A12) [Schorlemmer, 2007; Zechar et al., 2010]. According to Zechar et al. [2010], the observed log likelihood, given by equation (A11), “has a negative value, and values closer to zero indicate a more likely observation; in other words, such a value indicates that the forecast shows better agreement with the observation.”

[27] We then compare the log likelihood score of the observed locations in the spatial forecast with the log likelihood values that we would expect to find if the forecast were correct. To obtain this last parameter, we simulate over 1000 synthetic samples generated by a Poisson distribution and compute their spatial likelihood value. We calculate the quantile  $\zeta$  of the simulated values that are less than the observed score as shown by the expression

$$\zeta = \frac{|\{\hat{S}_x | \hat{S}_x \leq \hat{S}\}|}{|\{\hat{S}\}|} \quad (5)$$

where in a given set  $A$ ,  $|\{A\}|$  denotes the number of elements in the set  $\{A\}$  and  $\hat{S}_x$  corresponds to the joint log likelihood of the  $x$ th simulated catalog. A low  $\zeta$  value returned by equation (5) would indicate an inconsistency with the observations since it would imply that the measured likelihood value is much lower than the expected if the forecast was correct. A high value would indicate that the observed score is higher than what we expected, which is not ground for rejection, since it still shows that the observed events are exactly at correct locations [Zechar et al., 2010]. We reject models at the 95% confidence level when  $\zeta < 0.05$  and examine the rejection ratio  $R_\zeta$ , based on the above quantile  $\zeta$ , thought our evaluation period.

[28] For the  $T$  test application [Rhoades et al., 2011] the sample information gain per earthquake of a model A over model B is defined by

$$I_N(A, B) = \frac{1}{N} \sum_{i=1}^N (X_i - \Upsilon_i) - \frac{\hat{N}_A - \hat{N}_B}{N} \quad (6)$$

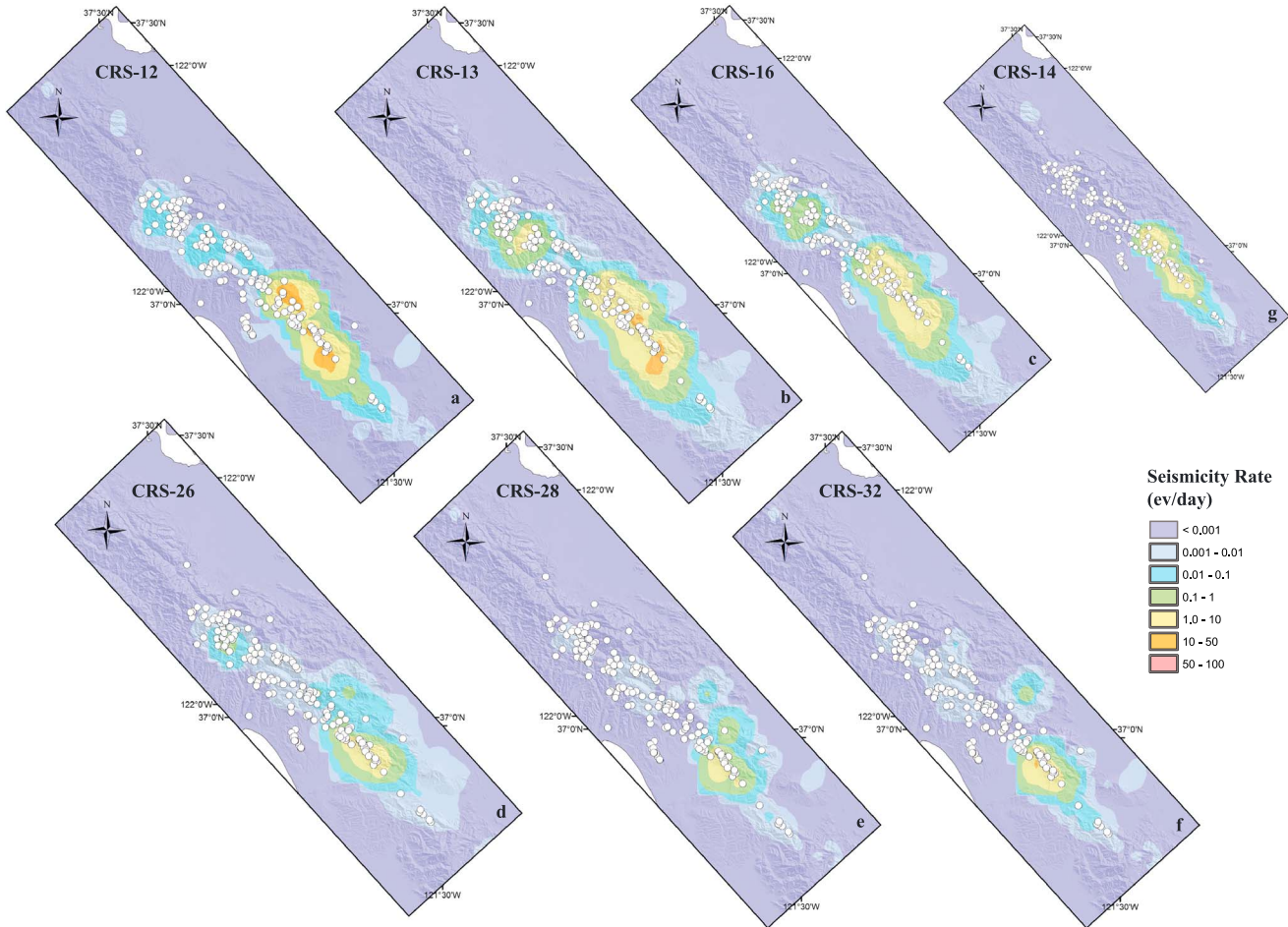
[29]  $I_N(A, B)$  is considered as the mean of a sample from a population with actual mean  $I(A, B)$ , where  $I_N(A, B)$  is the true information gain of model A over model B with  $X_i = \log \lambda_A(i)$  and  $Y_i = \log \lambda_B(i)$  the log likelihood value of a model A and B in the  $i$ th bin. In comparison with the  $R$  test, which tests whether model A can be rejected in favor of model B, the  $T$  statistic “measures the relative amount of information on future earthquakes provided by competing models, by computing a confidence interval for the information gain per earthquake” according to Rhoades et al. [2011] and it has also the benefit of being computationally less intensive.

## 5. Results

[30] We first discuss the forecast seismicity rate maps, describing earthquake occurrence in the first 10 days after the Loma Prieta mainshock, and comment on important features of our models.

### 5.1. Seismicity Rate Maps

[31] Selected CRS models are presented in Figure 2 with the aim of revealing major differences in the spatial distribution of forecasted seismicity due to different receiver



**Figure 2.** Maps of forecast daily seismicity rates with  $M \geq 3.0$  for physics-based models covering the first 10 days after the Loma Prieta mainshock for models based on (a) optimally oriented strike slip fault planes with San Andreas Fault (SAF) loading rate of 0.0065 MPa/yr and a gridded reference seismicity model, (b) optimally oriented strike slip fault planes with SAF loading rate of 0.0065 MPa/yr and a smoothed gridded reference seismicity model, (c) optimally oriented strike slip fault planes with SAF loading rate of 0.03 MPa/yr and a smoothed gridded reference seismicity model, (d) receiver planes derived from predominant geology with a SAF loading rate of 0.03 MPa/yr and a smoothed-gridded reference seismicity model, (e) receiver planes derived from predominant geology with a SAF loading rate of 0.07 MPa/yr and a gridded reference seismicity model, (f) shear stress changes on receiver planes derived from predominant geology with a SAF loading rate of 0.03 MPa/yr and a smoothed-gridded reference seismicity model, and (g) optimally oriented strike slip fault planes with SAF loading rate of 0.03 MPa/yr and a declustered background seismicity model. Models (Figure 2a–2e) are based on Coulomb stress change calculations with a friction coefficient of  $\mu = 0.2$ .

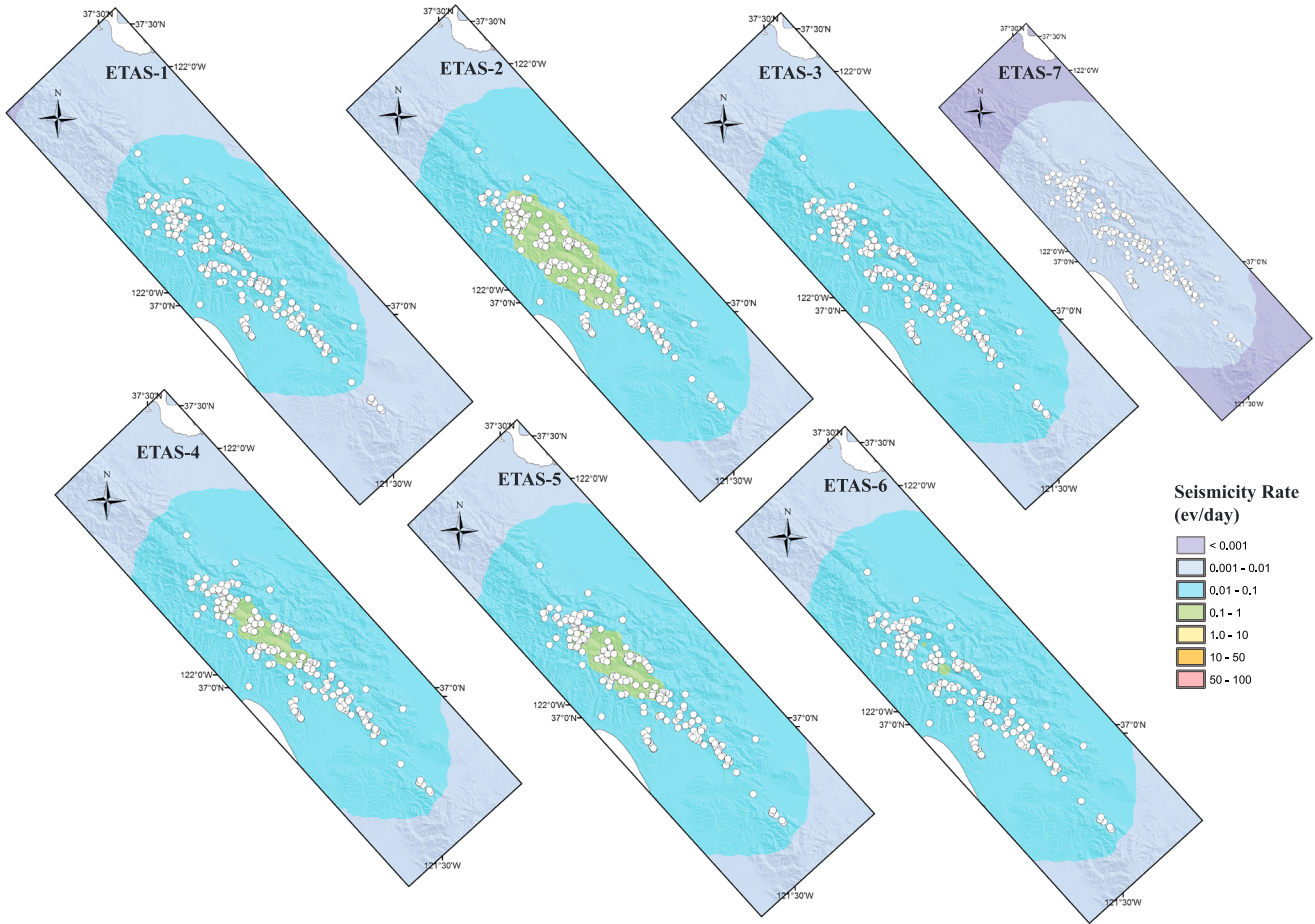
geometry (Figures 2a and 2b), loading rates (Figures 2b, 2c, and 2e), background seismicity formulations (Figures 2c and 2d), and friction coefficient (Figures 2b and 2f). For the statistical forecasts, results are compared between the ETAS-1 to ETAS-7 realizations (Figures 3a–3g) based on the different parameterization for  $a$ ,  $b$ ,  $p$ , and  $K$  values.

[32] For the CRS models, forecasts are based on the calculated static stress changes that are largest in the immediate vicinity of the mainshock plane. At the near source area, stress heterogeneity leads to negative Coulomb stresses changes, resulting in a decreased rate of seismicity, whereas at the ends of the mainshock fault, where the stress-increased lobes are calculated, the seismicity rates are increased. This dependency of the increased seismicity rates depending on

the fault model is one significant difference between physics-based and ETAS models. ETAS models tend to predict high forecast rates at the vicinity of the epicenter and do not have the capacity for rate reductions.

[33] When we compare CRS receiver fault formulations (optimally oriented strike slip faults (Figure 2c) and predominant geology (Figure 2d)), we find that the incorporation of thrust faults in the geology model aided forecasting of aftershocks on E-W striking reverse planes, whereas at the central San Andreas Fault, seismicity due to right/left lateral faulting subparallel to mainshock plane was adequately captured by optimally oriented strike slip faults. When different loading rates ( $\dot{\tau}$ ) are used, the resulting variability of forecasted rates is better represented in longer time intervals. For the





**Figure 3.** Maps of forecast daily seismicity rates with  $M \geq 3.0$  for the seven empirical-statistical ETAS models (see Table 2). It is worthy to note the difference in predicted seismicity rates between the ETAS-1 to ETAS-6 models and ETAS-7, which are updated daily and every 10 days, respectively.

examples shown in Figure 2 we are looking at the shortest term forecasting class, where the effect of  $\tau$  in the transient seismicity is less pronounced. However, we note that the observed seismicity of central to northern part of the rupture agrees with low loading rates around 0.0065 MPa/yr.

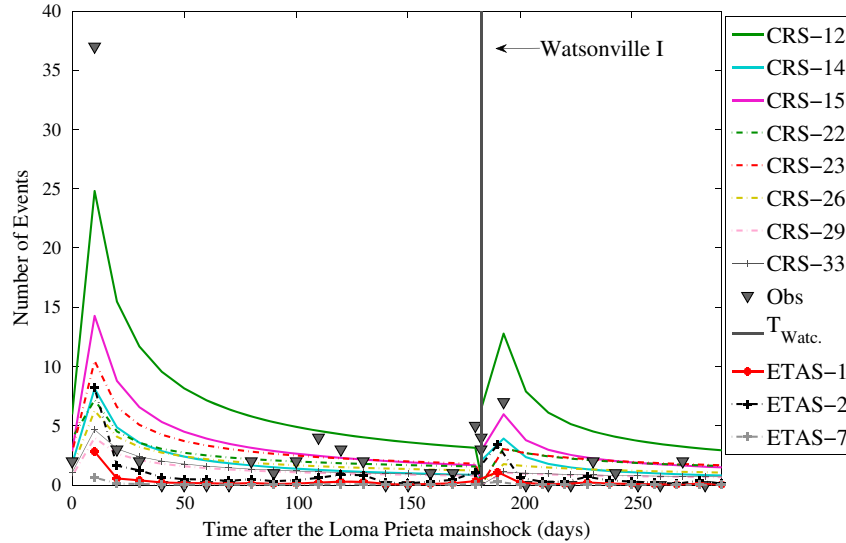
[34] Using a declustered catalog to derive our background model leads to lower forecast rates, especially where seismicity rates during the learning period (1974–1980) were already diminished, near the central and northern part of the Loma Prieta rupture as well as near the junction of San Andreas with Calaveras Fault (Figure 1a). When comparing the reference seismicity model formulations shown in Figures 2a and 2b, we get more distributed forecast rates when using smoothed formulations rather than gridded rates. Smoothed reference models improve the prediction of moderate rates at the northern edge of the Loma Prieta rupture (Figure 2b–2d). We note that the correlation distance of the smoothing operator should be compared with the spatial extent of the active faulting in the region. Forecast seismicity rates shown in Figures 2d–2f use predominant geology receiver planes for stress calculations, and also test the idea that adequate forecasting results can be achieved when incorporating shear (Figure 2f) instead of Coulomb stress changes (Figure 2e). We do not observe significant variability on the spatial distribution of seismicity rates between

Coulomb and shear stress change models, which is probably related with the low friction coefficient ( $\mu = 0.2$ ) for the Coulomb stress change calculations.

[35] For the ETAS model we observe that using a 10 day interval for the model update (Figure 3g) leads to lower forecasted rates along the Loma Prieta rupture. Moreover, the ETAS-2/4/5 models (Figure 3b, 3d, and 3e) predict higher rates at the near source region, which is related with larger  $K$  values and the trade off between  $\alpha$  and  $\gamma$  values. We also observe that the spatial extent of predicted daily rates with  $0.1 \leq N_{\text{daily}} \leq 1.0$  concentrates close to the edges of the Loma Prieta rupture for CRS models, following the form of the lobes related with Coulomb theory, whereas ETAS models appear more radially distributed around the triggering events, predicting adequate seismicity rates in the near-source region (see ETAS-2 in Figure 3b).

## 5.2. Log Likelihood Statistics

[36] In this section we present the results after the comparative performance evaluation for selected CRS and ETAS models for illustrations purposes based on log likelihood statistics for subregions A and B. We also provide, in the appendix of this article and as supplementary material, relevant results for all the models of this study, presented in Table A1.



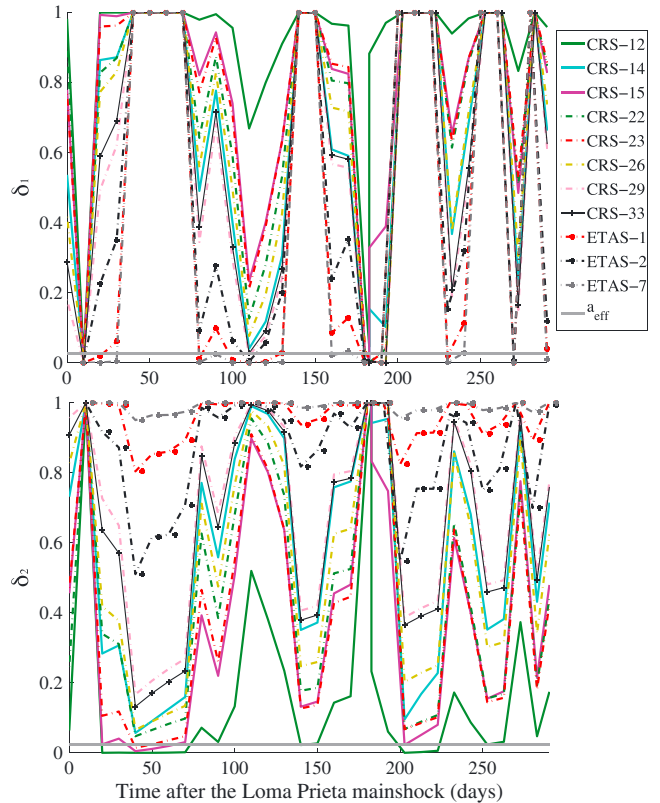
**Figure 4.** Observed (triangles) and predicted number of events with  $M_L \geq 3.0$  for selected models in the sub-region A within 10 day intervals for approximately one year following the Loma Prieta mainshock. We note that the first interval corresponds to  $\sim 37$  min between the Loma Prieta mainshock and the largest aftershock  $M5.1$ . Also, the time interval after the first Watsonville event is 112 min, constrained by the time of the second Watsonville event. A complete presentation of CRS and ETAS models developed in this study can be found in Figures A5 and A6, respectively.

### 5.2.1. Subregion A: $N$ Test

[37] In this paragraph we compare the forecasted number ( $N_{\text{fore}}$ ) to observed ( $N_{\text{obs}}$ ) of  $M \geq 3.0$  events at 10 day intervals for selected models during our evaluation period, beginning with the Loma Prieta mainshock and including the Watsonville events (Figure 4). We note that the interested reader can refer to Figures A5 and A6 for a complete presentation of CRS and ETAS models, respectively. We comment on model performance based on the representation of Figure 4 and the results of the modified  $N$  test (Figures 5 and 6).

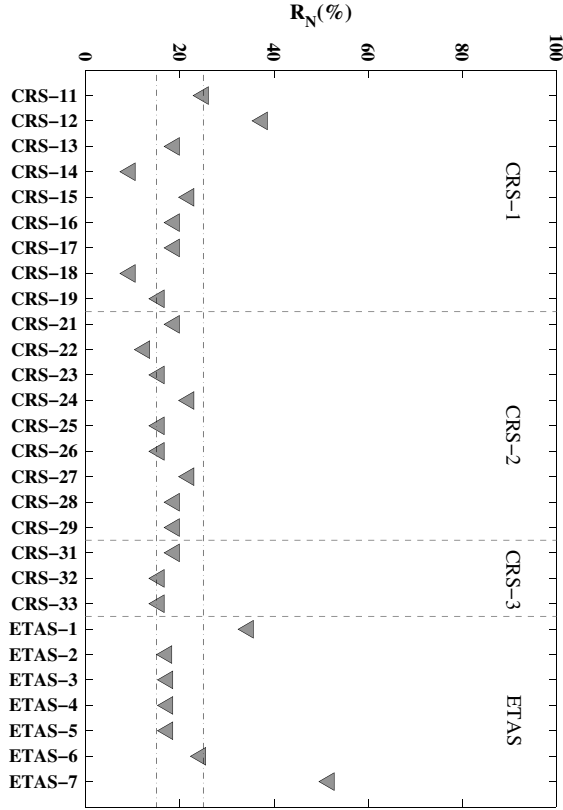
[38] All of our models underpredict aftershock rates during the first 10 days of the testing phase (at 10 day resolution; Figure 4). For ETAS models, there is significant variability in the number of forecast events between the first six ETAS ( $3 \leq N_F \leq 8$ ) model and ETAS-7 ( $N_F \approx 1$ ), which points out the importance of update interval for a successful short-term clustering model (see Figure 4 for ETAS-1/6/7 and Figure A6). We also note the difference between the ETAS-1 and the next 5 ETAS models (ETAS-2 to ETAS-6); in the ETAS-1 model we set  $\alpha = b = 1.0$  whereas ETAS-2 corresponds to a lower magnitude scaling formulation with  $\alpha < b$  (see Figure 4 for ETAS-1/2 and Figure A5b for ETAS-3 to 5). The latter assigns this way more importance to smaller magnitude events for cascading with an anticipated increase of productivity parameter  $K$  so that  $K_{\text{ETAS-1}} < K_{\text{ETAS-2}}$  (see Table 2). Furthermore, we observe from studying the quantile  $\delta_1$  that the ETAS-1 and ETAS-7 models underestimate the seismicity over the entire time period (Figure 5a).

[39] For CRS models, there is an obvious underestimation on behalf of all models immediately after the 1989 mainshock, which is caused by the low reference rate during the 1974–1980 learning period. However, as time progresses,



**Figure 5.** Quantile scores (top)  $\delta_1$  and (bottom)  $\delta_2$  of the incremental  $N$  test as a function of time for approximately one year following the Loma Prieta mainshock. The grey line indicates the 0.025 significance level beneath which the model is rejected. The two quantiles have complementary behavior and small values of  $\delta_1$  and  $\delta_2$  correspond to underestimation and overestimation of the number of observed events, respectively.





**Figure 6.** Rejection ratios  $R_N$  for the complete set of models based on the 10 day incremental test  $N$  test.

the physics-based models adjust up to the point of overforecasting at 20 days after the Watsonville events (see CRS-12 in Figure 4,  $t=200$  days) which is pointed out also by the quantile  $\delta_2 < 0.025$  (Figure 5b). Regarding CRS models we find that: (1) in short-term forecasting models based on optimally oriented strike slip planes (see CRS-14/15 models in Figures 4 and A5), predict approximately 32%–40% of the total number of observed events, (2) in long-term forecasting ( $t > 10$  days) models based on predominant geology bound the observations (Figure 4, see CRS-23/24/26). The above remark suggests that the underprediction in the first few days could be partly addressed by implementation of CRS models based on optimally oriented faults, but longer-term forecasts should be based on predominant fault structures. The CRS-25 model yields a balanced performance over the entire evaluation period exhibiting a complementary behavior in quantiles  $\delta_1$  and  $\delta_2$ , with the first ranging between 0.35 and 0.8.

[40] When we assess the rejection ratios  $R_N$  based on the 10 day incremental test  $N$  test (Figure 6) we note that: (1) within ETAS realizations, the ETAS-1, ETAS-6, and ETAS-7 models perform poorly ( $R_N > 24\%$ ), (2) within CRS implementation, the CRS-11/12 models have the worst performance ( $R_N > 25\%$ ), (3) the best performing models are CRS-14/18/19 and CRS-22/23/25/26 ( $10 < R_N < 15\%$ ) together with ETAS-2 to ETAS-5 ( $R_N = 17\%$ ), (4) CRS models based on shear stress changes have comparable performance with Coulomb stress based ones with rejection ratios  $R_N = 16$ –19% (see CRS-33 and CRS-29 in Figure 4), and (5) CRS models based on low uniform stressing rates across our study area, and resolved on optimally oriented faults, have higher rejection ratios

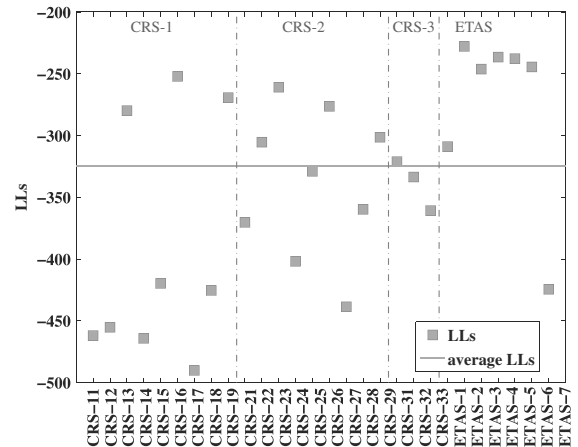
when compared with higher loading rate formulations with  $R_N > 19\%$  (see CRS-11/12/13 in Figure 6).

[41] In addition to the incremental  $N$  test discussed above, we also perform a cumulative  $N$  test (see also Figure A7). The cumulative  $N$  test sums over the number of forecast events at time  $t$  from the beginning of the testing period at 10 day resolution. We note that even if the cumulative test is useful for measuring the time-dependent performance of the model, the fact that it remembers previous incidental poor performances can be thought of as a disadvantage. In our study we find that the problem of early underprediction is inherited to later times, resulting to a total rejection ( $R_{NC} \geq 0.7$ ) of the 75% of forecasting models.

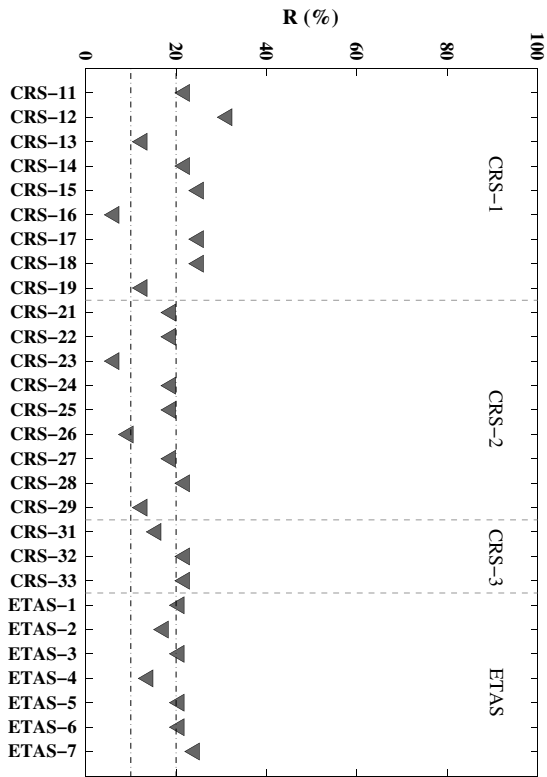
### 5.2.2. Subregion A: Consistency in Space ( $S$ Test) and Gain per Earthquake ( $T$ Test)

[42] In Figure 7 we present the joint log likelihood estimation to facilitate the comparison between forecast models (see also Table A2). Following the joint log likelihood metric for the entire evaluation period we find that: (1) the best performing models are the physics-based CRS-16/19, CRS-23/26 and the statistical models ETAS-2 to ETAS-6, with a 24 log-unit difference between the best performing ETAS and CRS model ( $LL_s^{ETAS-2} = -228$ ,  $LL_s^{CRS-16} = -228$ ) (2) for ETAS models there is a 18 log-units difference between the higher ( $LL_s^{ETAS-2} = -228$ ) and the lower joint log likelihood score ( $LL_s^{ETAS-1} = -308$ ) (3) for CRS models there is a 238 log-units difference between the higher ( $LL_s^{CRS-16} = -228$ ) and the lower joint log likelihood score ( $LL_s^{CRS-17} = -490$ ) and, (4) within the same generation and loading rate formulation there is a direction of improved performance going from declustered to gridded and finally to smoothed gridded reference seismicity (e.g., CRS-14/15/16 and CRS-17/18/19).

[43] Our  $N$  test conclusion for CRS forecasters, that CRS-2 models, present an improved performance for short-term forecasting, whereas for long-term forecasting CRS-3 models dominate, is further supported by the  $S$  test. Analytically, we rank the models for the short time class after the Loma Prieta mainshock ( $\text{Rank}^{\text{short,LP}}$ ) and we observe that CRS-1, CRS-2, and ETAS models are having three, two, and five positions among the first ten ranks, while for



**Figure 7.** Joint log likelihood values of the  $S$  test (Table A2) over the entire testing period. We note that the best performing models are the physics-based CRS-16/19, CRS-23/26 and the statistical models ETAS-2 to ETAS-6.



**Figure 8.** Rejection ratios  $R_{\zeta}$  for the complete set of models based on the  $S$  test.

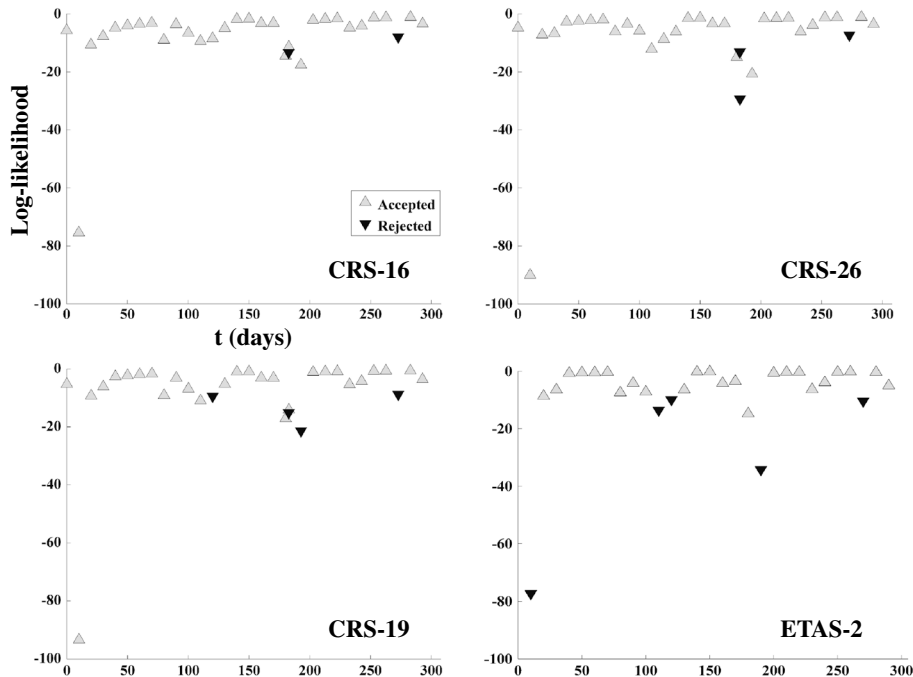
the long-term class CRS-2 and ETAS share those first ten positions (Table A2).

[44] We also calculate the rejection ratios  $R_{\zeta}$  for the 10 day  $S$  test and find that the CRS-16, CRS-23/26 models are rejected at less than 10%, followed closely by CRS-13/19, CRS-29, and ETAS-4 models ( $R_{\zeta} < 15\%$ ) and ETAS models, while the majority of the remaining models suffer rejection ratios of about 20% (Figure 8).

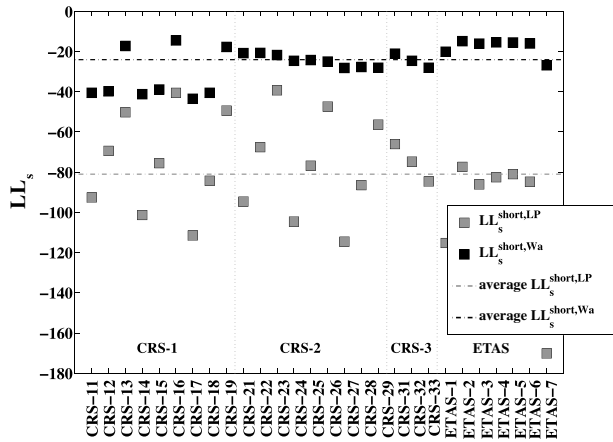
[45] In Figure 9 we present the log likelihood versus time for four best performing models during our entire evaluation period and we find that the majority of time intervals have log likelihood values smaller than 10 logarithmic units. We show that the forecast was rejected, based on the quantile  $\zeta$ , for only for few time intervals (red triangles), implying that the spatial consistency of the models is relatively static over the evaluation period. It is noteworthy that immediately after the Watsonville events ( $\sim t=180$  days), the CRS models perform poorly, and have smaller negative log likelihood values that are rejected (Figure 9).

[46] We evaluate the average log likelihood and standard deviation of the log likelihood sums over all spatial bins within the short-term and long-term forecasting classes after the Loma Prieta (LP) and the Watsonville (WS) mainshocks (Table A3). The results are used for a complementary relative ranking of the models within each time class and to address the question whether accounting for previous important ruptures in our study area improves the physics-based model performance.

[47] Under the  $LL_s^{\text{short, LP}}$  criterion, the best performing are the CRS-13/16/19 and CRS-23/26/29 (ranks 1–6), all of them based on smoothed gridded seismicity, and ETAS-2 model (rank 13). However, when we look the  $LL_s^{\text{short, WS}}$  values,



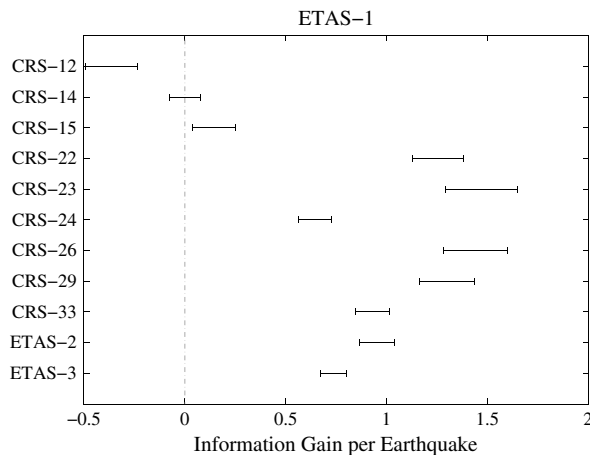
**Figure 9.** Log likelihood values of the  $S$  test versus time for models CRS-16, CRS-19, CRS-26, and ETAS-2 for approximately one year following the Loma Prieta mainshock. The time intervals for which a specific forecast model is rejected, based on quantile  $\zeta$ , are indicated with red triangles, whereas acceptance corresponds to green triangles.



**Figure 10.** Average negative log likelihood values of the  $S$  test for short-term periods following the Loma Prieta  $LL_s^{\text{short,LP}}$  and Watsonville  $LL_s^{\text{short,Wa}}$  events. For illustration purposes, log likelihood values have been saturated to  $-180$  units.

which correspond to the first 10 days after the Watsonville events, the CRS-16 model outperforms all forecast models, immediately followed by ETAS-1 to ETAS-6 models (Figure 10). It is important that the Watsonville sequence evolved within the testing subregion A, which favors short-term clustering models though earthquake cascading, resulting in higher ranks for ETAS models immediately after Watsonville events (see  $R^{\text{short, Wa}}$  based on  $LL_s^{\text{short, Wa}}$  in Table A3). In Figure 10 where we observe the smaller absolute log likelihood values for a short-period after Watsonville events in comparison with  $LL_s^{\text{short,LP}}$ , which points out that for physics-based models accounting for past major ruptures plays a crucial role.

[48] Our previous remark regarding the good performance of models based on predominant geology receivers is confirmed by comparing the first ten ranks under the  $LL_s^{\text{short,LP}}$  and



**Figure 11.** The  $T$  test results applied through our evaluation period for selected models. The plot shows the mean and 95% confidence interval of the information gain per earthquake when ETAS-1 is used as the reference model. Figure A9 presents the results of the  $T$  test application for the full range of models when ETAS-1 and CRS-13 are used as reference models.

$LL_s^{\text{long,LP}}$  criteria (see  $R^{\text{short,LP}}$  and  $LL_s^{\text{long,LP}}$  in Table A3), where CRS-1 models are absent in the long-term period and it is also illustrated in Figure A8, where we observe that for long-term after Loma Prieta, CRS-2 models are characterized by smaller absolute log likelihood values in respect with CRS-1 models.

[49] We also apply the  $T$  test which measures the relative amount of information on earthquakes provided by competing models, by computing a confidence interval for the information gain per earthquake [Rhoades et al., 2011]. We use two reference models, ETAS-1 and CRS-13, selected for their average general performance within their generation in the  $N$  and  $S$  test, described previously in this paper, and we compare each model with a specific reference model. The ETAS-1 model achieves a likelihood score  $-308$  (Table A2) over the entire time period, has a rejection ratio  $R_N=0.34$  and  $R_C=0.2$  at the incremental  $N$  test and  $S$  test, respectively, holding the 13th position over the entire time period. The CRS-13 model has a log likelihood score  $LL_s=-270$  (Table A2) and ranks among the best 15 models over the entire time period with rejection ratios for the incremental  $N$  test and  $S$  test less than 20%.

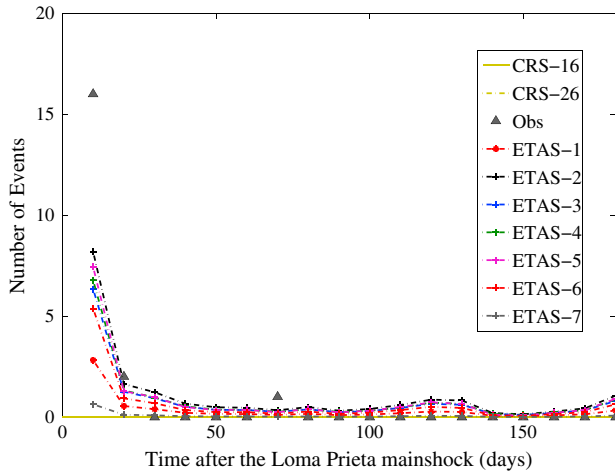
[50] In Figures 11 and A9, if a confidence interval for the information gain per earthquake does not intersect the vertical “zero” line, there is a significant difference in the information value of the two models. In Figure 11 we present the  $T$  test application for selected models and we observe that CRS-22/23/26/29, CRS-33, and ETAS-2 have a significant information gain exceeding 1, whereas CRS-12/14/15 provide less information value when compared with ETAS-1 model. In Figure A9 we present the results for the total number of models and we distinguish between three groups of models, depending on how informative they are in respect with ETAS-1 model; Group A with no significant gain in the information value with CRS-11/12/14/17 and ETAS-7, Group B with information gain reaching up to 0.8 with ETAS-15/18, CRS-21/24/27, CRS-33, and ETAS-3/4/5/6, and Group C with information gain greater than 0.8, corresponding to our most informative CRS-13/16/19, CRS-22/23/25/26/28/29, CRS-31/32, and ETAS-2. In case the CRS-13 model (Figure A9) serves as the reference model we observe that models CRS-16/19 and CRS-23/26/29 have information gain per earthquake of about 0.1–0.5.

[51] Based on our results on the  $T$  test (Figures 11 and A9), we select models from groups B (CRS-15, CRS-24, CRS-32) and C (CRS-16, CRS-26, ETAS-2) to apply the  $R$  test with ETAS-1 as reference models. In the  $R$  test, both the reference models  $H^r$  and  $H^i$  are used as null hypothesis to test against. In Table 3 we define the rejection ratios for both ways of testing  $R(LL^{ir})$  and  $R(LL^{ri})$ , respectively.

**Table 3.** Rejection Ratios Based on the Incremental  $R$  Test<sup>a</sup>

Model	$R(LL^{ri})$	$R(LL^{ir})$
CRS-15	0.31	0.41
CRS-16	0.17	0.21
CRS-24	0.24	0.27
CRS-26	0.20	0.28
CRS-32	0.21	0.27
ETAS-2	0.20	0.48

<sup>a</sup>Analytically, for the rejection ratio  $R(LL^{ri})$  the reference model ETAS-1 forms the null hypothesis, whereas for  $R(LL^{ir})$  the null hypothesis is formed by each model presented in this table.



**Figure 12.** Observed (triangles) and predicted number of  $M \geq 3.0$  for selected models in the subregion A within 10 day intervals following the Loma Prieta mainshock. The first interval corresponds to 37 min between the Loma Prieta mainshock and the largest ( $M=5.1$ ) aftershock. The CRS models predict number of events  $< 10^{-3}$ , which leads to severe underestimation of seismic activity for the near source area, mainly due to the low reference seismicity during our learning phase and the stress shadow imposed after the mainshock.

[52] We find that five out of six models provide better forecasts than ETAS-1 since they present  $R(L^{ir}) < 0.3$ , indicating that the test hypothesis is rejected less than 30% of our tests with the exception of CRS-15. The latter is also true when the reference model is used as the null hypothesis with the exception of the models CRS-15 and

ETAS-2 models, having rejection ratios  $R(L^{ri})=41\%$  and  $48\%$ , respectively, leading to a double rejection between the two models.

### 5.3. Log Likelihood Statistics in Subregion B

[53] We now examine subregion B, which lies close to the epicenter of the mainshock. We choose this test region because we want to address the question whether physics-based forecast models hold any predictive power in the near source region, and compare their performance against ETAS models. From the theoretical point of view, we anticipate that physics-based models are likely to fail because they oversimplify mainshock rupture heterogeneity, often not captured in the source representation.

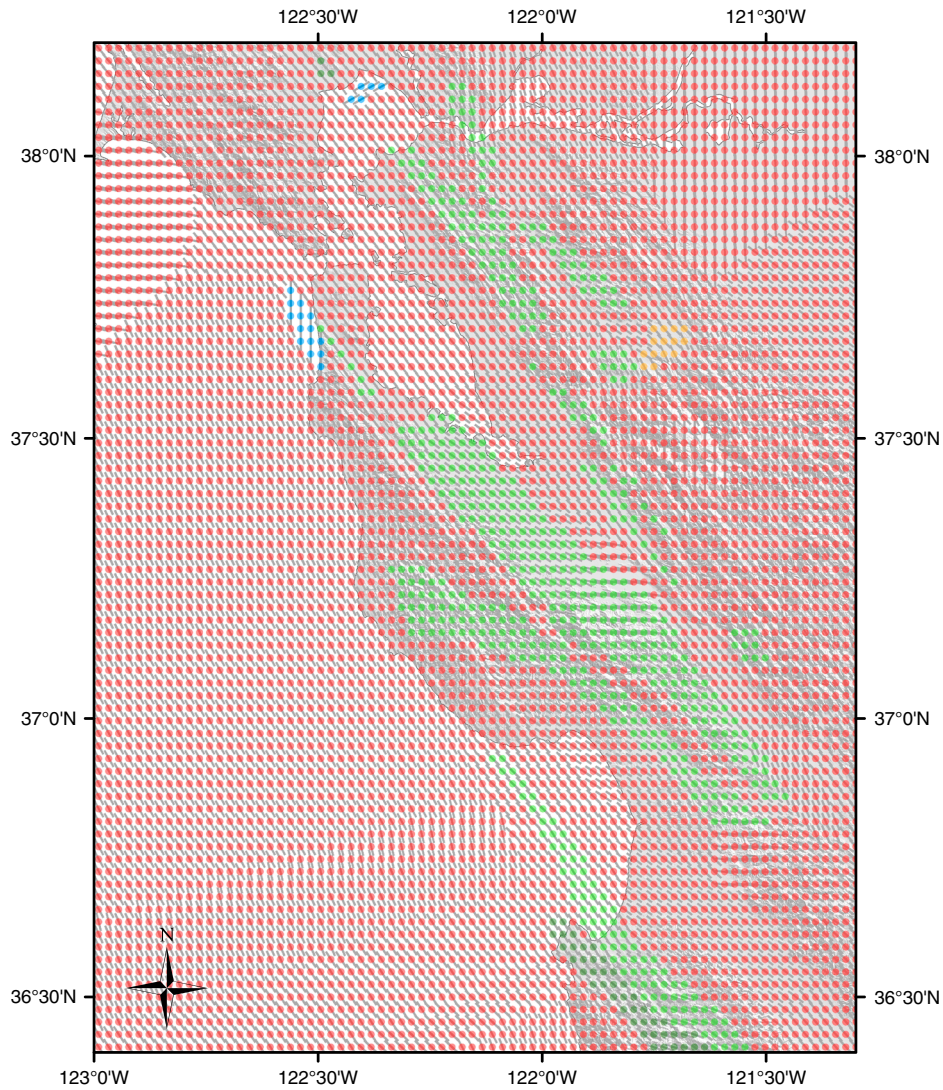
[54] We restrict our comparison among the best performing short-term forecast models identified in subregion A, and in locations within  $\sim 10$  km of the Loma Prieta epicenter (Figure 12). When we consider predominant geology fault planes as receivers, these locations are calculated to be under a stress decrease following the mainshock. During the first 10 days, the ETAS-2 model slightly underpredicts ( $N_{\text{ETAS-2}}=8 \approx N_1=9 < N_{\text{Obs}}=16 < N_2=26$ ), where  $N_1$  and  $N_2$  correspond to the lower and upper bounds given from equations (A9) and (A10), assuming the quantiles  $\delta_1 = \delta_2 = 0.025 = a_{\text{eff}}$ . Evaluating the ETAS models performance based on  $\delta_1, \delta_2$  quantiles we find rejection ratios less than 0.04 with the exception of ETAS-7, which is rejected at 10% of the tests over the entire evaluation period. On the other hand, CRS models fail for two reasons: (1) the low reference seismicity rate prior to the mainshock and (2) the calculated stress decrease from the mainshock which does not correspond to an observed rate decrease; a common problem in

**Table A1.** Main Features and Free Parameters of the Physics-Based Forecast Models (CRS)<sup>a</sup>

Model-Generation Number	Learning Phase Formulation	Friction Coefficient $\mu$	Receiver Plane Formulation	Reference	SAF Loading Rate (MPa/yr)
CRS-11	DE	0.2	OOPs	This study	0.0064
CRS-12	GR	0.2	OOPs	This study	0.0064
CRS-13	SMGR	0.2	OOP	This study	0.0064
CRS-14	DE	0.2	OOPs	P02	0.03
CRS-15	GR	0.2	OOPs	P02	0.03
CRS-16	SMGR	0.2	OOPs	P02	0.03
CRS-17	DE	0.2	OOPs	SmSan03	0.07
CRS-18	GR	0.2	OOPs	SmSan03	0.07
CRS-19	SMGR	0.2	OOPs	SmSan03	0.07
CRS-21	DE	0.2	PrG	This study	0.0064
CRS-22	GR	0.2	PrG	This study	0.0064
CRS-23	SMGR	0.2	PrG	This study	0.0064
CRS-24	DE	0.2	PrG	P02	0.03
CRS-25	GR	0.2	PrG	P02	0.03
CRS-26	SMGR	0.2	PrG	P02	0.03
CRS-27	DE	0.2	PrG	SmSan03	0.07
CRS-28	GR	0.2	PrG	SmSan03	0.07
CRS-29	SMGR	0.2	PrG	SmSan03	0.07
CRS-31	GR	0.0	PrG	This study	0.0064
CRS-32	GR	0.0	PrG	P02	0.03
CRS-33	GR	0.0	PrG	SmSan03	0.07

<sup>a</sup>Each CRS is named after the Generation (in bold) and its number, e.g., CRS-12, is the second model of Generation 1; the three CRS Generations ( $i=1..3$ ) correspond to models based on Coulomb stress changes resolved on Optimally Oriented (strike-slip) Planes for Failure (OOPs), Predominant Geology (PrG) receivers and shear stress changes resolved on PrG grid. It is noted that DE, GR, and SMGR correspond to background seismicity models, gridded into  $2.5 \text{ km} \times 2.5 \text{ km}$  cells, derived from a declustered catalog (DE) [Reasenber, 1985], a gridded (GR) and a smoothed gridded model (SMGR) using a Gaussian operator with correlation distance 5 km. For SAF, stressing rates follow calculations by Smith and Sandwell [2003], Smith-Konter and Sandwell [2009], and Parsons [2002] noted here as SmSan03 and P02, respectively.





**Figure A1.** Grid of discrete faults covering our study area. Geometrical characteristics have been assigned based on prevailing deformation styles of major faults and focal mechanism data (Figure A2) available during the learning phase (1974–1980) [e.g., *Herd, 1979*] of our forecast experiment. Line segments represent the hypothetical fault trace whereas color points reveal the sense of faulting. Analytically, red, light and dark green, and blue dots correspond to strike slip, thrust, reverse, and normal faulting, respectively.

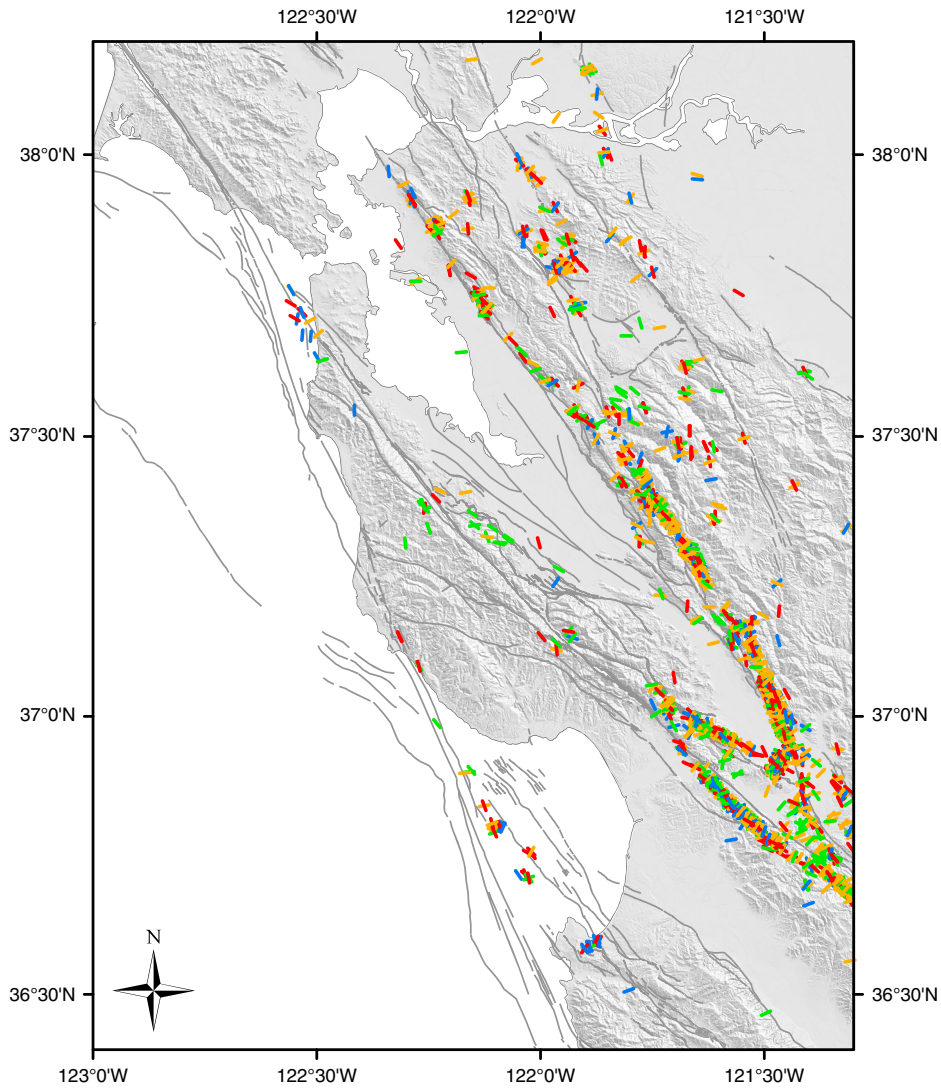
areas close to mainshocks [e.g., *Parsons et al., 2012*]. We observe intervals that predict within acceptable limits for CRS-models at periods longer than 20 days, but we are reluctant to consider this a success, since the overall underforecasting behavior is paired here with the variability of observed events within certain intervals where  $N_{\text{obs}} = 0$ .

## 6. Conclusions

[55] In this study we evaluate retrospective earthquake forecasts for the period between 1980 and 2009 for Northern California, focusing primarily on the first year after the 1989  $M_w = 6.9$  Loma Prieta earthquake including the effects of the largest  $M = 5.1$  aftershock and the two Watsonville events ( $M_L = 5.4$  and  $M_L = 5.1$  within 2 h) that occurred 40 min and 182 days after Loma Prieta. We investigate

the performance of short ( $\leq 10$  days after the mainshock) and long-term forecasts (10–300 days) in the near-source region and at greater distances with our goal being to reveal the strength and weaknesses of physics-based and short-term clustering models in time and space.

[56] Within distances of few rupture lengths, the total number of observed earthquakes is forecasted adequately by the CRS-14/18/19, CRS-23/25/26, and ETAS-2/3/4/5 models with corresponding rejection ratios for the incremental  $N$  (number) test over the entire evaluation period less than 18%. We find that: (1) among physics-based models, the ones based on smoothed gridded seismicity and stressing rates on San Andreas of about 0.03–0.07 Mpa/yr present the higher predictive power (CRS-16/19 and CRS-23/26/29), (2) at short-term forecasts optimally oriented receivers perform better capturing the initial high variability of the stress field, represented by the diversity of focal mechanisms, whereas at



**Figure A2.** PPFIT solutions for earthquakes within the learning phase of our forecast experiment between 1974 and 1980, taken from the ANSS catalog. Line segments represent the identified as causative fault plane with red, orange, green, and blue representing right and left-lateral, reverse and normal faulting, respectively.

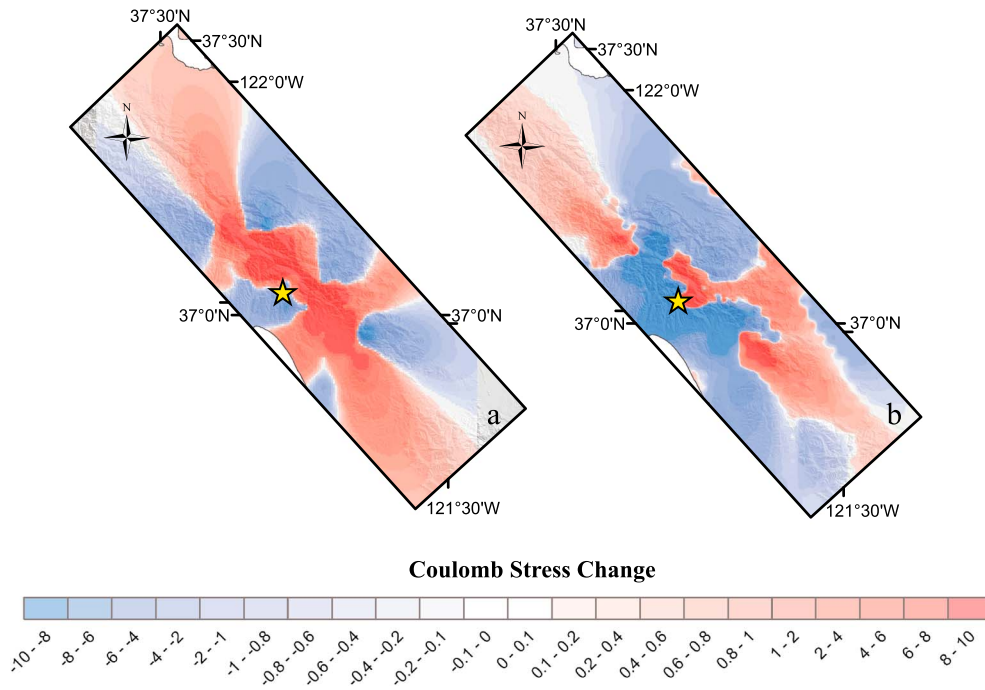
long-term predominant geology receivers have a higher predictive power, (3) ETAS models updated daily and incorporating smaller magnitude events ( $M \geq 2.5$ ) for cascading are among the 10 first ranks over the entire evaluation period, (4) ETAS models receive a performance boost after the Watsonville events with epicenters within our evaluation area which underlines how important is the proximity of the mainshock for such models, and (5) physics-based models based on shear stress changes have comparable performance with their Coulomb stress-based counterparts.

[57] The near-source area, lying under a stress shadow [Harris and Simpson, 1996] after the Loma Prieta mainshock, is a setting in which physics-based forecasts are known to struggle [e.g., Parsons et al., 2012]. Not surprisingly, here the ETAS models outperform all CRS models in the  $N$  test since the combination of low reference seismicity rate during the learning phase and the calculated stress decrease result in severe underforecasting. Spatial consistency tests for ETAS

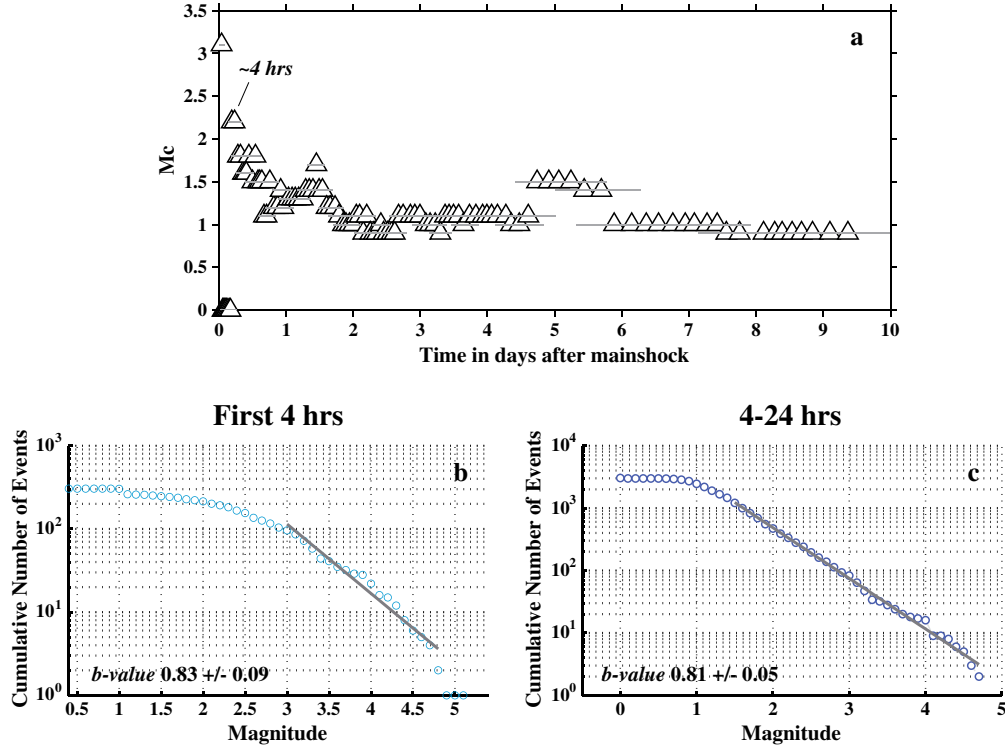
show that we have balanced performance for short-to-medium term forecasting in the near-source region.

[58] It is evident, especially for CRS models, that the forecasts are problematic when the seismicity during the learning phase is nonexistent or extremely low, and future research has to incorporate time-varying background rate in physics-based and statistical forecasting models to overcome this problem. Furthermore, we note that during any retrospective experiment related with earthquake catalogs spanning the last 30–40 years, we inevitably test the spatial performance of our models with higher-quality data than what we used to develop them. Usually, forecast experiments target regions with aftershock sequences of mainshocks between  $M6.5$  and  $M7.0$  but for such major earthquakes, there is an important variability from sequence to sequence. The variability is related with many parameters such as the presence of precursory phenomena (e.g., foreshocks), the stress drop of the mainshock in respect to the regional stress field, the heterogeneity of the source and the triggering potential

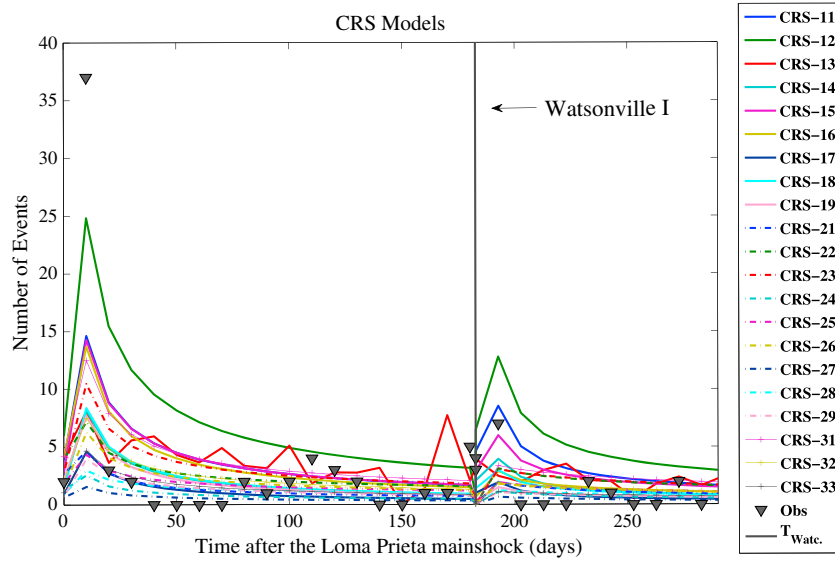




**Figure A3.** Calculated Coulomb stress changes following the  $M_w=6.9$  1989 Loma Prieta mainshock using the *Beroza* [1991] source model resolved on (left) optimally oriented fault planes (CRS-1) and (right) predominant geology grid (CRS-2).



**Figure A4.** Magnitude of completeness  $M_c$  following the Loma Prieta mainshock. (a)  $M_c$  versus time for the first 10 days, (b) frequency-magnitude distribution for the first 4 h, and (c) frequency-magnitude distribution for the remaining first day (5–24 h). We observe that for the first few hours the catalog is complete down to  $M_L 3.0$  ( $b$  value  $0.83 \pm 0.09$ ) while later in the same day it reaches  $M_L 1.5$  ( $b$  value  $0.81 \pm 0.05$ ).



**Figure A5.** Observed (triangles) and predicted number of events with  $M_L \geq 3.0$  for CRS models in the subregion A within 10 day intervals for approximately one year following the Loma Prieta mainshock. We note that the first interval corresponds to  $\sim 37$  min between Loma Prieta mainshock and the largest aftershock  $M5.1$ . Also, the time interval after the first Watsonville event is 112 min, constrained by the time of the second Watsonville event.

among neighboring faults influencing the performance of physics-based and statistical models in different ways.

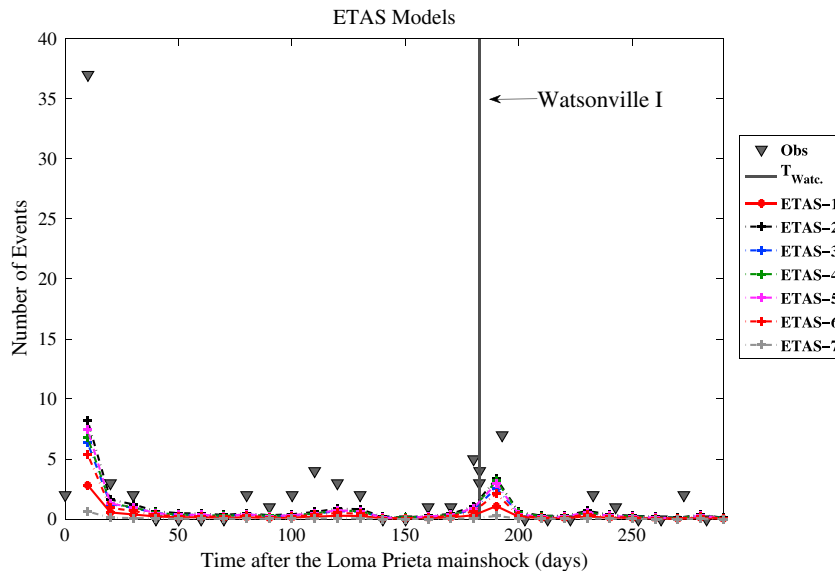
[59] From the scientific perspective, the aim is to improve our insight in a prospective manner by understanding each time which underlying physical hypothesis lead to a higher predictability during our retrospective experiment. In this paper we have extensively studied Northern California, focusing on the Loma Prieta mainshock, but in order to generalize our results, more aftershock sequences from various geotectonic environments should be modeled. In this way we add to our knowledge behind the physical

mechanism of earthquake triggering, which is of paramount importance for developing better forecasts in the future.

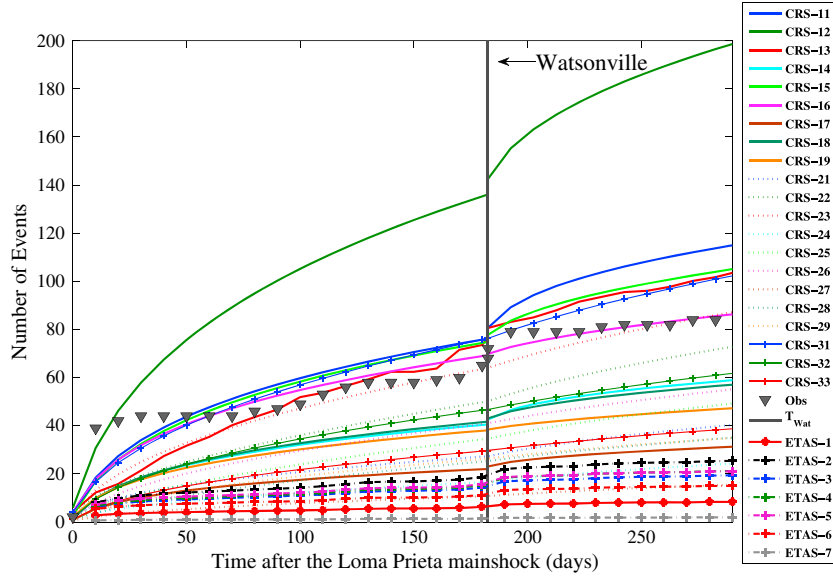
**Appendix A: Equations for the Rate/State and ETAS Implementation, Development of Predominant Geology Fault Grid, Definition of Statistical Tests, and Supplementary Figures**

**A1. Rate-and-State Equations**

[60] We use the expression for seismicity rate  $R$  as a function of state variable  $\gamma$  under secular tectonic shear stressing



**Figure A6.** Observed (triangles) and predicted number of events with  $M_L \geq 3.0$  for ETAS models in the subregion A within 10 day intervals for approximately one year following the Loma Prieta mainshock.



**Figure A7.** Cumulative observed (triangles) and predicted number of events with  $M_L \geq 3.0$  in the sub-region A for approximately one year following Loma Prieta mainshock. We note that the first interval corresponds to  $\sim 37$  min between Loma Prieta mainshock and the largest aftershock  $M5.1$ . Also, the time interval after the first Watsonville event is 112 min, constrained by the time of the second Watsonville event.

rate  $\tau$ . Under constant shear stressing rate at each location  $\gamma$  reaches a steady state, and is expressed as

$$\gamma_o = \frac{1}{\tau} \quad (\text{A1})$$

[61] At steady state, the seismicity rate  $R$  equals the background rate  $r$  since  $R$  is given by

$$R = \frac{r}{\gamma\tau} \quad (\text{A2})$$

[62] If there is no stress perturbation, then the seismicity rate is constant. When there is a perturbation, the state variable of the system  $\gamma_{n-1}$  before the event evolves coseismically to a new value  $\gamma_n$

$$\gamma_n = \gamma_{n-1} \exp\left(\frac{-\Delta CFF}{A\sigma}\right) \quad (\text{A3})$$

where  $A\sigma$  is the rate/state constitutive parameter  $A$  times multiplied by the effective normal stress, assumed to be 0.05 MPa [Toda *et al.*, 2005]. The seismicity rate during the time  $\Delta t$  is transient, and given sufficient time, recovers, providing a new state variable for the system given by

$$\gamma_{n+1} = \left[\gamma_n - \frac{1}{\tau}\right] \exp\left[\frac{-\Delta t\tau}{A\sigma}\right] + \frac{1}{\tau} \quad (\text{A4})$$

[63] The duration of the transient effects is inversely proportional to the tectonic loading rate, implying that if we give sufficient time, even in cases of slow stressing rates, the transient seismicity will eventually disappear.

## A2. ETAS Development

[64] For ETAS-1 model: For parameter values  $a = b = 1.0$ , the apparent branching ratio  $n_a$ , is given by the equation

$$n_a = \frac{kb \ln(10)(M_{\max} - m_d)}{1 - 10^{-b(M_{\max} - M_{\min})}} \quad (\text{A5})$$

[65] And the relation between  $n$  and  $n_a$  is

$$n_a = n \left( \frac{M_{\max} - m_d}{M_{\max} - M_{\min}} \right) \quad (\text{A6})$$

[66] From the implementation of the above equation we take  $n = 0.67$ ,  $n_a = 0.59$ , and  $K = 0.0619$ .

[67] For ETAS-2 model: For parameter values  $\alpha < b = 1.0$ , the apparent branching ratio  $n_a$  is given by the equation

$$n_a = \frac{kb}{b - \alpha} \left( \frac{10^{-(b-\alpha)(m_d - M_{\min})} - 10^{-(b-\alpha)(M_{\max} - M_{\min})}}{1 - 10^{-b(M_{\max} - M_{\min})}} \right) \quad (\text{A7})$$

and the relation between  $n$  and  $n_a$  is

$$n_a = n \left( \frac{10^{(b-\alpha)(M_{\max} - m_d)} - 1}{10^{(b-\alpha)(M_{\max} - M_{\min})} - 1} \right) \quad (\text{A8})$$

[68] In the case of  $\alpha = 0.8$  we take  $n = 0.67$ ,  $n_a = 0.51$ , and  $K = 0.1543$ . The values are in close agreement with Werner *et al.* [2011], who used  $M \geq 2.0$  for target magnitude range  $M \geq 3.95$ . We consider an  $a$  value ranging between 0.7 and 0.9 for ETAS-2 up to ETAS-6, respectively, and calculate in each case the apparent and true branching ratios as well as  $K$  value using the equations above.

**Table A2.** Joint Log Likelihood (LL<sub>s</sub>) for the Full Testing Period<sup>a</sup>

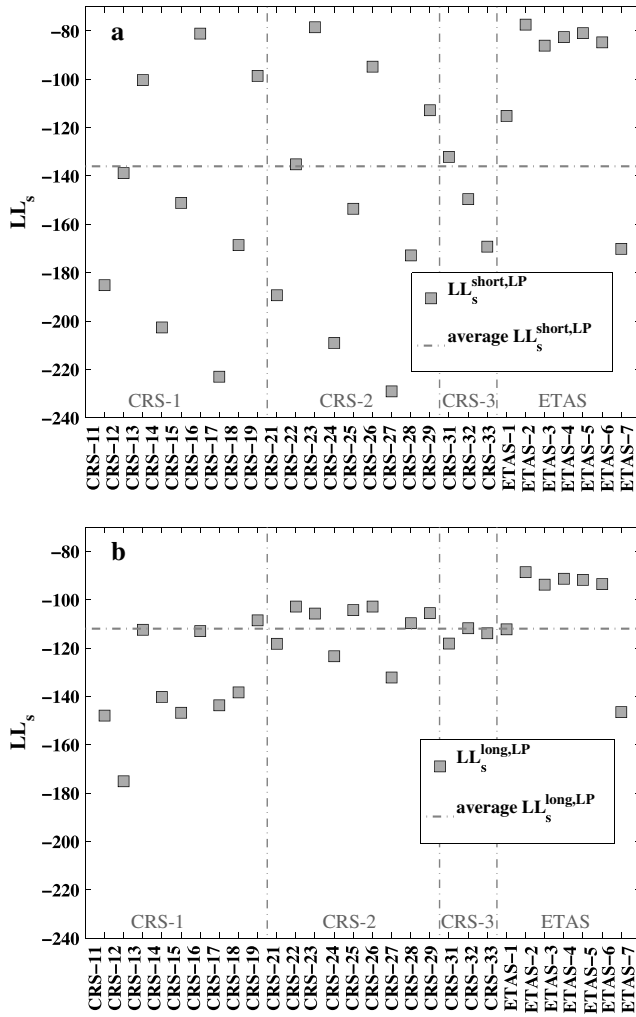
Model	LL <sub>s</sub>	Rank	LL <sub>s</sub> <sup>short, LP</sup>	Rank <sup>short, LP</sup>	LL <sub>s</sub> <sup>long, LP</sup>	Rank <sup>long, LP</sup>
CRS-11	-462.00	26	-185.14	23	-147.94	27
CRS-12	-455.30	25	-138.79	15	-175.10	28
CRS-13	-279.86	10	-100.32	10	-112.42	15
CRS-14	-464.12	27	-202.62	25	-140.21	24
CRS-15	-419.74	21	-151.12	17	-146.73	26
CRS-16	-252.04	6	-81.12	4	-112.96	16
CRS-17	-490.14	28	-223.02	27	-143.64	25
CRS-18	-425.39	23	-168.56	19	-138.37	23
CRS-19	-269.37	8	-98.61	9	-108.56	12
CRS-21	-370.33	19	-189.28	24	-118.27	19
CRS-22	-305.39	12	-135.20	14	-102.86	7
CRS-23	-260.79	7	-78.50	2	-105.76	11
CRS-24	-401.84	20	-209.09	26	-123.39	21
CRS-25	-328.92	12	-153.55	18	-104.32	9
CRS-26	-276.27	9	-94.74	8	-102.88	8
CRS-27	-438.60	24	-229.07	28	-132.18	22
CRS-28	-359.65	17	-172.84	22	-109.74	13
CRS-29	-301.38	11	-112.77	11	-105.52	10
CRS-31	-321.10	14	-132.06	13	-118.12	18
CRS-32	-333.63	16	-149.56	16	-111.76	14
CRS-33	-360.71	18	-169.18	20	-113.92	17
ETAS-1	-308.83	13	-115.19	12	-92.15	6
ETAS-2	-227.88	1	-77.33	1	-73.83	1
ETAS-3	-246.10	5	-86.05	7	-77.71	5
ETAS-4	-236.44	2	-82.51	5	-75.95	2
ETAS-5	-237.74	3	-80.94	3	-76.26	3
ETAS-6	-244.44	4	-84.71	6	-77.53	4
ETAS-7	-424.40	22	-170.14	21	-119.64	20

<sup>a</sup>LL<sub>s</sub><sup>short, LP</sup> and LL<sub>s</sub><sup>long, LP</sup> correspond to the joint log likelihood within the short (<10 days) and long-term forecasting classes after the mainshock until Watsonville occurrence, respectively.

**Table A3.** Average Log Likelihood (LL<sub>s</sub>) and Ranks (R)<sup>a</sup>

Models	LL <sub>s</sub> <sup>short, LP</sup>	LL <sub>s</sub> <sup>long, LP</sup>	σ <sup>LP</sup>	LL <sub>s</sub> <sup>short, Wa</sup>	LL <sub>s</sub> <sup>long, Wa</sup>	σ <sup>Wa</sup>	R <sup>short, LP</sup>	R <sup>long, LP</sup>	R <sup>short, Wa</sup>	R <sup>long, Wa</sup>
CRS-11	-92.56	-4.90	9.0	-40.48	-3.16	4.2	21	24	25	25
CRS-12	-69.39	-8.22	7.4	-39.78	-4.46	3.8	9	28	24	27
CRS-13	-50.16	-6.68	2.5	-17.20	-3.22	1.9	5	27	7	26
CRS-14	-101.30	-3.38	9.1	-41.20	-1.95	4.6	23	1	27	20
CRS-15	-75.56	-4.90	7.3	-38.92	-2.74	4.2	11	23	23	24
CRS-16	-40.56	-4.81	2.4	-14.46	-1.92	2.0	2	22	1	19
CRS-17	-111.50	-3.84	9.3	-43.47	-1.02	5.1	25	3	28	12
CRS-18	-84.28	-3.58	7.4	-40.53	-1.70	4.6	16	2	26	18
CRS-19	-49.30	-4.16	3.1	-17.71	-1.01	2.6	4	8	8	11
CRS-21	-94.64	-3.86	7.6	-20.72	-1.51	1.3	22	4	11	14
CRS-22	-67.59	-4.47	3.8	-20.62	-2.44	0.8	8	17	10	21
CRS-23	-39.25	-4.77	2.4	-21.68	-2.57	1.6	1	20	13	22
CRS-24	-104.54	-4.05	7.7	-24.55	-0.88	1.8	24	7	15	8
CRS-25	-76.77	-4.35	4.1	-24.25	-1.60	1.3	12	13	14	16
CRS-26	-7.37	-4.64	2.9	-24.97	-1.53	2.1	3	18	17	15
CRS-27	-114.53	-4.23	8.0	-28.11	-0.53	2.3	26	10	22	5
CRS-28	-86.42	-4.30	4.4	-27.69	-1.04	1.7	20	12	19	13
CRS-29	-56.38	-4.78	3.4	-28.08	-0.91	2.6	6	21	21	9
CRS-31	-66.02	-4.75	3.9	-21.08	-2.66	0.8	7	19	12	23
CRS-32	-74.78	-3.91	4.2	-24.62	-1.63	1.3	10	5	16	17
CRS-33	-84.58	-3.96	4.6	-28.02	-0.97	1.8	17	6	20	10
ETAS-1	-115.19	-5.15	5.3	-20.06	-0.19	4.2	27	25	9	1
ETAS-2	-77.33	-4.17	3.9	-14.79	-0.60	3.3	13	9	2	7
ETAS-3	-86.05	-4.40	4.3	-16.03	-0.45	3.5	19	16	6	3
ETAS-4	-82.51	-4.27	4.2	-15.41	-0.53	3.4	15	11	3	6
ETAS-5	-80.94	-4.35	4.2	-15.47	-0.40	3.7	14	14	4	2
ETAS-6	-84.71	-4.39	4.2	-15.99	-0.47	3.5	18	15	5	4
ETAS-7	-170.14	-6.50	7.0	-26.82	-0.04	5.4	28	26	18	26

<sup>a</sup>We note that for short and long-term classes after the Loma Prieta mainshock (LL<sub>s</sub><sup>short, LP</sup>, LL<sub>s</sub><sup>long, LP</sup>) and Watsonville (LL<sub>s</sub><sup>short, Wa</sup>, LL<sub>s</sub><sup>long, Wa</sup>) events ~170 days after. Our estimates for long-term class after Loma Prieta corresponds to the time period 10 days after Loma Prieta mainshock until the time of the Watsonville events.

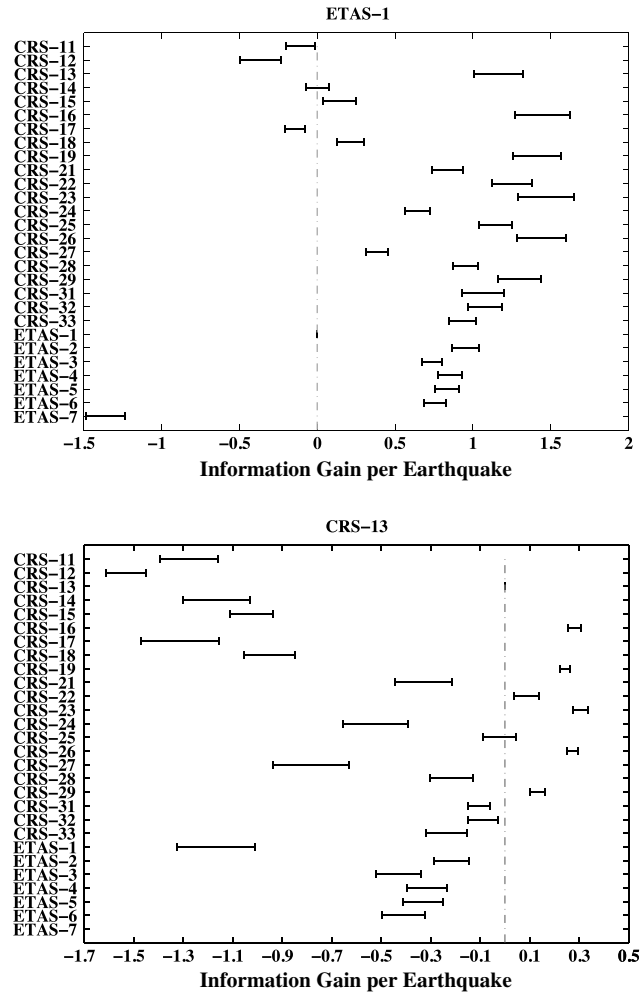


**Figure A8.** Joint log likelihood values of the  $S$  test for (a) short ( $t \leq 10$  days) and (b) long-term forecasting ( $t \geq 10$  days) (Table A2) following the Loma Prieta mainshock until the Watsonville I event.

**A3. Development of Predominant Geology Grid**

[69] We have set the 3-D fault plane grid (Figure A1) to comply with the following concepts: (1) we only use seismological data to influence our definitions of faulting style from the same period of the forecast learning period (1974–1980) and (2) we describe primary geological features that are consistent with the long term deformation history (major faults and recent topography). An important task for our receiver fault model is to explain regional seismicity changes; it thus cannot only be defined by sparsely distributed faults. Instead, we cover our study region with discrete fault sections on a  $2.5 \times 2.5$  km grid, so that every point has a defined receiver fault orientation that is either directly associated with the mapped faults and seismicity inside the learning phase, or is interpolated/extrapolated from those data. Every fault section is 2 km long, and can either be thought of in most cases as part of a longer fault system, or representative of sources of “background” seismicity occurring on small fractures or other unmapped faults while its direction was taken from the closest 2 km fault segment.

[70] We assess whether rupture variability is adequately expressed in our fault plane grid by comparing with PFFIT solutions [Reasenber and Oppenheimer, 1985] between 1974 and 1980 (Figure A2). We find that there is noteworthy compressive strain accommodated by vertical deformation expressed with reverse or thrust events near the strike slip junctions of the San Andreas peninsula with southern Calaveras and Hayward with northern Calaveras. From this analysis, inclusion of the Sargent and Mission faults is necessary because they accommodate deformation within blocks bounded by major strike slip faults; as an aside, knowledge not available from our test period confirms this [Andrews et al., 1993; Manaker et al., 2005]. The broadly interpolated view of the San Francisco Bay region geology we employ was available during our test period [e.g., Herd, 1979], and can be confirmed by later reports. For example, reverse and thrust belts are described more fully for their geometrical characteristics in the U.S. Geological Survey (USGS) Complete Reports on Quaternary Faults [http://geohazards.usgs.gov/cfusion/qfault/index.cfm].



**Figure A9.** The  $T$  test results applied through our evaluation period for selected models. The plot shows the mean and 95% confidence interval of the information gain per earthquake when ETAS-1 and CRS-13 are used as reference models.

[71] Other aspects of note from the PPFIT interpretation of 1974–1980 seismological data include N-S strike slip faulting near the junction between the northern and central Calaveras fault, and distributed normal faulting offshore of San Francisco and in the San Pablo Bay area. We thus have allowed for small scale normal faulting in these two cases; the first with N-S trending normal faults dipping to the NE at the northernmost edge of the peninsula section of the San Andreas fault segment, and second, at the dilatational step over between the right-lateral Hayward and Rodgers Creek faults, expressed through normal faults striking NE-SW, and dipping to the southeast.

**A4. Definition of Statistical Tests**

**A4.1. Modified  $N$  Test**

[72] The modified  $N$  test evaluates the consistency between the forecast and observed number of events within a test area. Zechar et al. [2010] improved the original  $N$  test metric by introducing the following equations,

$$\delta_1 = 1 - F(N_{\text{Obs}} - 1 | N_F) \tag{A9}$$

$$\delta_2 = F(N_{\text{Obs}} | N_F) \tag{A10}$$

where  $F(x|\mu)$  is the right-continuous Poisson cumulative distribution function with expectation  $\mu$  evaluated at  $x$  and  $N_F$  is the forecast number of events determined by the model.

**A4.2.  $S$  Test**

[73] Analytically, the expression for estimating the log likelihood  $L$  of observing  $\omega$  events at a given expectation  $\lambda$  for a model  $j$  is defined by the logarithm of the probability  $p(\omega|\lambda)$ , which is given by

$$L(\omega|\lambda) = \log p(\omega|\lambda) = -\lambda + \omega \log \lambda - \log \omega! \tag{A11}$$

[74] The joint log likelihood represents the sum of log likelihood values over all bins  $b_i$  and is given by the expression

$$L(\Omega|\Lambda) = \sum_{(i,j) \in R} (-\lambda(i,j) + \omega(i,j) \log(\lambda(i,j)) - \log(\omega(i,j)!)) \tag{A12}$$

[75] **Acknowledgments.** The authors would like to thank Jeanne Hardebeck and Fred Pollitz for constructive reviews and comments on an early version of the manuscript. We are grateful to the Associate Editor and two anonymous reviewers and Sandy Steacy for their detailed comments. Earthquake catalogs were made available through ANSS and Lawrence-Livermore websites (<http://www.ncedc.org/anss/catalog-search.html> and <http://www.ldeo.columbia.edu/~felixw/NCAeqDD/>, last accessed January 2013). For static stress change calculations we have used Coulomb 3.2 software (available at <http://earthquake.usgs.gov/research/modeling/coulomb/>, last accessed January 2013) and DLC software, provided by R. Simpson. Variable slip source models were available through SCRMOD site (<http://www.seismo.ethz.ch/static/srcmod/>, last accessed January 2013). The research presented in this paper has been supported by the Earthquake hazards Program of the USGS.

**References**

Andrews, D., D. Oppenheimer, and J. Lienkaemper (1993), The Mission link between the Hayward and Calaveras Faults, *J. Geophys. Res.*, *98*, 12,083–12,095.  
 Beroza, G. C. (1991), Near-source modeling of the Loma Prieta earthquake: Evidence for heterogeneous slip and implications for earthquake hazard, *Bull. Seismol. Soc. Am.*, *81*, 1603–1621.  
 Beroza, G. C., and P. Spudich (1988), Linearized inversion for fault rupture behavior: Application to the 1984, Morgan Hill, California, earthquake, *J. Geophys. Res.*, *93*, 6275–6296.

Beroza, G. C., and M. D. Zoback (1993), Mechanism diversity of the Loma Prieta aftershocks and the mechanics of mainshock-aftershock interaction, *Science*, *259*, 210–213.  
 Cocco, M., S. Hainzl, F. Catalli, B. Enescu, A. M. Lombardi, and J. Woessner (2010), Sensitivity study of forecasted aftershock seismicity based on Coulomb stress calculation and rate- and state-dependent frictional response, *J. Geophys. Res.*, *115*, B05307, doi:10.1029/2009JB006838.  
 Dieterich, J. (1994), A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.*, *99*, 2601–2618.  
 Dietz, L. D., and W. L. Ellsworth (1990), The October 17, 1989, Loma Prieta, California, earthquake and its aftershocks: Geometry of the sequence from high-resolution locations, *Geophys. Res. Lett.*, *17*, 1417–1420.  
 Dietz, L. D., and W. L. Ellsworth (1997), Aftershocks of the 1989 Loma Prieta earthquake and their tectonic implications, in *The Loma Prieta, California Earthquake of October 17, 1989—Aftershocks and Post-Seismic Effects*, Profess. Pap. 1550-D, edited by P. A. Reasenberg, pp. D5–D47, U.S. Geol. Surv., Menlo Park, Calif.  
 Enescu, B., J. Mori, M. Miyazawa, and Y. Kano (2009), Omori-Utsu law  $c$ -values associated with recent moderate earthquakes in Japan, *Bull. Seismol. Soc. Am.*, *99*, 884–891.  
 Felzer, K. (2008), Uniform California earthquake rupture forecast, Appendix I: Calculating California seismicity rates, *USGS Open File Report 2007-1437L*, 41 pp., Pasadena, Calif.  
 Field, E. H., et al. (2009), Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2), *Bull. Seismol. Soc. Am.*, *99*, 2053–2107, doi:10.1785/0120080049.  
 Geist, E. L., and D. J. Andrews (2000), Slip rates on San Francisco Bay area faults from anelastic deformation of the continental lithosphere, *J. Geophys. Res.*, *105*, 25,543–25,552.  
 Hainzl, S., B. Enescu, M. Cocco, J. Woessner, F. Catalli, R. Wang, and F. Roth (2009), Aftershock modeling based on uncertain stress calculations, *J. Geophys. Res.*, *114*, B05309, doi:10.1029/2008JB006011.  
 Hanks, T., and H. Kanamori (1979), A moment magnitude scale, *J. Geophys. Res.*, *84*, 2348–2350, doi:10.1029/JB084iB05p02348.  
 Hardebeck, J. L. (2013), Uniform California earthquake rupture forecast, Appendix S: Constraining ETAS parameters from the UCERF3 catalog and validating the ETAS model for  $M \geq 6.5$  earthquakes, 24 pp.  
 Hardebeck, J. L., and A. J. Michael (2004), Stress orientations at intermediate angles to the San Andreas Fault, California, *J. Geophys. Res.*, *109*, B11303, doi: 10.1029/2004JB003239.  
 Harris, R. A., and R. W. Simpson (1996), In the Shadow of 1857—The effect of the Great Ft. Tejon earthquake on subsequent earthquakes in Southern California, *Geophys. Res. Lett.*, *23*, 229–232.  
 Harte, D., and D. Vere-Jones (2005), The entropy score and its uses in earthquake forecasting, *Pure Appl. Geophys.*, *162*(6-7), 1229–1253.  
 Helmstetter, A. (2003), Is earthquake triggering driven by small earthquakes?, *Phys. Rev. Lett.*, *91*(5), 058,501, doi:10.1103/PhysRevLett.91.058501.  
 Helmstetter, A., and D. Sornette (2002), Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, *107*(B10), 2237, doi:10.1029/2001JB001580.  
 Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005), Importance of small earthquakes for stress transfer and earthquake triggering, *J. Geophys. Res.*, *110*, B05S08, doi:10.1029/2004JB003286.  
 Herd, D. G. (1979), Neotectonic framework of central coastal California and its implications to microzonation of the San Francisco Bay region, *U.S. Geol. Surv. Circ.*, *807*, 3–12.  
 Jordan, T. H., and L. M. Jones (2010), Operational earthquake forecasting: Some thoughts on why and how, *Seismol. Res. Lett.*, *81*, 571–574, doi:10.1785/gssrl.81.4.571.  
 Jordan, T. H., Y. T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka, and J. Zschau (2011), Operational earthquake forecasting. State of knowledge and guidelines for utilization, *Ann. Geophys.*, *54*(4), doi:10.4401/ag-5350.  
 Kagan, Y. Y., and L. Knopoff (1977), Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. Inter.*, *14*, 97–108.  
 Kilb, D., M. Ellis, J. Gombert, and S. Davis (1997), On the origin of diverse aftershock mechanisms 1989 Loma Prieta earthquake, *Geophys. J. Int.*, *128*, 557–570.  
 King, G. C. P., R. S. Stein, and J. Lin (1994), Static stress changes and the triggering of earthquakes, *Bull. Seismol. Soc. Am.*, *84*(3), 935–953.  
 Lachenbruch, A. H., and J. H. Sass (1980), Heat flow and energetics of the San Andreas Fault zone, *J. Geophys. Res.*, *85*, 6185–6222.  
 Lombardi, A. M., and W. Marzocchi (2010), The assumption of Poisson seismic-rate variability in CSEP/RELM experiments, *Bull. Seismol. Soc. Am.*, *100*, 2293–2300, doi:10.1785/0120100012.  
 Manaker, D., A. Michael, and R. Bürgmann (2005), Subsurface structure and kinematics of the Calaveras-Hayward fault stepover from three-dimensional Vp and seismicity, San Francisco Bay region, California, *Bull. Seismol. Soc. Am.*, *95*, 446–470, doi:10.1785/0120020202.



- Marsan, D. (2005), The role of small earthquakes in redistributing crustal elastic stress, *Geophys. J. Int.*, *163*, 141–151, doi:10.1111/j.1365-246X.2005.02700.x.
- Marzocchi, W., and A. M. Lombardi (2008), A double branching model for earthquake occurrence, *J. Geophys. Res.*, *113*, B08317, doi:10.1029/2007JB005472.
- Marzocchi, W., J. D. Zechar, and T. H. Jordan (2012), Bayesian forecast evaluation and ensemble earthquake forecasting, *Bull. Seismol. Soc. Am.*, *102*(6), 2574–2584, doi:10.1785/01201110327.
- Mount, V. S., and J. Suppe (1992), Present-day stress orientations adjacent to active strike-slip faults: California and Sumatra, *J. Geophys. Res.*, *97*, 11,995–12,013, doi:10.1029/92JB00130.
- Mueller, R. J., and M. J. S. Johnston (2000), Co-seismic and post-seismic deformation along the San Andreas Fault south of the October 18, 1989, Loma Prieta M7.1 Earthquake Rupture, *US Geological Survey Open File Report*, 00-146, 35 pp.
- Murray-Moraleda, J. R., and R. W. Simpson (2009), Geodetically inferred coseismic and postseismic slip due to the M5.4 31 October 2007 Alum Rock earthquake, *Bull. Seismol. Soc. Am.*, *99*(5), 2784–2800.
- Nanjo, K. Z., et al. (2012), Predictability study on the aftershock sequence following the 2011 Tohoku-Okii, Japan, earthquake: First results, *Geophys. J. Int.*, *191*(2), 653–658.
- Ogata, Y. (1988), Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. Stat. Assoc.*, *83*, 9–27.
- Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, *50*, 379–402.
- Ogata, Y., and J. Zhuang (2006), Space-time ETAS models and an improved extension, *Tectonophysics*, *413*, 13–23, doi:10.1016/j.tecto.2005.10.016.
- Oppenheimer, D. H. (1990), Aftershock slip behavior of the 1989 Loma Prieta, California earthquake, *Geophys. Res. Lett.*, *17*, 898–916.
- Parsons, T. (2002), Post-1906 stress recovery of the San Andreas Fault system from 3-D finite element analysis, *J. Geophys. Res.*, *107*(B8), 2162, doi:10.1029/2001JB001051.
- Parsons, T., R. S. Stein, R. W. Simpson, and P. A. Reasenberg (1999), Stress sensitivity of fault seismicity: A comparison between limited-offset oblique and major strike-slip faults, *J. Geophys. Res.*, *104*, 20,183–20,202.
- Parsons, T., Y. Ogata, J. Zhuang, and E. L. Geist (2012), Evaluation of static stress change forecasting with prospective and blind tests, *Geophys. J. Int.*, *188*, 1425–1440, doi:10.1111/j.1365-246X.2011.05343.x.
- Reasenberg, P. (1985), Second-order moment of central California seismicity, 1969–1982, *J. Geophys. Res.*, *90*(B7), 5479–5495, doi:10.1029/JB090iB07p05479.
- Reasenberg, P., and D. Oppenheimer (1985), FPFIT, FPLOT, and FPPAGE: Fortran computer programs for calculating and displaying earthquake fault-plane solutions, *Technical report 85-739*, 109 pp., U.S. Geol. Survey.
- Reasenberg, P. A., and R. W. Simpson (1992), Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake, *Science*, *255*, 1687–1690.
- Rhoades, D. A. (2007), Application of the EEPAS model to forecasting earthquakes of moderate magnitude in Southern California, *Seismol. Res. Lett.*, *78*(1), 110–115.
- Rhoades, D. A. (2013), Mixture models for improved earthquake forecasting with short-to-medium time horizons, *Bull. Seismol. Soc. Am.*, *103*(4), 2208–2215.
- Rhoades, D. A., D. Schorlemmer, M. C. Gerstenberger, A. Christophersen, J. D. Zechar, and M. Imoto (2011), Efficient testing of earthquake forecasting models, *Acta Geophys.*, *59*, 728–747, doi:10.2478/s11600-011-0013-5.
- Rice, J. R. (1992), Fault stress states, pore pressure distributions, and the weakness of the San Andreas Fault, in *Fault Mechanics and Transport Properties of Rocks: A Festschrift in Honor of W. F. Brace*, edited by B. Evans and T. Wong, pp. 475–503, Academic Press, San Diego, Calif.
- Saichev, A., and D. Sornette (2005), Vere-Jones' self-similar branching model, *Phys. Rev. E*, *72*, 056,122.
- Saichev, A., and D. Sornette (2006), Renormalization of branching models of triggered seismicity from total to observable seismicity, *Eur. Phys. J. B*, *51*, 443–459.
- Schorlemmer, D., M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007), Earthquake likelihood model testing, *Seismol. Res. Lett.*, *78*(1), 17–29, doi:10.1785/gssrl.78.1.17.
- Smith, B., and D. T. Sandwell (2003), Coulomb stress along the San Andreas Fault system, *J. Geophys. Res.*, *108*(B6), 2296, doi:10.1029/2002JB002136.
- Smith-Konter, B., and D. T. Sandwell (2009), Stress evolution of the San Andreas Fault System: Recurrence interval versus locking depth, *Geophys. Res. Lett.*, *36*, L13304, doi:10.1029/2009GL037235.
- Sornette, D., and M. J. Werner (2005a), Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Bath's law, and observed aftershock sequences, *J. Geophys. Res.*, *110*, B08304, doi:10.1029/2004JB003535.
- Sornette, D., and M. J. Werner (2005b), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res.*, *110*, B09303, doi:10.1029/2005JB003621.
- Steady, S., S. S. Nalbant, J. McCloskey, C. Nostro, O. Scotti, and D. Baumont (2005), Onto what planes should Coulomb stress perturbations be resolved?, *J. Geophys. Res.*, *110*, B05S15, doi:10.1029/2004JB003356.
- Steady, S., M. Gerstenberger, C. Williams, D. Rhoades, and A. Christophersen (2013), A new hybrid Coulomb/statistical model for forecasting earthquake rates, *Geophys. J. Int.*, doi:10.1093/gji/ggt404.
- Stein, R. S. (1999), The role of stress transfer in earthquake occurrence, *Nature*, *402*, 605–609.
- Stein, R. S., A. A. Barka, and J. H. Dieterich (1997), Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering, *Geophys. J. Int.*, *128*, 594–604.
- Toda, S., R. S. Stein, K. Richards-Dinger, and S. Bozkurt (2005), Forecasting the evolution of seismicity in Southern California: Animations built on earthquake stress transfer, *J. Geophys. Res.*, *110*, B05S16, doi:10.1029/2004JB003415.
- Townend, J. (2006), What do faults feel? Observational constraints on the stresses acting on seismogenic faults, in *Earthquakes: Radiated Energy and the Physics of Faulting*, AGU Monograph Series, vol. 170, edited by R. Abercrombie et al., pp. 313–327, AGU, Washington, D. C., USA.
- Utsu, T. (1961), A statistical study on the occurrence of aftershocks, *Geophys. Mag.*, *30*, 521–605.
- Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, *113*, B08311, doi:10.1029/2007JB005479.
- Wells, D. L., and K. J. Coppersmith (1994), New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.*, *84*, 974–1002.
- Werner, M. J. (2008), On the fluctuations of seismicity and uncertainties in earthquake catalogs: Implications and methods for hypothesis testing, PhD thesis, 308 pp., University of California Los Angeles.
- Werner, M. J., and D. Sornette (2008), Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. Geophys. Res.*, *113*, B08302, doi:10.1029/2007JB005427.
- Werner, M. J., J. D. Zechar, W. Marzocchi, S. Wiemer, and the CSEP-Italy Working Group (2010), Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, *Ann. Geophys.*, *53*(3), 11–30.
- Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011), High-resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.*, *101*, 1630–1648.
- Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, F. Catali, B. Enescu, M. Cocco, M. C. Gerstenberger, and S. Wiemer (2011), A retrospective comparative forecast test on the 1992 Landers sequence, *J. Geophys. Res.*, B05305, doi:10.1029/2010JB007846.
- Zechar, J. D., M. C. Gerstenberger, and D. A. Rhoades (2010), Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. Seismol. Soc. Am.*, *100*, 1184–1195, doi:10.1785/0120090192.
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, *97*, 369–380.