

Earthquake precursors: activation or quiescence?

John B. Rundle,^{1,2} James R. Holliday,³ Mark Yoder,³ Michael K. Sachs,³ Andrea Donnellan,⁴ Donald L. Turcotte,⁵ Kristy F. Tiampo,⁶ William Klein⁷ and Louise H. Kellogg⁵

¹Departments of Physics and Geology, University of California, Davis, CA, USA

²Santa Fe Institute, Santa Fe, NM, USA

³Department of Physics, University of California, Davis, CA, USA. E-mail: yoder@physics.ucdavis.edu

⁴Earth and Space Science Division, Jet Propulsion Laboratory, Pasadena, CA, USA

⁵Department of Geology, University of California, Davis, CA, USA

⁶Department of Earth Science, University of Western Ontario, Ontario, Canada

⁷Department of Physics, Boston University, Boston, MA, USA

Accepted 2011 June 28. Received 2011 June 6; in original form 2010 September 21

SUMMARY

We discuss the long-standing question of whether the probability for large earthquake occurrence (magnitudes $m > 6.0$) is highest during time periods of smaller event activation, or highest during time periods of smaller event quiescence. The physics of the activation model are based on an idea from the theory of nucleation, that a small magnitude earthquake has a finite probability of growing into a large earthquake. The physics of the quiescence model is based on the idea that the occurrence of smaller earthquakes (here considered as magnitudes $m > 3.5$) may be due to a mechanism such as critical slowing down, in which fluctuations in systems with long-range interactions tend to be suppressed prior to large nucleation events. To illuminate this question, we construct two end-member forecast models illustrating, respectively, activation and quiescence. The activation model assumes only that activation can occur, either via aftershock nucleation or triggering, but expresses no choice as to which mechanism is preferred. Both of these models are in fact a means of filtering the seismicity time-series to compute probabilities. Using 25 yr of data from the California–Nevada catalogue of earthquakes, we show that of the two models, activation and quiescence, the latter appears to be the better model, as judged by backtesting (by a slight but not significant margin). We then examine simulation data from a topologically realistic earthquake model for California seismicity, Virtual California. This model includes not only earthquakes produced from increases in stress on the fault system, but also background and off-fault seismicity produced by a BASS–ETAS driving mechanism. Applying the activation and quiescence forecast models to the simulated data, we come to the opposite conclusion. Here, the activation forecast model is preferred to the quiescence model, presumably due to the fact that the BASS component of the model is essentially a model for activated seismicity. These results lead to the (weak) conclusion that California seismicity may be characterized more by quiescence than by activation, and that BASS–ETAS models may not be robustly applicable to the real data.

Key words: Time series analysis; Persistence, memory, correlations, clustering; Probabilistic forecasting.

1 INTRODUCTION

Many years ago, Gardner & Knopoff (1974) asked the question: ‘Is the sequence of large earthquakes, with aftershocks removed, Poissonian?’. Their one-word abstract was ‘Yes’. They came to this answer by removing earthquakes from the southern California catalogue until the remaining event occurrences were consistent with Poisson statistics. Since that time, a number of models have been proposed that assume earthquakes are best characterized by ‘sta-

tionary’ or ‘homogeneous’ Poisson statistics with an aftershock component. In this case, precursory clustering or quiescence of background seismicity is considered to be a purely random occurrence of a system with no memory. Variations in small earthquake rate prior to a large earthquake are explained as a consequence of random fluctuations.

More generally, the question of whether the probability for earthquake occurrence is highest during periods of smaller event activation, or highest during periods of smaller event quiescence has

been an ongoing debate. Classic examples of this debate include papers by Mogi (1969) and Kanamori (1981). Other examples include Rundle & Kanamori (1987); Tiampo *et al.* (2002a, 2002b); Chen *et al.* (2005); Ogata (2005, 2006, 2007); Holliday *et al.* (2005, 2007); Yen *et al.* (2006); Huang (2006); Chen & Wu (2006); Huang (2008); Mignan & Di Giovambattista (2008); Shearer & Lin (2009); Kossobokov (2006); Bowman & Sammis (2004) and Bufe & Varnes (1993).

Most of the models in these papers have been primarily statistical in nature, although it is possible to place both activation and quiescence into the context of modern nucleation models (nucleation references such as Langer 1967; Gunton & Droz 1983; Gunton *et al.* 1983; Rundle 1989, 1993; Klein & Unger 1993; Rundle *et al.* 1997; Shcherbakov *et al.* 2005). In such a context, the physics of the activation model are based on an idea from the theory of nucleation—that a small earthquake has a finite probability of growing into a large earthquake, so that more small events imply a larger probability for occurrence of a large earthquake. A comment on this model has been stated as: ‘the greatest probability for a large earthquake is the moment after it occurs’ (E.H. Field, US Geological Survey, personal communication, 2010). Examples of this type of model are the ETAS (Ogata 2005), BASS (Holliday *et al.* 2007; Turcotte *et al.* 2007) and STEP (Gerstenberger *et al.* 2005) models, which are statistical models using the Omori and Gutenberg–Richter laws.

An important caveat to the preceding discussion is that the analysis in this paper cannot distinguish activation resulting from the physics of nucleation from activation resulting from triggering of activity on neighbouring faults (see also the discussions in Hardebeck *et al.* 2008 and Greenhough *et al.* 2009). Here, we only seek to analyse whether such activation of small earthquakes can be seen as a precursor to large earthquakes. As part of this analysis, we also analyse a data set generated by a Virtual California (VC) numerical simulation (Rundle *et al.* 2005; Yikilmaz 2010) that includes a BASS–ETAS model to simulate the smaller, subgrid-scale earthquakes. This model explicitly includes the physics of activation.

By contrast, the physics of the quiescence model is based on the idea that the occurrence of smaller earthquakes may be due to a mechanism such as critical slowing down (Ma 1974; Klein & Unger 1983). Here, fluctuations in systems with long-range interactions such as elastic systems tend to be suppressed prior to large nucleation events. An example of such a model is the seismic gap model. Other models include both activation and quiescence, such as the Pattern Informatics model (Tiampo *et al.* 2002a, 2002b; Holliday *et al.* 2005), which looks only for deviations from the average (activation or quiescence), giving equal weight to both.

We note that we use the term ‘forecast’ as opposed to ‘prediction’. We take the former to mean the computation of event probabilities. The latter is taken to mean the specification of a unique outcome, for example, the statement that ‘an earthquake will occur on Tuesday of next week at 9:30 p.m. at location X, and will be of magnitude Y’. The difference can be compared to the statement that ‘the next coin flip will have a 50 per cent chance of being a head’, compared to the statement that ‘the next coin flip will be a head’.

To summarize our results, we construct simple models for large earthquake probabilities based on simple models of quiescence and activation in the rate of small earthquake activity. We ‘backtest’ these models using earthquake data from the ANSS catalogue in California during the years 1985–present to determine which model is more consistent with the data. We find that neither the activation nor the quiescence model provides significant forecast skill, from the standpoint of a reliability/attributes (R/A) test, or a receiver operating characteristic (ROC) test. Our conclusion is that the rate of

small earthquake occurrence is not well correlated with the impending occurrence of a larger earthquake.

2 MODELS

We use the general formalism of reliability engineering and hazard analysis (Ebeling 1997; NIST 2010). If time between failures of a system is t , the probability density function for time between failures is $P(t)$ and the cumulative distribution is $P(t)$,

$$P(t) = \int_0^t p(t) dt. \quad (1)$$

Then, the instantaneous hazard rate or failure rate $h(t)$ is defined as

$$h(t) = \frac{p(t)}{1 - P(t)} = \frac{p(t)}{S(t)}. \quad (2)$$

Here, $S(t)$ is defined as the survivor function. The failure rate $h(t)$ is the conditional rate of failure, given that the system has survived to time t . Since

$$h(t) = -\frac{d \log\{S(t)\}}{dt}, \quad (3)$$

we have the identity

$$P(t) = 1 - \exp\left\{-\int_0^t h(t) dt\right\}. \quad (4)$$

The cumulative hazard function $H(t)$ is defined as

$$H(t) = \int_0^t h(t) dt. \quad (5)$$

Note that for a (homogeneous) Poisson process (NIST 2010) having rate parameter λ (number of events per time interval), the hazard rate $h(t) = \lambda$, $H(t) = \lambda t$ and the $P(t) = 1 - \exp\{-\lambda t\}$.

The conditional probability $P(t|\Delta t)$ that a failure will occur in a time Δt in the future, given that it has not failed before t , is given by

$$P(t|\Delta t) = \frac{P(t + \Delta t) - P(t)}{1 - P(t)} \quad (6)$$

or

$$P(t|\Delta t) = 1 - \exp\{-\Delta H(t, \Delta t)\}, \quad (7)$$

where $\Delta H(t, \Delta t) = H(t + \Delta t) - H(t)$.

The basic idea of our models is to use small earthquakes, those with magnitudes m_C at the level of catalogue completeness, to compute expected properties of large earthquakes having magnitude m . The steps in this method are the following:

(1) We first compute the expected (average) number N_C of ‘small’ earthquakes larger than the cut-off magnitude m_C corresponding to 1 ‘large’ earthquake having magnitude $> m$.

$$N_C = 10^{b(m-m_C)}. \quad (8)$$

We note that eq. (8) is just the statement, for example, that if $b = 1$, then there must be 1000 $m > 3$ earthquakes for every $m > 6$ earthquake.

In eq. (8), we assume that b has been computed by fitting seismicity data over a time interval $\{t_0, t\}$, t being the present time and t_0 is a time at least 30 yr into the past. With an averaging time interval of at least 30 yr, b will be a relatively slowly varying function of time over intervals of a year or less. Thus, N_C will also vary relatively slowly with time intervals of a year or less.

(2) Compute a (non-declustered, non-homogeneous) time-varying ‘Poisson’ rate of small earthquakes $v_C(t)$. This can be done using a variety of methods. Here, we use a double-averaging method in which we first compute the number of small earthquakes $\Delta n_C(t)$ over a time interval $\{t - T_C, t\}$, where T_C is an (assumed) averaging time interval. The rate $v_C(t)$ is then defined by

$$v_C(t) = \frac{1}{t - t_0} \int_{t_0}^t \frac{\Delta n_C(t)}{T_C} dt. \quad (9)$$

Typically, we use $T_C = 5$ yr which is approximately five times the interoccurrence time of $m > 6$ earthquakes in the California region. Parameters in the model such as T_C were backtested and validated using the procedures described later, implying that the modelled probabilities are optimal. We also point out that the double averaging described in (9) also contributes to the slowly varying b -value and the stability and reliability of the results. We also point out that we did not decluster the catalogue [usually justified by the Gardner-Knoppoff (1974) result] since we explicitly do not assume that the process is a homogenous Poisson process.

(3) Next we compute the expected $n_E(t)$ and observed $n_O(t)$ numbers of small earthquakes in a moving time window of length T_W , and compute their ratio $R(t) = n_O(t)/n_E(t)$.

If $R(t) > 1$, then there are more small earthquakes observed than were expected during the previous time interval T_W . We take this condition to correspond to ‘activation’. If $R(t) < 1$, then there were fewer small earthquakes observed than were expected during the previous time interval T_W . We take this condition to correspond to ‘quiescence’.

Computation of $n_E(t)$ and $n_O(t)$ is described later.

The time window we select can be termed the ‘Poisson window’:

$$T_W = \frac{N_C}{2v_C}. \quad (10)$$

The ratio N_C to v_C corresponds to the average time interval between large earthquakes. The factor of 2 in the denominator corresponds to the introduction of a Nyquist sampling frequency, in the sense that we are ‘sampling’ the small earthquakes at a frequency $1/T_W$ that is twice the average ‘Nyquist frequency’ (recurrence frequency) at which the large earthquakes occur. We find that the definition of T_W yields relatively robust outcomes, although the results do depend on the particular definition to some degree. This is a point that we plan to investigate in more detail in the future.

(4) Using the definition (8), the expected number of small earthquakes during the time interval T_W is then $n_E(t) = N_C/2$.

To compute the observed number of small earthquakes $n_O(t)$, we note that the simplest method would be to pass a moving window of length T_W and amplitude 1 over the time-series of number of small earthquakes versus time, that is, use a boxcar window function.

However, such a window will cause known problems, particularly at the trailing edge (at time $t - T_W$), where the departure of large clusters of events (in time) from the window will cause large fluctuations in the ratio $R(t)$.

For a window function, it is more desirable to use a cosine or other taper on the trailing edge. For the leading edge at time t , to include the triggering of larger events that is sometimes observed during heightened (‘foreshock’) activity, we retain the use of a sharp (step function) edge. Thus, we use a window function

$$F(t) = \left[\sin\left(\frac{\tau - t + T_W/2}{T_W}\right)\pi + 1 \right], \quad (11)$$

which is valid on $\tau \in \{t - T_W, t\}$. This function is used to define the observed number of earthquakes $n_O(t)$. It is normalized so that

the area under $F(t)$ is the same as the area under the boxcar window function.

(5) With these equations, we construct two non-homogeneous Poisson forecast models. The first is the ‘activation’ model that assumes that a large earthquake is more likely to occur during a period of heightened anomalous activity of small earthquakes.

$$P_A(t|\Delta t) = 1 - \exp\{-\Delta H_A(t, \Delta t)\}. \quad (12)$$

Here, $\Delta H_A(t, \Delta t)$ is the cumulative conditional hazard rate function for activation, equal to

$$\Delta H_A(t, \Delta t) = \frac{R(t)f v_C \Delta t}{N_C}. \quad (13)$$

The factor f is discussed later.

The second model is the ‘quiescence’ model that assumes that a large earthquake is more likely during a period of diminished anomalous activity of small earthquakes

$$P_Q(t|\Delta t) = 1 - \exp\{-\Delta H_Q(t, \Delta t)\}, \quad (14)$$

where $\Delta H_Q(t, \Delta t)$ is the cumulative conditional hazard rated function for quiescence

$$\Delta H_Q(t, \Delta t) = \frac{f v_C \Delta t}{N_C R(t)}. \quad (15)$$

Note that the ratio v_C/N_C is the Poisson rate for large earthquakes extrapolated from the rate of small earthquakes.

In both the activation and quiescence models, we have included a factor f that will be determined by optimizing the forecast using standard verification tests (‘backtesting’ the forecasts). The product $f v_C = v_O$ can then be considered to be the appropriate (and optimal) Poisson rate, determined by optimizing the forecasts. This is a process of data assimilation using the verification tests in a retrospective testing mode.

Note that the factor $R(t)$ in (13) and (15) plays opposing roles. In the activation model, $\Delta H_A(t, \Delta t)$ is proportional to the activation ratio $R(t)$, whereas in the quiescence model, $\Delta H_Q(t, \Delta t)$ is inversely proportional to $R(t)$.

3 DATA ASSIMILATION WITH BACKTESTING AND VERIFICATION

The testing schema that we use for backtesting consists of methods that have been developed in the validation and verification community (Joliffe & Stephenson 2003; Casati *et al.* 2008). For our purposes, these tests are of two types: R/A diagrams (Murphy 1973; Hsu & Murphy 1986; Murphy 1988) and ROC tests (Green & Swets 1966; Joliffe & Stephenson 2003; Kharin & Zwiers 2003). We consider each in turn. Note that the R/A test is conditioned on the forecasts (i.e. given the forecast probability Y , what actually happened?). However, the ROC test is conditioned on the observations (i.e. given that X occurred, what was the forecast?).

3.1 Attributes/reliability (A/R)

A/R diagrams are based on a partition of the Briar Score (Winkler & Murphy 1968; Atger 2004; Mason 2004).

$$QR = 1/K \sum_{k=1}^K (r_k - d_k)^2. \quad (16)$$

Here, K represents the number of forecasts made, $k = 1, \dots, K$, r_k is the probability computed on the k th occasion, $d_k = 1$ if an event

occurs within the forecast window and $d_k = 0$ otherwise. For our purposes, a forecast is made on every time step from t_0 to t , in time steps of about $\delta t = 3.6$ d, so $K = \frac{t-t_0}{\delta t}$.

Furthermore, Hsu & Murphy (1986) describe a decomposition of QR which leads to three terms.

$$\begin{aligned} QR = \bar{d}(1-\bar{d}) + (1/K) \sum_{t=1}^T K_t(r_t - \bar{d}_t)^2 \\ -(1/K) \sum_{t=1}^T K_t(\bar{d}_t - \bar{d})^2. \end{aligned} \quad (17)$$

Here, we have divided the probability density up into T subsamples, where the subsample t consists of the K_t forecasts for which $r_k = r_t$, and $K = \sum_t K_t$. Note that \bar{d}_t is the relative frequency of occurrence of the event when $r_k = r_t$ and

$$\bar{d} = (1/K) \sum_t K_t \bar{d}_t. \quad (18)$$

In eq. (17) above, the first term is interpreted as the quadratic score for a static forecast having sample frequency \bar{d} over the K instances (Hsu & Murphy 1986) and is generally denoted as $QR^* = \bar{d}(1-\bar{d})$. The second term is the average squared difference between forecast probability r_k and relative frequency, thus it is a measure of the ‘reliability’ of the forecast. If $r_t = \bar{d}_t$ then the forecast is perfectly reliable, and smaller values of this term represent better reliability. The third term is regarded as a measure of attribute ‘resolution’ (Murphy & Daan 1985). It measures the difference between the subsample \bar{d}_t and the overall \bar{d} , thus the importance of subsample t . No resolution is indicated when $\bar{d} = \bar{d}_t$ and larger values of the third term indicate better resolution. Resolution indicates a forecaster’s ability to divide the sample into subsamples for which the forecast probabilities differ from the static probability.

Following Hsu & Murphy (1986), we define the following indicators of forecast quality:

$$\text{Reliability} : (1/K) \sum_{t=1}^T K_t(r_t - \bar{d}_t)^2, \quad (19)$$

$$\text{Sample Skill} : \frac{(QR^* - QR)}{QR^*}, \quad (20)$$

$$\text{Resolution} : (1/K) \sum_{t=1}^T K_t(\bar{d}_t - \bar{d})^2. \quad (21)$$

3.2 ROC and area skill

ROC methods were developed in a variety of fields beginning in the 1950s, an example being Green & Swets (1966). In the context of signal detection theory, these methods were further developed by Mason (1982) and Murphy & Winkler (1987). A succinct summary is given by Kharin & Zwiers (2003). In these methods, a threshold D is applied to a probabilistic forecast, transforming the forecast into a binary, or ‘yes/no’ forecast. For values $D \geq r_k$, the event is considered to be ‘forecast = yes’, whereas for $D < r_k$, the event is considered to be ‘forecast = no’. During the forecast period, it is determined O whether the event was actually observed ‘observed = yes’ or not ‘observed = no’.

Put another way, the forecasts are rank ordered from highest probability value to lowest probability value, and the threshold D decreases downward progressively from highest to lowest value. As the threshold decreases, it is then determined whether an observation occurs at a forecast value above or below the threshold.

A contingency table is then constructed for each value of the threshold D in which the various elements of the table are the number of forecasts that have the appropriate characteristics:

	Observed = Yes	Observed = No
Forecast = Yes	a	b
Forecast = No	c	d

The success or ‘hit’ rate is defined to be: $H = a/(a+c)$. The false alarm rate is defined to be: $F = b/(b+d)$. The total number of forecasts is $n = a+b+c+d$.

The ROC diagram is then constructed by plotting H against F as the threshold value R decreases. Both H and F vary between the limits [0,1]. A ‘good’ forecast has $H(F) > F$. A random forecast is characterized by the condition $H = F$ that is represented on the ROC diagram by a diagonal line connecting the point (0,0) at the lower left corner to the point (1,1) at the upper right corner.

The ‘Area Skill Score’ (Kharin & Zwiers 2003) is the area under the curve $H(F)$, thus it takes on values from 0 to 1. Here, we modify this definition slightly to be the area under the curve $H(F)-F$, thus it takes on values from -0.5 to 0.5. Values <0 indicate a better forecast of event non-occurrence. Values >0 indicate a better forecast of event occurrence.

3.3 Estimation of uncertainty

There are several ways to estimate uncertainty in the computation of reliability, skill and resolution. One of these is via Monte Carlo methods (Bevington & Robinson 2006), in which random models are generated and evaluated, with standard deviations being computed from deviations of model parameters from the optimal value. By random models, we mean models in which, for example, the parameter f in (13) and (15) is assigned a value based on a random number generator, and the performance of the suites of model is used to define the envelope of uncertainty.

A second method of estimating uncertainty is with bootstrap methods, in which large earthquake times are sampled with replacement (Efron & Tibshirani 1993). Estimates of reliability or other parameters are computed for suites of models having the optimal parameters, and the envelope of uncertainty is computed. Here, we implement this method by resampling the large earthquake times 200 distinct times, and recomputing the verification measures as described earlier. The 200-member ensemble of values of these measures then defines the statistical sample that allows us to estimate sample uncertainties.

Using the bootstrap method, we compute both the mean and standard deviation of the statistical measures of reliability, sample skill, etc. for the 200 samples. For a measure of uncertainty, we choose the larger of: (1) the computed standard deviation or (2) the difference between the computed mean from the 200 samples, and the computed value for the actual data.

4 DATA AND RESULTS

The region of interest for our study is the area of California and Nevada between latitudes 31.5° and 42° N, and between longitudes -126° and -113° W. We consider the time interval from 1985 January 1 to present (2011 January 26). We use seismicity data from the ANSS catalogue, and assume a catalogue completeness level of magnitude $m_C \geq 3.5$. For data analysis and plotting purposes, we

use a time step equal to 0.005 yr, about 43.83 hr. Larger earthquakes closer together in time than this value will not appear independently in the time-series plots, although they are treated as separate events in the verification analyses. The analysis seeks to forecast the large earthquakes having magnitudes $m > 6.0$ during the time period of interest using the small earthquakes having $m > 3.5$ during the time period of interest. For the simulation data, the recurrence time (~ 12 months) of large events is somewhat longer than the recurrence time (~ 11 months) of the real data. For this reason, we use a forecast time of 5.5 months for the simulated data, in place of the forecast time of 4 months used for the real data.

Using eq. (12) together with (14), we compute and show 4-month forecasts for the activation model ($P_A(t|\Delta t)$, Fig. 1a) and the quiescence model ($P_Q(t|\Delta t)$, Fig. 1b). It should be noted that in producing Figs 1(a) and (b), it is assumed that the forecast probability is a coherent property of the entire region, rather than of an individual location. This assumption follows from the observations of a

regional-scale correlation length in seismically active areas (Jaume 1999, 2000; Zoller *et al.* 2001).

In Figs 1(a) and (b), the time-series of probability values is shown by the red line. The vertical dashed blue lines indicate the times and probabilities of the major earthquakes having magnitude $m \geq 6$. At the bottom of each of the time-series is a record of the regional activity of small earthquakes. The largest values of activity are associated with occurrence of the $m \geq 6$ earthquakes in the top figure. Note that the parameter f for each model is chosen by a reliability analysis as described later. For the activation model, we find $f = 0.9$, and for the quiescence model, we find $f = 0.6$.

It can be seen in Fig. 1(a) that, with a few exceptions, the probability $P_A(t|\Delta t)$ time-series computed for the activation model typically trends to lower values as the next major earthquake approaches. After the earthquake occurs, the probability ‘increases’ sharply, thus the probability is generally highest ‘just after’ the earthquake occurs. The primary exception to this pattern are the large earthquakes that

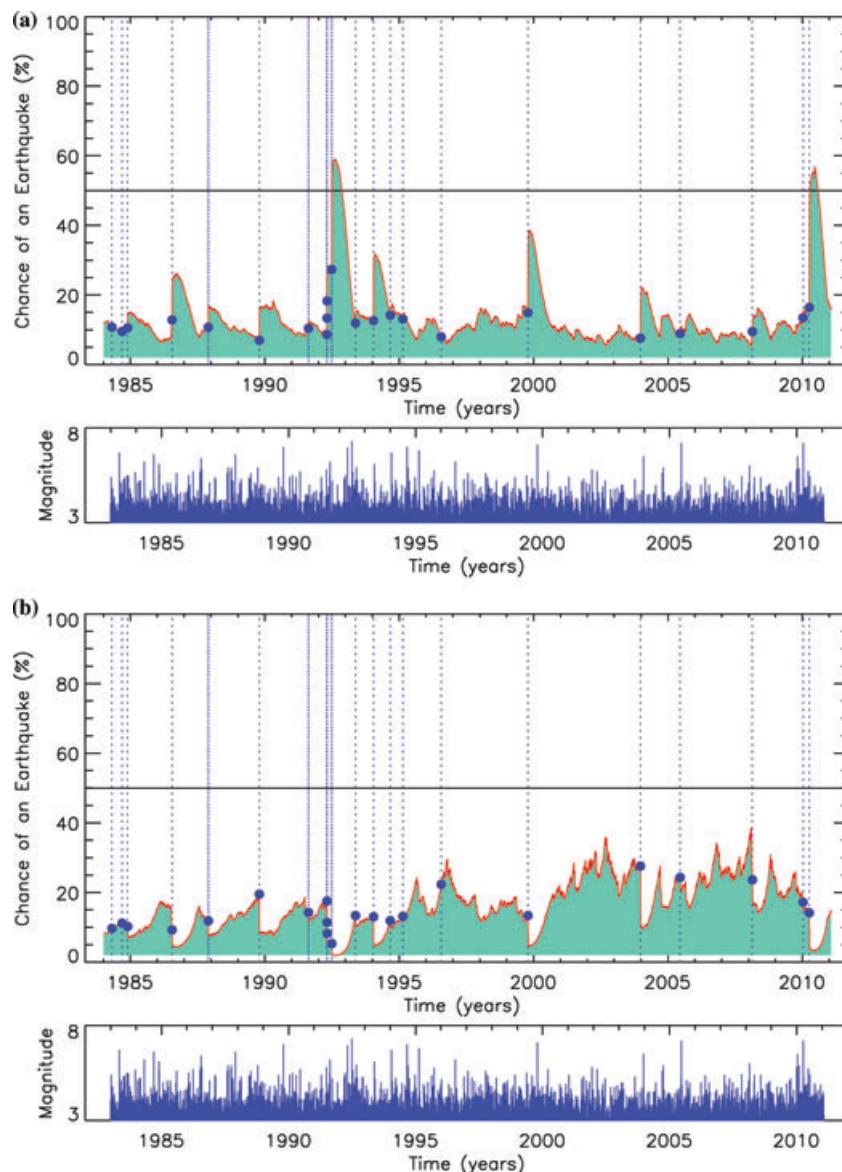


Figure 1. Time-series of 4-month forecasts for California and Nevada region, from 1985–present. Top panel in each figure shows the 4-month forecast probability (per cent) as a function of time (red curve). Bottom panel shows the magnitude of the events as a function of time. In each top panel, the blue dots and blue dotted vertical lines denote the occurrence of the large $M \geq 6$ earthquakes. (a) Activation model. (b) Quiescence model.

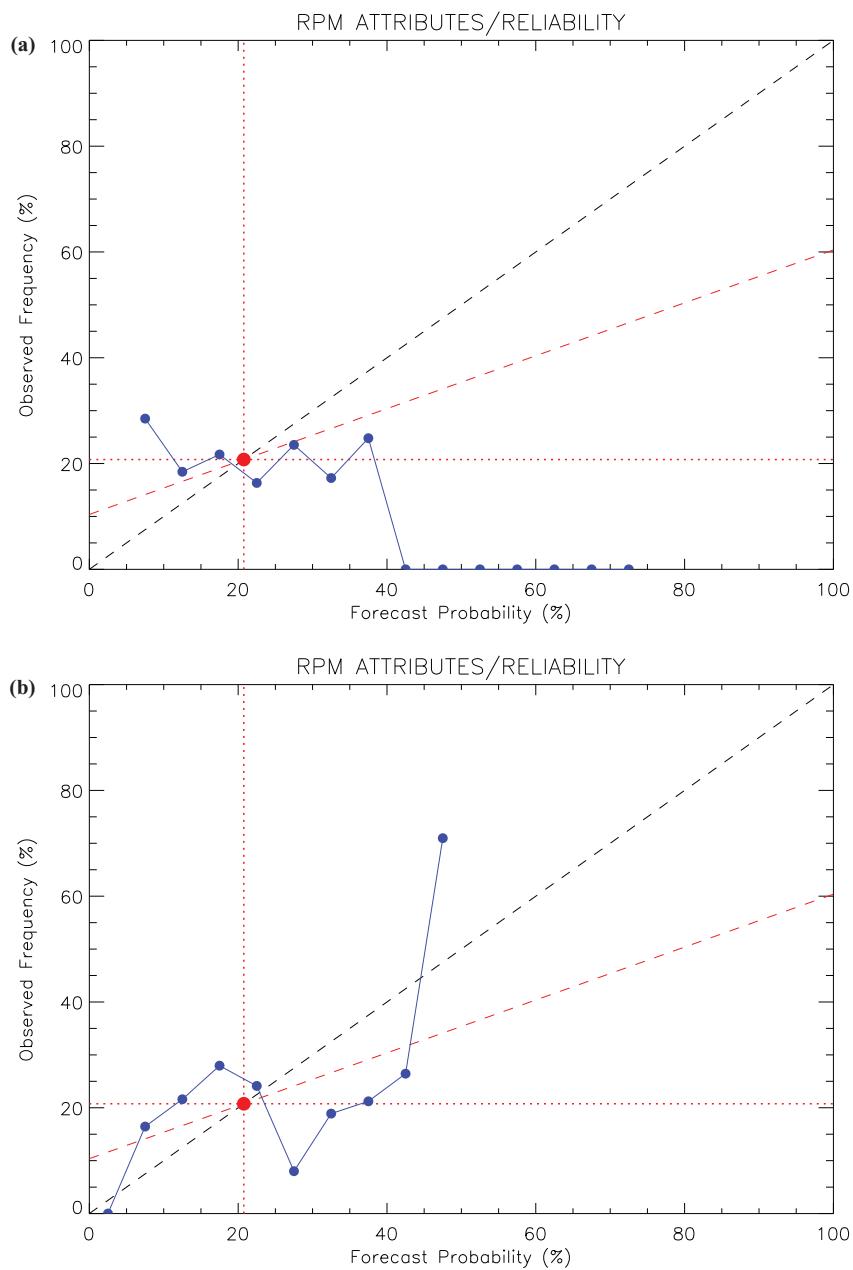


Figure 2. Attributes/Reliability diagrams for aftershock and quiescence models. The 45° dashed diagonal line ($\text{slope} = 1$) is the line of best reliability (observed frequency = forecast probability). The dashed line of slope $\frac{1}{2}$ is the line of best sample skill. The dashed horizontal line represents zero resolution. (a) Activation model (b) Quiescence model.

are closely clustered in time in 1992, the Joshua Tree ($m6.1$)–Cape Mendocino ($m7.2$)–Landers ($m7.3$)–Big Bear ($m6.5$) events. Again this figure illustrates the idea that the activation model supposes that large earthquakes should occur during intervals of greatest activity. Note also that there are a number of examples of ‘false alarms’, where the probability time-series increases suddenly in a fluctuation, but there is no corresponding $m \geq 6$ earthquakes. A case in point occurs at the beginning of 1998.

By contrast, Fig. 1(b) shows the probability and activity time-series for the quiescence model $P_Q(t|\Delta t)$. Unlike the activation model, 4-month probabilities computed from the quiescence model ‘decrease’ sharply just after a large earthquake (but not to zero, due to the possibility of a large aftershock). Like the activation model,

the quiescence model also produces false alarms, for example, at year 2002 May, and 2002 July. In this model, the sequence of events in 1992 occurs at progressively decreasing probability. Also, the 1999 October 16 Hector Mine earthquake occurs at a low value of probability, indicating the appearance of increasing activity just prior to the event.

To select the optimal values of parameter f in eqs (13) and (15), we perform a grid search and select the value leading to the smallest value for the reliability measure eq. (19). For these values of f , the associated A/R diagrams are shown in Figs 2(a) and (b) (Hsu & Murphy 1986). The ‘climatology’ point (\bar{d}, \bar{d}) is shown as a large dot, at the intersection of slope 1, the horizontal and vertical reference lines and the line of slope 0.5, which is the ‘no skill’

line. The horizontal dashed line is the ‘no resolution’ line, and the vertical dashed line is what we may call the ‘no sensitivity’ line where $r_t = \bar{d}$.

It can be seen from Fig. 2(a) that the reliability values for the activation model do not follow the diagonal line closely, and that there are large values of forecast probability, greater than 35 per cent, at which the observed frequency is zero. This is an example of ‘overforecasting’. In fact, for values of forecast probability less than about 35 per cent, the reliability values appear to follow the horizontal ‘no resolution’ line that passes through the climatology point. By contrast, the quiescence model in Fig. 2(b) indicates that the reliability values do a better job of following the diagonal line, and that the overforecasting problem is not in evidence.

The density function of probability measure corresponding to the A/R diagram of Figs 2(a) and (b) are shown in Figs 3(a) and (b). These figures show the relative number of forecasts r_t in each of 20 aggregated, equally spaced bins. This type of analysis indicates the relative importance of each forecast bin to the overall evaluation of the forecast. From these figures, it can be seen that for the activation model, most of the forecast probability is in the bins from 5 to 20 per cent. For the quiescence model, the forecast probability is much more evenly spaced between 0 and 45 per cent (better ‘dynamic range’).

Figs 4(a) and (b) show the corresponding ROC functions. The activation ROC curve in Fig. 4(a) follows the diagonal line in general, but is for the most part below the diagonal, meaning that it is

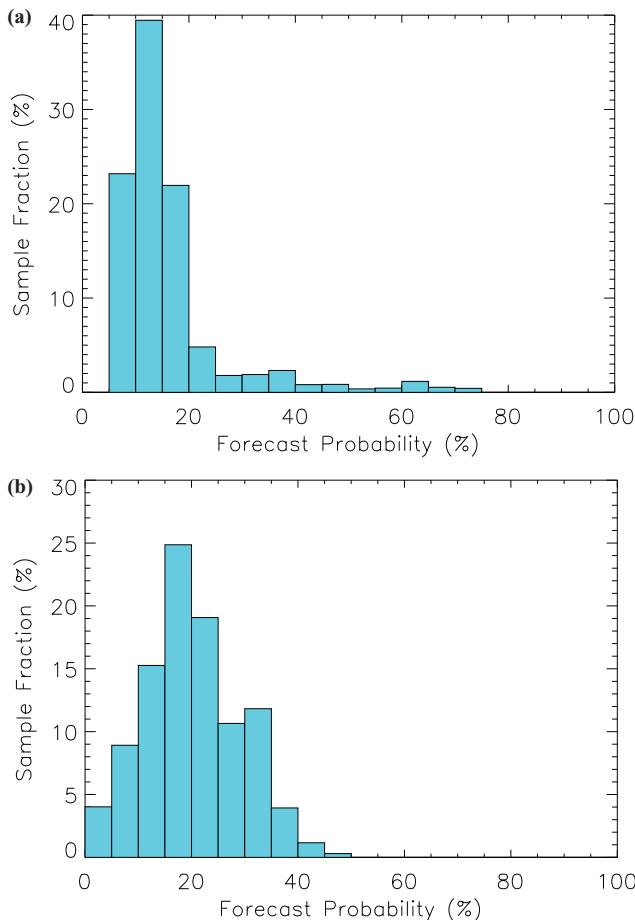


Figure 3. Density plots for forecast probability measure for aftershock and quiescence models. Wider range of values implies optimal dynamic range. (a) Activation model. (b) Quiescence model.

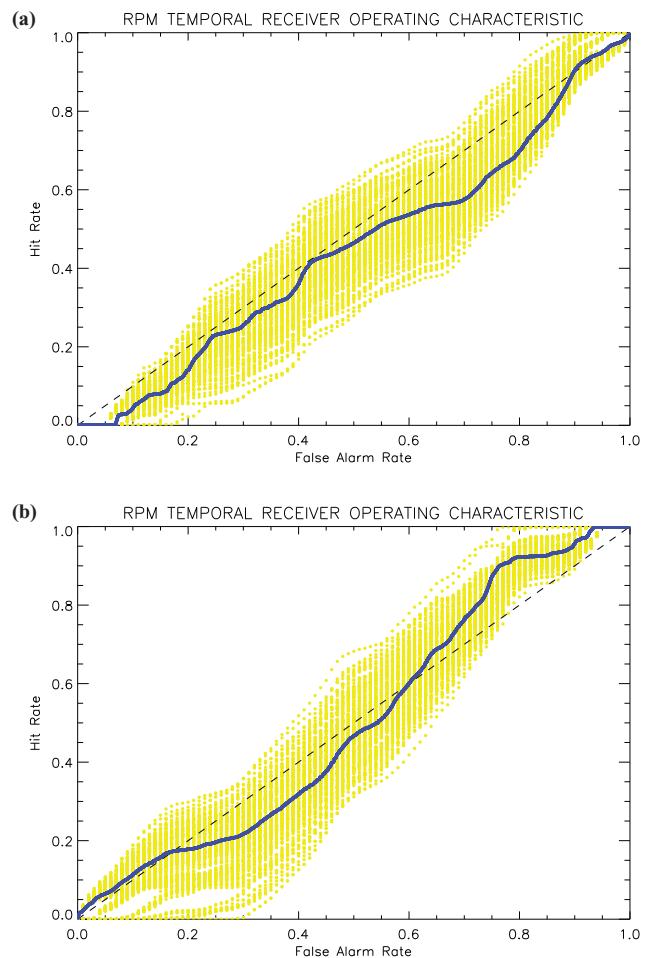


Figure 4. Receiver Operating Characteristic diagrams for aftershock and quiescence models applied to real data. Diagonal line represents no skill (random guessing). An ROC curve that lies above the diagonal represents a positive skill forecast or a forecast of occurrence. An ROC curve below the diagonal represents a negative skill forecast or a forecast of non-occurrence. Yellow bands are ROC diagrams for the 200 bootstrap tests. (a) Activation model. (b) Quiescence model.

better at forecasting non-occurrence of large earthquakes. The area skill score, the area between the ROC curve and the diagonal line, is negative, and has the value -0.046 .

The quiescence ROC in Fig. 4(b) also follows the diagonal, but spends about half the interval above the diagonal line, indicating a somewhat better forecast for occurrence than the activation ROC in Fig. 4(a). The area skill score for this model is -0.007 , still negative, indicating overall a slightly better forecast of non-occurrence.

The results of these analyses, together with estimates of uncertainties, are summarized in Table 1. Here, we give values of reliability, resolution, sample skill and area skill, together with estimates of uncertainty. These uncertainties were computed as described earlier by a bootstrap procedure. We computed 200 members of the bootstrap ensemble, sampling the times of the $m \geq 6$ earthquakes with replacement. We then computed the means and standard deviations as described earlier.

The results are shown in Table 1. Comparing values for the quiescence model to those of the activation model, taking account of the preferred ranges of parameters, it can be seen that the quiescence model performs better than the activation model, although both are within the uncertainty.

Table 1. Results from the verification computations for reliability, resolution, sample skill, and area skill, as described in the text for California Data from 1985 to present (January 26, 2011). Uncertainties are obtained from the bootstrap analysis, and represent the larger of (1) the absolute value of the difference between observed value and bootstrap mean or (2) the computed bootstrap standard deviation.

	Reliability	Resolution	Sample skill	Area skill
Range (Good to bad)	0 to ∞	1 to 0	1 to $-\infty$	+0.5 to -0.5
Activation model	0.028 ± 0.005	0.004 ± 0.001	-0.139 ± 0.02	-0.050 ± 0.052
Quiescence model	0.013 ± 0.006	0.005 ± 0.002	-0.039 ± 0.048	-0.006 ± 0.047

In addition to the analysis of the California data, we also investigated the question of whether ETAS or BASS type models are capable of producing either accelerated (activation) or decelerated (quiescence) seismic energy release. This can be regarded as a somewhat more sophisticated null hypothesis than assuming a stationary Poisson process. Another question we seek to answer

is whether the time-series of Fig. 1 are so dominated by overlapping aftershock sequences that the result is the appearance of a statistically stationary process.

To answer these questions, we employed a state-of-the-art earthquake simulator, VC (Rundle *et al.* 2005; Yikilmaz 2010) to generate a series of simulated earthquakes to which we applied the

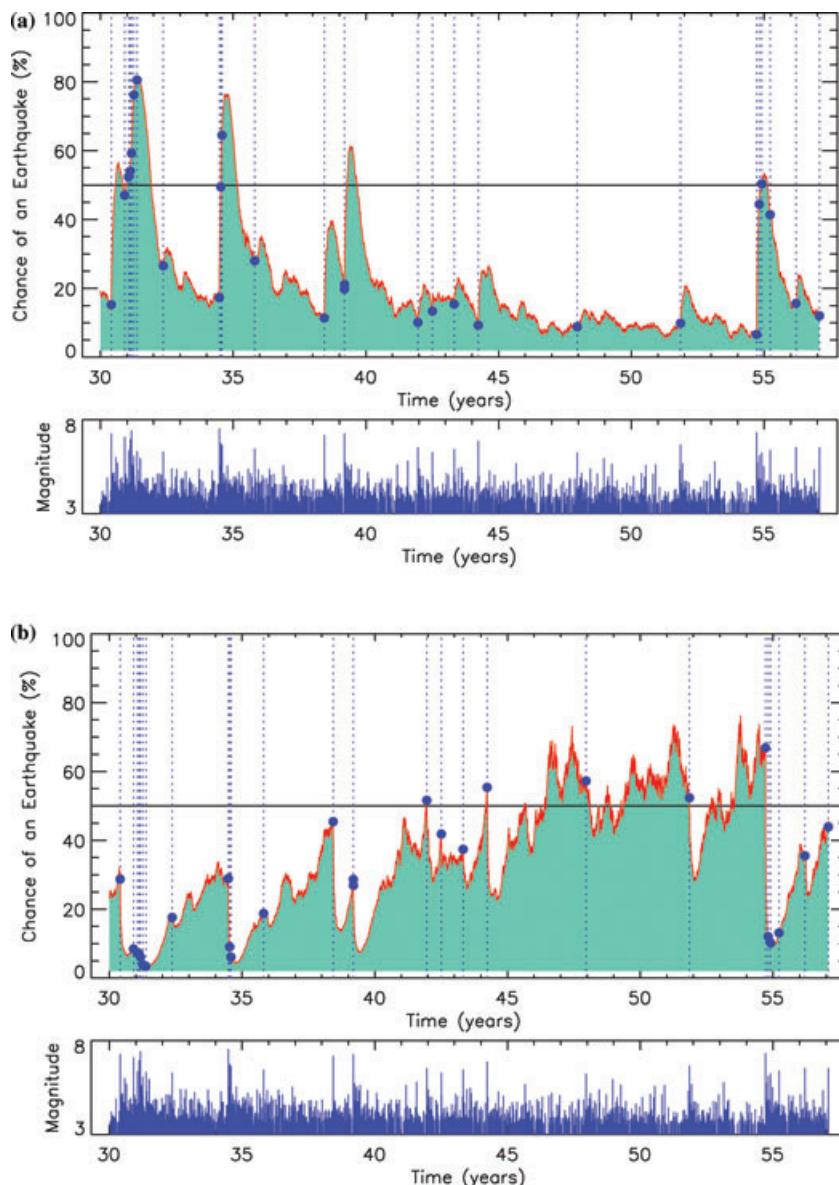


Figure 5. Time-series of 5.5-month forecasts for simulation data from Virtual California + BASS model + Poisson background. Top panel in each figure shows the 5.5-month forecast probability (per cent) as a function of time (red curve). Bottom panel shows the magnitude of the events as a function of time. In each top panel, the blue dots and blue dotted vertical lines denote the occurrence of the large $M \geq 6$ earthquakes. (a) Activation model. (b) Quiescence forecast model.

models and verification regimen as described in the foregoing. The significance of this test using the VC model is that (1) there is no question of completeness, since we can observe all earthquakes produced and (2) the dynamics of the model are fully known, since it is a simulation.

The latest version of VC includes a BASS–ETAS model to simulate earthquakes down to magnitude $m = 3.5$ that are driven by two processes. The first process is subgrid-scale earthquakes driven by the cellular automaton VC simulation, which by itself produces on-fault earthquakes down to about $m = 5.5$. The second process is off-fault earthquakes, for which there is a background Poisson rate, producing off-fault earthquakes from magnitudes $m \sim 6.0$ down to the lower cut-off of $m = 3.5$. We consider the full California fault model similar to that described by the WGCEP (Working Group on

California Earthquake Probabilities; <http://wgcep.org>, last accessed 2010 June) and the QuakeSim group (<http://quakesim.jpl.nasa.gov>, last accessed 2010 June). Details of these models are described in papers currently in preparation.

We note that the simulation model has more $m > 6$ earthquakes during the ~ 25 yr shown than does the actual data (27 for the simulation versus 22 for the data). It was not possible to find an exact statistical match for the California–Nevada data for the same time period, number of earthquakes, b -value and other variables at the present time, given the large number of VC and ETAS–BASS parameters involved. To do so will require the development of more sophisticated data assimilation procedures. Moreover, we used a 5.5-month forecast period for the simulation rather than the 4-month forecast period for the California–Nevada data, a value that was

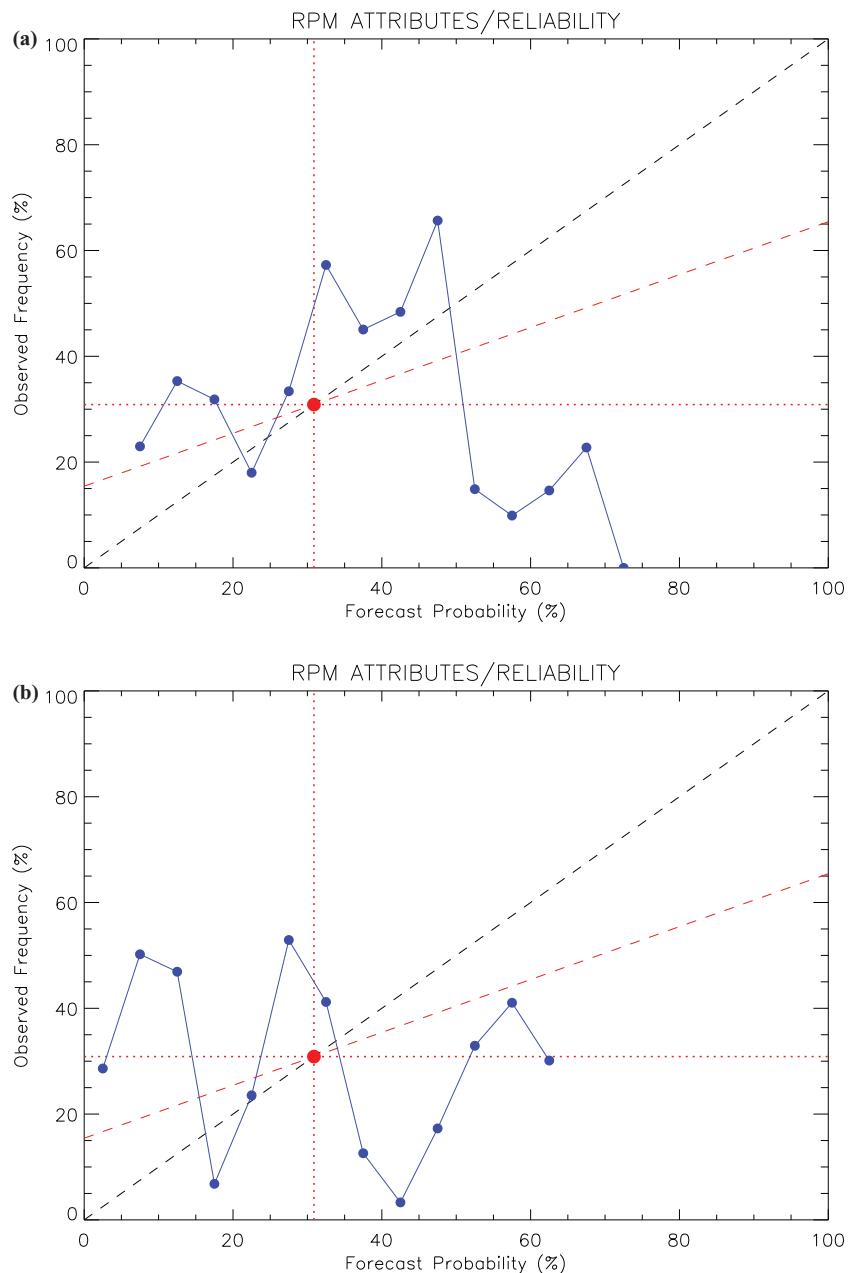


Figure 6. Attributes/Reliability diagrams for aftershock and quiescence models 5.5-month forecasts for simulated data from Virtual California + BASS model + Poisson background. The 45° dashed diagonal line (slope = 1) is the line of best reliability (observed frequency = forecast probability). The dashed line of slope $\frac{1}{2}$ is the line of best sample skill. The dashed horizontal line represents zero resolution. (a) Activation model. (b) Quiescence model.

dictated by the desire to optimize, as much as possible, the forecast statistics. Both of these factors lead to the higher probabilities apparent in Fig. 5 as compared to Fig. 1. However, the major point remains, that the simulation forecast based on activation has better reliability and skill than the simulation forecast based on quiescence, a conclusion that is opposite to that reached with the real data.

We applied the verification tests as described earlier and found the results shown in Figs 5–8 and summarized in Table 2. It is interesting to observe that for the VC simulation, the activation forecast model is now preferred on the basis of minimum reliability (smaller reliability error is better). This is not surprising, since the VC model explicitly includes a BASS activation-type process. Although the resolution for the activation forecast model is not as good for the VC data as for the real California data, the area skill and sample skill are both better.

The fact that the activation forecast is preferred for an activation-based (VC) model seems to suggest that the verification tests are revealing interesting information about the earthquake process, and may be revealing similar quality of information about the real California data set. Since the real data apparently favour a quiescence-type forecast model, our results appear to indicate that BASS–ETAS models may not be as widely applicable as currently thought.

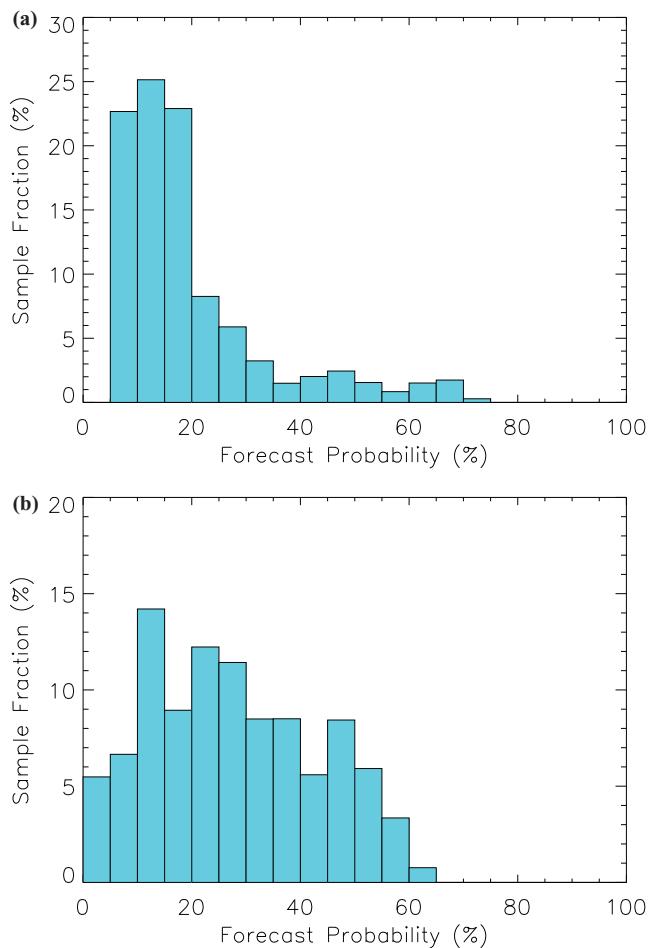


Figure 7. Density plots for forecast probability measure for aftershock and quiescence models for 5.5-month forecasts for simulation data from Virtual California + BASS model + Poisson background. Wider range of values implies optimal dynamic range. (a) Activation model. (b) Quiescence model.

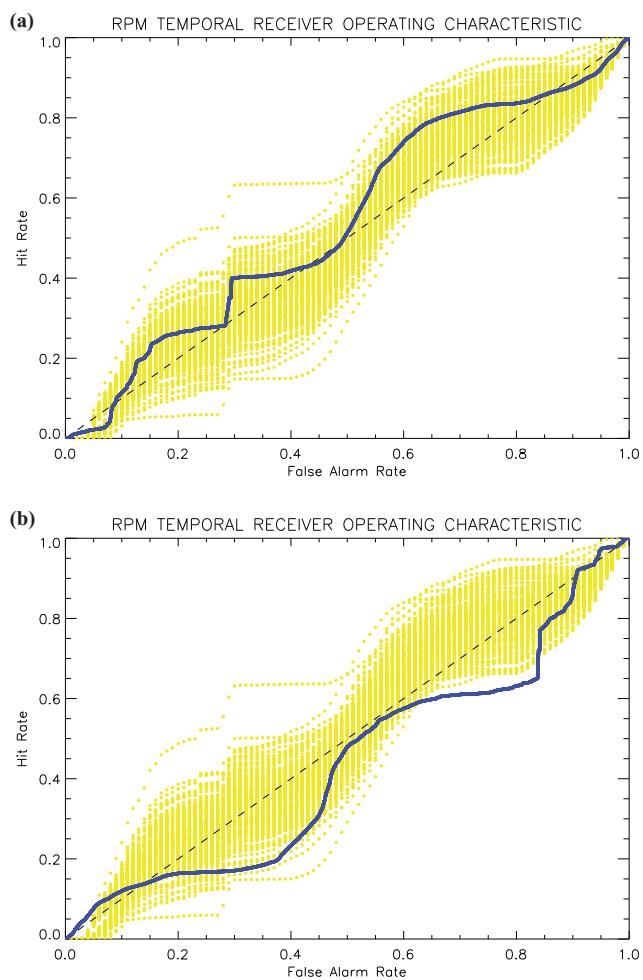


Figure 8. Receiver Operating Characteristic diagrams for aftershock and quiescence models for 5.5-month forecasts from Virtual California + BASS model + Poisson background. Diagonal line represents no skill (random guessing). An ROC curve that lies above the diagonal represents a positive skill forecast or a forecast of occurrence. An ROC curve below the diagonal represents a negative skill forecast or a forecast of non-occurrence. Yellow bands are ROC diagrams for the 200 bootstrap tests. (a) Activation model. (b) Quiescence model.

We may also conclude that the basic assumption of earthquakes as a homogeneous Poisson process is questionable. Recall from the introduction that Gardner & Knopoff (1974) arrived at the conclusion that earthquakes are Poisson by ‘removing’ the statistical non-Poisson component. Recall also that for a process to be Poisson, the events described must be independent and statistically identical (e.g. radioactive decay, arrival of cars at a restaurant and arrival of internet packets at a server). The Gutenberg–Richter distribution describes the frequency of events larger than a magnitude m . One may question whether an earthquake of $m = 6$ is statistically identical to a nearby event having $m = 7.5$. In addition, recent research indicates that main shock earthquakes in a region are characterized by correlations and a correlation length and are thus not independent (Rundle *et al.* 1997; Jaume & Sykes 1999; Hainzl *et al.* 2000; Zoller *et al.* 2001; Bowman & Sammis 2004). Thus, the basic assumptions associated with earthquakes as Poisson processes seem to be called into question.

Table 2. Results from the verification computations for the reliability, resolution, sample skill, and area skill for the 25 yr from the Virtual California + BASS + Poisson background simulations. Definitions as in Table 1.

	Reliability	Resolution	Sample skill	Area skill
Range (Good to bad)	0 to ∞	1 to 0	1 to $-\infty$	+0.5 to -0.5
Activation model	0.042 ± 0.021	0.021 ± 0.013	-0.097 ± 0.034	-0.016 ± 0.038
Quiescence model	0.067 ± 0.016	0.021 ± 0.006	-0.215 ± 0.066	-0.044 ± 0.039

5 SUMMARY

We have considered two types of forecast models for large earthquakes. Both use the rate of occurrence of small earthquakes to forecast the occurrence of large earthquakes. The first is an activation model, in which a high recent rate of activity of small earthquakes is assumed to imply a high probability of a future large earthquake. The second is a quiescence model, in which a low recent rate of activity of small earthquakes is assumed to imply a high probability of a future large earthquake. We applied these models to data sets from California, as well as to simulation data produced by a VC model.

For the activation model, the occurrence of a large earthquake is typically immediately followed by a large increase in probability for a future large earthquake ('the highest probability for the large earthquake is seen immediately after it happens'). For the quiescence model, the occurrence of a large earthquake typically leads to a sudden drop in probability for a future large earthquake ('the lowest probability for the large earthquake is seen immediately after it happens'). Models of the activation type include the ETAS (Ogata 2005), BASS (Holliday *et al.* 2007) and STEP (Gerstenberger *et al.* 2005) models. While quiescence as a precursor has been proposed (Wyss *et al.* 1996), there are apparently no quiescence models in the literature comparable to the quiescence model described here.

We compared the two models using standard verification tests described in the literature. We also computed errors using the standard bootstrap methods. Using 25 yr of catalogue seismicity data for California and Nevada, we find that there is a small preference for the quiescence model, although the difference in computed verification values is typically near the noise level. For the simulation data, we found essentially the opposite conclusion, raising questions about the applicability of the BASS-ETAS activation model approach. Research using similar models is underway to further clarify these conclusions.

ACKNOWLEDGMENTS

Research by JBR and JRH was supported by NASA grant NNX08AF69G to UC Davis. Research by MY and MKS was supported by a contract from JPL to UC Davis #1291967.

REFERENCES

- Atger, F., 2004. Relative impact of model quality and ensemble deficiencies on the performance of ensemble based probabilistic forecasts evaluated through the Brier score, *Nonlinear Process. Geophys.*, **11**, 399–409.
- Bevington, P.R. & Robinson, D.K., 2006. *Data Reduction and Error Analysis for the Physical Sciences*, McGraw Hill, New York, NY.
- Bowman, D.D. & Sammis, C.G., 2004. Intermittent criticality and the Gutenberg-Richter distribution, *Pure appl. Geophys.*, **161**, 1945–1956.
- Bufe, C.G. & Varnes, D.J., 1993. Predictive modeling of the seismic cycle in the San Francisco Bay region, *J. geophys. Res.*, **98**, 9871–9883.
- Casati, B. *et al.*, 2008. Forecast verification: current status and future directions, *Meteorol. Appl.*, **15**, 3–18.
- Chen, C.C. & Wu, Y.X., 2006. An improved region-time-length algorithm applied to the 1999 Chi-Chi, Taiwan earthquake, *Geophys. J. Int.*, **166**, 1144–1147.
- Chen, C.C., Rundle, J.B., Holliday, J.R., Nanjo, K.Z., Turcotte, D.L., Li, S.-C. & Tiampo, K.F., 2005. The 1999 Chi-Chi, Taiwan, earthquake as a typical example of seismic activation and quiescence, *Geophys. Res. Lett.*, **32**, L22315, doi:10.1029/2005GL023991.
- Ebeling, C.E., 1997. *An Introduction to Reliability and Maintainability Engineering*, McGraw-Hill, Boston.
- Efron, B. & Tibshirani, R., 1993. *An Introduction to the Bootstrap*, Chapman and Hall/CRC, Boca Raton, FL.
- Gardner, J.K. & Knopoff, L., 1974. Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, *Bull. seism. Soc. Am.*, **64**, 5.
- Gerstenberger, M.C., Weimer, S., Jones, L.M. & Reasenberg, P., 2005. Real time forecasts of tomorrow's earthquakes in California, *Nature*, **435**, 328–331.
- Green, D.M. & Swets, J.M., 1966. *Signal Detection Theory and Psychophysics*, John Wiley & Sons, New York, NY.
- Greenhough, J., Bell, A. & Main, I.G., 2009. Comment on 'Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes' by Arnaud Mignan and Rita Di Giovambattista, *Geophys. Res. Lett.*, **36**, L17303, doi:10.1029/2009GL039846.
- Gunton, J.D. & Droz, M., 1983. *Introduction to the Theory of Metastable and Unstable States*, Springer Lecture Notes in Physics 183, Springer-Verlag, Berlin.
- Gunton, J.D., San Miguel, M. & Sahni, P.S., 1983. The dynamics of first order phase transitions, in *Phase Transitions and Critical Phenomena*, Vol. 8, pp. 269–467, eds Domb, C. & Lebowitz, J., Academic Press, London.
- Hainzl, S., Zoller, G., Kurths, J. & Zschau, J., 2000. Seismic quiescence as an indicator for large earthquakes in a system of self-organized criticality, *Geophys. Res. Lett.*, **27**, 597–600.
- Hardebeck, J.L., Felzer, K.R. & Michael, A.J., 2008. Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant, *J. geophys. Res.*, **113**, B08310, doi:10.1029/2007JB005410.
- Holliday, J.R., Nanjo, K.Z., Tiampo, K.F., Rundle, J.B. & Turcotte, D.L., 2005. Earthquake forecasting and its verification, *Nonlinear Process. Geophys.*, **12**, 965–977.
- Holliday, J.R., Turcotte, D.L. & Rundle, J.B., 2007. BASS, an alternative to ETAS, *Geophys. Res. Lett.*, **34**, L12303, doi:10.1029/2007GL029696.
- Hsu, W.-R. & Murphy, A.H., 1986. The attributes diagram: a geometrical framework for assessing the quality of probability forecasts, *Int. J. Forecast.*, **2**, 285–293.
- Huang, Q., 2006. Search for reliable precursors: a case study of the seismic quiescence of the 2000 western Tottori prefecture earthquake, *J. geophys. Res.*, **111**, doi:10.1029/2005JB003982.
- Huang, Q., 2008. Seismicity changes prior to the M_s 8.0 Wenchuan earthquake in Sichuan, China, *Geophys. Res. Lett.*, **35**, L23308, doi:10.1029/2008GL036270.
- Jaume, S.C., 2000. Changes in earthquake size-frequency distributions underlying accelerating seismic moment/energy release, *Geophys. Monogr.*, **120**, 199–210.
- Jaume, S.C. & Sykes, L.R., 1999. Evolving towards a critical point: a review of accelerating seismic moment/energy release prior to large and great earthquakes, *Pure appl. Geophys.*, **155**, 279–305.

- Jolliffe, I.T. & Stephenson, D.B., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley & Sons, Chichester.
- Kanamori, H., 1981. The nature of seismicity patterns before large earthquakes, in *Earthquake Prediction—An International Review*, Series 4, pp. 1–19, ed. Ewing, M., AGU, Washington, DC.
- Kharin, V.V. & Zwiers, F.W., 2003. On the ROC score of probability forecasts, *J. Clim.*, **16**, 4145–4150.
- Klein, W. & Unger, C., 1983. Pseudospinodals, spinodals and nucleation, *Phys. Rev. B*, **28**, 445–448.
- Kossobokov, V.G., 2006. Testing earthquake prediction methods: the west Pacific short-term forecast of earthquakes with magnitude $M_w > 5.8$, *Tectonophysics*, **413**, 25–31.
- Langer, J.S., 1967. Theory of the condensation point, *Ann. Phys.*, **41**, 108–157.
- Ma, S.-K., 1974. *Modern Theory of Critical Phenomena*, Benjamin-Cummings, Reading, MA.
- Mason, I.B., 1982. A model for assessments of weather forecasts, *Aust. Meteorol. Mag.*, **30**, 291–303.
- Mason, S.J., 2004. On using ‘climatology’ as a reference strategy in the Brier and ranked probability skill scores. *Mon. Weather Rev.*, **132**, 1891–1895, doi:10.1175/1520-0493(2004)132<1891:OUCAAR>2.0.CO;2.
- Mignan, A. & Di Giovambattista, R., 2008. Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes, *Geophys. Res. Lett.*, **35**, L15306, doi:10.1029/2008GL035024.
- Mogi, K., 1969. Some features of recent seismic activity in and near Japan (2), Activity before and after great earthquakes, *Bull. Earthq. Res. Inst. Univ. Tokyo*, **47**, 395–417.
- Murphy, A.H., 1973. A new vector partition of the probability score, *J. Appl. Meteorol.*, **12**, 595–600.
- Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Weather Rev.*, **116**, 2417–2424.
- Murphy, A.H. & Daan, H., 1985. Forecast evaluation, in *Probability, Statistics and Decision Making in the Atmospheric Sciences*, eds Murphy, A.H. & Katz, R.W., Westview Press, Boulder, CO.
- Murphy, A.H. & Winkler, R.L., 1987. A general framework for forecast verification, *Mon. Weather Rev.*, **115**, 1330–1338.
- NIST/SEMATECH e-Handbook of Statistical Methods, 2010. Available at: <http://www.itl.nist.gov/div898/handbook/> (last accessed 2010 June).
- Ogata, Y., 2005. Synchronous seismicity changes in and around the northern Japan preceding the 2003 Tokachi-oki earthquake of M8.0, *J. geophys. Res.*, **110**, B08305, doi:10.1029/2004JB003323.
- Ogata, Y., 2006. Monitoring of anomaly in the aftershock sequence of the 2005 earthquake of M7.0 off coast of the western Fukuoka, Japan, by the ETAS model, *Geophys. Res. Lett.*, **33**, L01303, doi:10.1029/2005GL024405.
- Ogata, Y., 2007. Seismicity and geodetic anomalies in a wide area preceding the Niigata-Ken-Chuetsu earthquake of 23 October 2004, central Japan, *J. geophys. Res.*, **112**, B10301, doi:10.1029/2006JB004697.
- Rundle, J.B., 1989. A physical model for earthquakes. 3. Thermodynamical approach and its relation to nonclassical theories of nucleation, *J. geophys. Res.*, **94**, 2839–2855.
- Rundle, J.B., 1993. Magnitude-frequency relations for earthquakes using a statistical-mechanical approach, *J. geophys. Res.*, **98**, 21 943–21 949.
- Rundle, J.B. & Kanamori, H., 1987. Application of an inhomogeneous stress (patch) model to complex subduction zone earthquakes: a discrete interaction matrix approach, *J. geophys. Res.*, **92**, 2606–2616.
- Rundle, J.B., Gross, S., Klein, W. & Turcotte, D.L., 1997. The statistical mechanics of earthquakes, *Tectonophysics*, **277**, 147–164.
- Rundle, J.B. et al., 2005. A simulation-based approach to forecasting the next great San Francisco earthquake, *Proc. Natl. Acad. Sci.*, **102**, 15363–15367, doi:10.1073/pnas.0507528102.
- Shcherbakov, R., Yakovlev, G., Turcotte, D.L. & Rundle, J.B., 2005. Model for the distribution of aftershock interoccurrence times, *Phys. Rev. Lett.*, **95**, 218501, doi:10.1103/PhysRevLett.95.218501.
- Shearer, P.M. & Lin, G., 2009. Evidence for Mogi doughnut behavior in seismicity preceding small earthquakes in southern California, *J. geophys. Res.*, **114**, B01318, doi:10.1029/2008JB005982.
- Tiampo, K.F., Rundle, J.B., McGinnis, S. & Klein, W., 2002a. Pattern dynamics and forecast methods in seismically active regions, *Pageoph*, **159**, 2429–2467.
- Tiampo, K.F., Rundle, J.B., McGinnis, S., Gross, S. & Klein, W., 2002b. Mean field threshold systems and phase dynamics: an application to earthquake fault systems, *Europhys. Lett.*, **60**, 481–487.
- Turcotte, D.L., Holliday, J.R. & Rundle, J.B., 2007. BASS, an alternative to ETAS, *Geophys. Res. Lett.*, **34**, L12303, doi:10.1029/2007GL029696.
- Winkler, R.L. & Murphy, A.H., 1968. ‘Good’ probability assessors, *J. appl. Meteorol.*, **7**, 751–758.
- Wyss, M., Shimazaki, K. & Urabe, T., 1996. Quantitative mapping of a precursory seismic quiescence to the Izu-Oshima 1990(M6.5) earthquake, Japan, *Geophys. J. Int.*, **127**, 735–743.
- Yen, J.-Y., Chen, K.-S., Chang, C.-P. & Ng, S.-M., 2006. Deformation and ‘deformation quiescence’ prior to the Chi-Chi earthquake evidenced by DInSAR and groundwater records during 1995–2002 in Central Taiwan, *Earth, planets Space*, **58**, 805–813.
- Yikilmaz, M.B., 2010. Studies of fault interactions and regional seismicity using numerical simulations, *PhD thesis*, University of California, Davis, CA.
- Zoller G., Hainzl S. & Kurths J., 2001. Observation of growing correlation length as an indicator for critical point behavior prior to large earthquakes, *J. geophys. Res.*, **106**, 2167–2175.