LETTERS

Real-time forecasts of tomorrow's earthquakes in California

Matthew C. Gerstenberger[1], Stefan Wiemer[2], Lucile M. Jones[1] & Paul A. Reasenberg[3]

Despite a lack of reliable deterministic earthquake precursors, seismologists have significant predictive information about earthquake activity from an increasingly accurate understanding of the clustering properties of earthquakes[1–4]. In the past 15 years, time-dependent earthquake probabilities based on a generic short-term clustering model have been made publicly available in near-real time during major earthquake sequences. These forecasts describe the probability and number of events that are, on average, likely to occur following a mainshock of a given magnitude, but are not tailored to the particular sequence at hand and contain no information about the likely locations of the aftershocks. Our model builds upon the basic principles of this generic forecast model in two ways: it recasts the forecast in terms of the probability of strong ground shaking, and it combines an existing time-independent earthquake occurrence model based on fault data and historical earthquakes[5] with increasingly complex models describing the local time-dependent earthquake clustering[1,2]. The result is a time-dependent map showing the probability of strong shaking anywhere in California within the next 24 hours. The seismic hazard modelling approach we describe provides a better understanding of time-dependent earthquake hazard, and increases its usefulness for the public, emergency planners and the media.

We calculate the time-dependent seismic hazard by combining stochastic models derived from the Gutenberg–Richter relationship[6] and the modified Omori law[7]: starting from a background model of the long-term, spatially varying Poisson hazard[5], we add clustering models of varying complexity depending on the amount of currently available aftershock data. Most of the time and over most of California, a recent earthquake large enough to generate any appreciable increase in hazard will not have occurred and we then assume that the background model adequately represents the hazard. After an event of local magnitude $M_l = 3$ or larger, the additional hazard associated with possible aftershocks[2] is added to the background hazard until this time-dependent contribution decays below the background level. We do not limit an aftershock to a magnitude smaller than its preceding mainshock[8], so that a foreshock is simply an event whose aftershock happens to be bigger than itself; this permits the use of the same statistical description for foreshock and aftershock hazard.

Our clustering model is a composite of up to three models of increasing complexity: a generic-clustering model derived from the original Reasenberg and Jones model[1,2], a sequence-specific model, and a spatially heterogeneous model. In the generic clustering model, the rate at time t of aftershocks with magnitude $M \geq M_c$ is given by:

$$\lambda(t) = 10^{a' + b(M_m - M)}/(t + c)^p$$

where a', b, c and p are constants and M_m is the mainshock magnitude. If the ongoing earthquake sequence produces a sufficient number of aftershocks, a sequence-specific model is calculated in addition to the generic model; this model uses the a posteriori parameter values ($\hat{a}'$, $\hat{b}$ and $\hat{p}$) estimated for the sequence. Concurrent with the sequence-specific model, a third, spatially heterogeneous[9,10] model is calculated, whose parameters (a'_i, b_i and p_i) are calculated independently at the nodes of a 5-km grid covering the region, with parameter estimates for each node based on the seismicity within 15 km of it. However, for any given node, it is not always possible to make parameter estimates owing to the need for a minimum number of events to obtain a robust estimate of the parameters. See the online Supplementary Methods section for a full treatment of the calculations.

We do not limit an aftershock to occur at the same location as its mainshock. In the first two models, the mainshock is assumed initially to be a point source. This is modified to be a finite source, if and when the geometry of the source is known. The total rate of aftershocks is determined by the model and is distributed across an area that extends one-half of a fault length from the source. The spatial density of aftershocks is assumed to be proportional to $1/r^2$, where r is the horizontal distance from the mainshock's source. In the spatially heterogeneous model, the spatial variations in the rate are calculated directly from the distribution of aftershocks that have already occurred.

We calculate the expected rate of $4 \leq M \leq 8$ aftershocks at each node and for each model and combine the rates using Akaike weights, based on the corrected Akaike Information Criterion (AICc)[11]. The weights are based on the model's likelihood score, the amount of data and the number of free parameters that must be estimated, thus ensuring a gradual transition from the generic-clustering to the sequence-specific and spatially heterogeneous model as sufficient data warranting a more complex model become available. Because the AICc prefers fewer free parameters, the spatially heterogeneous model will be strongly weighted only in active areas.

The hazard of earthquake shaking at any site depends on the expected rate of earthquake sources nearby, their magnitudes and distances from the site, propagation effects (primarily geometric spreading and seismic attenuation) and local soil conditions at the site. We use an empirical attenuation function[12] to estimate peak ground acceleration (PGA) from a possible aftershock. We currently ignore local site effects and convert the PGA to an equivalent Modified Mercalli Intensity (MMI)[13]. The seismic hazard at a location is expressed as the probability that ground shaking exceeding MMI VI will occur there in the next 24 hours. We assume an isotropic distribution of shaking from any future earthquake as aftershocks and foreshocks do not have to have the same focal mechanism as their mainshock.

A snapshot of the probability of exceeding MMI VI in California on 28 July 2004 is shown in Fig. 1c. This is a composite of the time-independent background from geological information and

[1]US Geological Survey, 525 S. Wilson Ave, Pasadena, California 91106, USA. [2]ETH Zürich, Institute of Geophysics, 8093 Zürich, Switzerland. [3]US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA.

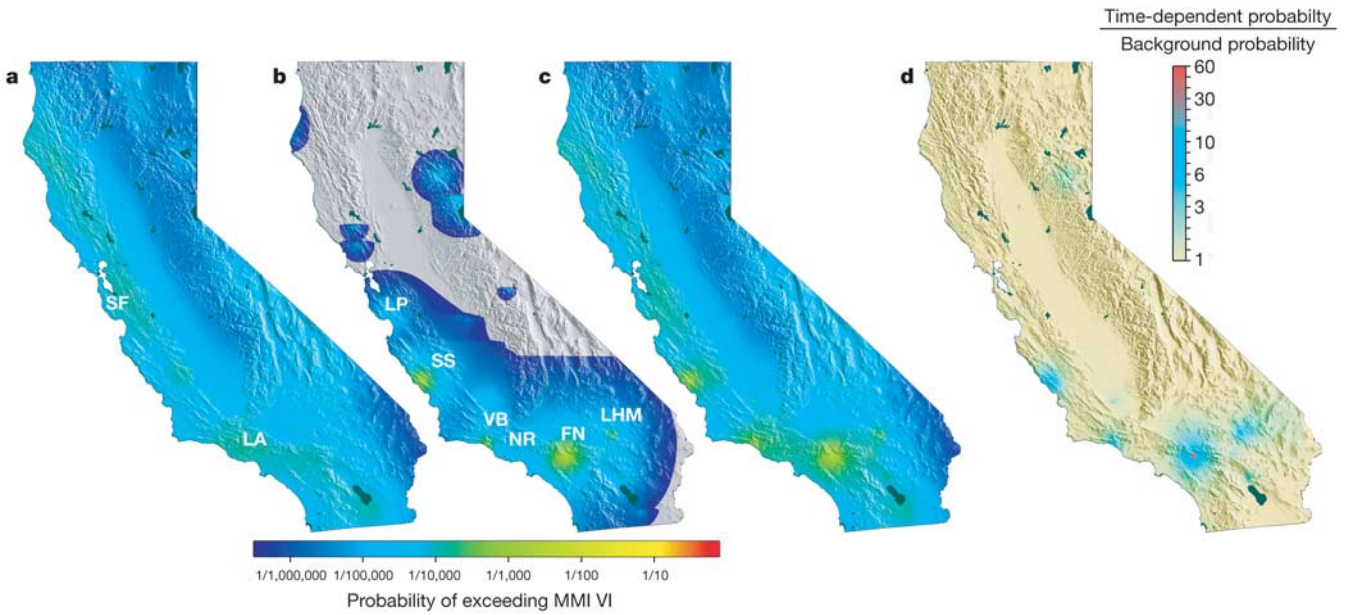

Figure 1 | Maps of California, showing the probability of exceeding MMI VI over the next 24-h period. The period starts at 14:07 Pacific Daylight Time on 28 July 2004. **a**, The time-independent hazard based on the 1996 USGS hazard maps for California. SF and LA are the locations of San Francisco and Los Angeles, respectively. **b**, The time-dependent hazard which exceeds the background including contributions from several events: 22 December 2003, San Simeon (SS, $M_w = 6.5$), a $M_l = 4.3$ earthquake 4 days earlier near Ventura (VB), a $M_l = 3.8$ event 30 min before the map was made near San Bernardino (FN), the 1999 Hector Mine $M_w = 7.1$ earthquake in the Mojave desert (LHM), and the 1989 $M_w = 6.9$ Loma Prieta (LP) earthquake. **c**, The combination of these two contributions, representing the total forecast of the likelihood of ground shaking in the next 24-hour period. **d**, The ratio of the time-dependent contribution to the background.

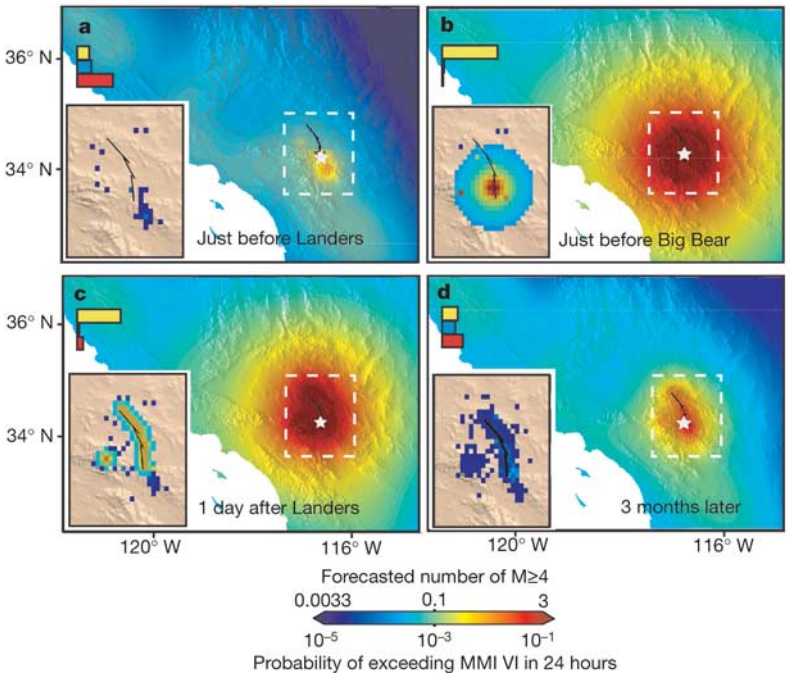

Figure 2 | Hazard maps calculated for four 24-hour intervals in the vicinity of the 1992 Landers ($M = 7.3$) earthquake. The colour shows the probability of exceeding MMI VI shaking intensity during interval (labels below colour scale). The black line indicates the Landers rupture and the white star is the Landers epicentre. The inset maps correspond to area within dashed white lines and indicate expected number of $M \geq 4.0$ aftershocks in the 24-hour interval (labels above colour scale). Histograms on upper left of figures indicate relative mean weight over inset area given to generic model (yellow), sequence-specific model (blue) and spatially heterogeneous model (red). **a**, Interval beginning 1 min before Landers mainshock shows increased hazard south of the Landers epicentre associated with the foreshocks to Landers (largest $M_l = 3.0$) and aftershocks of the ($M_w = 6.1$) Joshua Tree earthquake two months earlier. **b**, Interval beginning 2 hours after the mainshock and 1 hour before the ($M_w = 6.3$) Big Bear aftershock. **c**, Interval beginning 24 h after Landers mainshock. **d**, Interval beginning three months after mainshock.

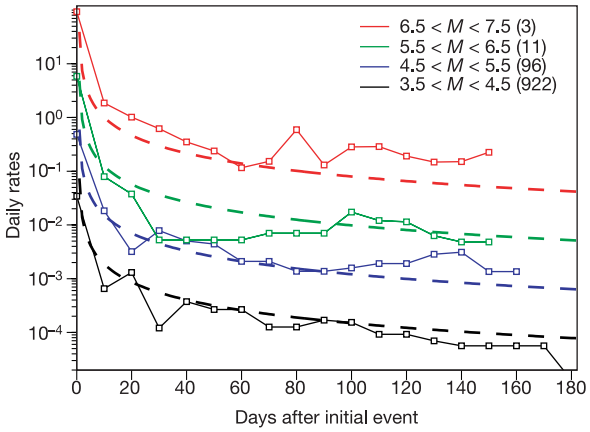

Figure 3 | Calculated and observed rates of events $M \geq 4$ in 24-hour intervals following mainshocks occurring between 1988 and 2002 in southern California. Dashed lines show the rates forecasted by the generic California clustering model (without cascades) for the mainshock magnitude (M) shown. For this test a simple circular aftershock zone implementation (solid lines) gives the observed rates of $M \geq 4.0$ aftershocks following all mainshocks with magnitude within 0.5 units of M. The aftershock zones are defined as the areas within one rupture length of the mainshock epicentre.

historical earthquake data (Fig. 1a) and the time-dependent contributions (Fig. 1b). The ratio of the time-dependent contribution to the background probability is shown in Fig. 1d.

We treat every earthquake ($M \geq 3$) as a potential mainshock and calculate an aftershock model for it. At any time, more than one earthquake may be contributing to the expected rate of aftershocks at a given location, most commonly after a large aftershock. In these cases, the earthquake producing the highest expected rate at a location is adopted, and its rate assigned to the node. Thus, a cascade of aftershock sequences[3] can be modelled, with secondary and tertiary sequences represented where and when they exceed the rate of the primary sequence. Having translated all models into common units of daily rate of events for each magnitude bin at each location, we have a straightforward procedure for deciding how to terminate a cluster: in our definition, a sequence remains active as long as at any one location a higher number of events ($M \geq 4$) is forecast than is contained in the background model at the same location.

The application of our approach to the 1992 Landers sequence is demonstrated in Fig. 2. Just before the mainshock, the background model dominates the earthquake hazard, except in the Landers region where foreshocks ($M_l \leq 3.0$) had just occurred and near the moment magnitude (M_w) = 6.1 Joshua Tree earthquake that occurred two months earlier (Fig. 2a). In the first hours after the mainshock, the generic clustering model dominates and is isotropic relative to the mainshock's epicentre (Fig. 2b). By one day after the mainshock, a mainshock rupture model is available, allowing fault-specific versions of both the generic and sequence-specific models to be calculated (Fig. 2c). Also, the $M_w = 6.3$ Big Bear aftershock has occurred, and is reflected in the forecast rate of aftershocks. However, the hazard is still nearly isotropic owing to the smoothing inherent in the hazard calculation and the shape of the combined aftershock zones. Three months later (Fig. 2d), the spatially heterogeneous model is used in most nodes. The modelled rate of aftershocks is lower but still heterogeneous, producing a heterogeneous hazard forecast.

As a first test, we verified that the generic clustering model[1,2] describes the average clustering activity of California reasonably well. Using data from 1988–2002, after the period used to initially develop the model and thus independent data, we compute the average daily rate of events following an earthquake of a given size (Fig. 3).

Next, using a likelihood-based testing algorithm developed as part of the Regional Earthquake Likelihood Models group of the Southern California Earthquake Centre (D. Schorlemmer, M.C.G., D. J. Jackson & S.W.; manuscript in preparation) we compared our model against successively more challenging null hypotheses. First, we showed that the background model of the 1996 National Hazard map could be rejected with a significance of $<1\%$. This is hardly surprising, as this model does not include earthquake clustering. In the second and third tests, we compared the full composite model to stages of the model, the second using only the generic clustering model, and the third using sequence-specific parameters when they could be determined. Again, we were able to reject both partial models with $<1\%$ significance. We can thus state that the composite model did a better job of predicting the seismicity of 1992–2001 in the region of Fig. 2 than the National Hazard maps, an average clustering model, or a sequence-specific parameter model.

Many earthquake-forecasting case studies have been made[14], but few have been tested systematically against a null hypothesis[15]. We believe strongly that quantitative testing is key to making progress in earthquake forecasting research. Our time-dependent hazard model has been tested against the null hypothesis of the long-term hazard model and represents our empirical understanding of aftershock behaviour. It could be a null hypothesis for future statistical or physical models in time-dependent forecast-related research.

The first hours to days after a mainshock are the most hazardous time, with the daily probabilities of exceeding MMI VI that approach 1 after the largest possible mainshocks. The need to characterize the relative and absolute increase in hazard is therefore at its greatest when emergency response activities are underway. Because the hazard is highest immediately after the mainshock, it is important to update quickly, preferably in an automatic system. Reliable real-time monitoring of larger aftershocks sequences is now possible in California, as demonstrated during the 1999 $M_w = 7.1$ Hector Mine event[16]; however, further improvements in monitoring and processing are likely once the new Advanced National Seismic System (ANSS) is implemented. This should enhance our model's forecasting ability because we will be able to move more quickly from the isotropic to a more realistic, heterogeneous model. Similar to real-time seismicity maps and ShakeMaps[13], 24-hour forecast maps such as the one shown in Fig. 1 are now available through the web (http://pasadena.wr.usgs.gov/step).

Received 19 January; accepted 7 April 2005.

1. Reasenberg, P. A. & Jones, L. M. Earthquake aftershocks: Update. *Science* **265,** 1251–1252 (1994).
2. Reasenberg, P. A. & Jones, L. M. Earthquake hazard after a mainshock in California. *Science* **243,** 1173–1176 (1989).
3. Helmstetter, A. & Sornette, D. Sub-critical and supercritical regimes in epidemic models of aftershock sequences, D. *J. Geophys. Res.* **107,** 2237, doi:10.1029/2001JB001580 (2002).
4. Felzer, K. R., Abercrombie, R. E. & Ekstrom, G. A common origin for aftershocks and multiplets. *Bull. Seismol. Soc. Am.* **94,** 88–99 (2004).
5. Frankel, A. *et al.* National seismic hazard maps. *U. S. Geological Survey Open-File Report*, 96–532 (1996).
6. Gutenberg, B. & Richter, C. F. Magnitude and energy of earthquakes. *Ann. Geofis.* **9,** 1–15 (1954).
7. Utsu, T., Ogata, Y. & Matsu'ura, R. S. The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth* **43,** 1–33 (1995).
8. Reasenberg, P. A. Foreshock occurrence before large earthquakes. *J. Geophys. Res.* **104,** 4755–4768 (1999).
9. Wiemer, S., Gerstenberger, M. & Hauksson, E. Properties of the aftershock sequence of the 1999 Mw 7.1 Hector Mine earthquake: implications for aftershock hazard. *Bull. Seismol. Soc. Am.* **92,** 1227–1240 (2002).
10. Wiemer, S. & Katsumata, K. Spatial variability of seismicity parameters in aftershock zones. *J. Geophys. Res.* **104,** 13135–13151 (1999).
11. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference A Practical Information-Theoretic Approach* 488 (Springer, New York, 2002).
12. Boore, D. M., Joyner, W. B. & Fumal, T. E. Equations for estimating horizontal response spectra and peak acceleration from western North America earthquakes: A summary of recent work. *Seismol. Res. Lett.* **68,** 128–153 (1997).
13. Wald, D. J., Quitoriano, V. & Heaton, T. H. TriNet "ShakeMaps": rapid generation of instrumental ground motion and intensity maps for earthquakes

in southern California. *Earthquake Spectra* **15,** 537–556 (1999).
14. Toda, S., Stein, R. S., Reasenberg, P. A. & Dieterich, J. H. Stress transferred by the Mw = 6.9 Kobe, Japan, shock: effect on aftershocks and future earthquake probabilities. *J. Geophys. Res.* **103,** 24543–24565 (1998).
15. Geller, R. J. Earthquake prediction: a critical review. *Geophys. J. Int.* **131,** 425–450 (1997).
16. Wiemer, S. Introducing probabilistic aftershock hazard mapping. *Geophys. Res. Lett.* **27,** 3405–3408 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are very grateful for the reviews from E. Field, D. Jackson, R. Simpson and A. Cornell. The work was supported by the Southern California Earthquake Center, ETH Zurich and the US Geological Survey.

Author Contributions This work was completed as part of the PhD thesis of M.C.G. under the supervision of S.W. at ETH-Zürich. L.M.J. and P.A.R. provided significant contributions to the model development.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C.G. (mattg@usgs.gov).