Geophysical Journal International

Geophys. J. Int. (2012) **191,** 653–658 doi: 10.1111/j.1365-246X.2012.05626.x

Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, earthquake: first results

K. Z. Nanjo,[1,2] H. Tsuruoka,[1] S. Yokoi,[1] Y. Ogata,[3] G. Falcone,[4] N. Hirata,[1] Y. Ishigaki,[5] T. H. Jordan,[6] K. Kasahara,[1] K. Obara,[1] D. Schorlemmer,[6,7] K. Shiomi[2] and J. Zhuang[3]

[1]*Earthquake Research Institute, University of Tokyo, Tokyo, Japan. E-mail: nanjo@eri.u-tokyo.ac.jp*
[2]*National Research Institute for Earth Science and Disaster Prevention, Ibaraki, Japan*
[3]*Institute of Statistical Mathematics, Tokyo, Japan*
[4]*Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy*
[5]*Japan Meteorological Agency, Tokyo, Japan*
[6]*Southern California Earthquake Center, University of Southern California, Los Angeles, CA, USA*
[7]*GFZ German Research Centre for Geosciences, Potsdam, Germany*

Accepted 2012 July 23. Received 2012 July 21; in original form 2011 August 15

SUMMARY

Although no deterministic and reliable earthquake precursor is known to date, we are steadily gaining insight into probabilistic forecasting that draws on space–time characteristics of earthquake clustering. Clustering-based models aiming to forecast earthquakes within the next 24 hours are under test in the global project 'Collaboratory for the Study of Earthquake Predictability' (CSEP). The 2011 March 11 magnitude 9.0 Tohoku-Oki earthquake in Japan provides a unique opportunity to test the existing 1-day CSEP models against its unprecedentedly active aftershock sequence. The original CSEP experiment performs tests after the catalogue is finalized to avoid bias due to poor data quality. However, this study differs from this tradition and uses the preliminary catalogue revised and updated by the Japan Meteorological Agency (JMA), which is often incomplete but is immediately available. This study is intended as a first step towards operability-oriented earthquake forecasting in Japan. Encouragingly, at least one model passed the test in most combinations of the target day and the testing method, although the models could not take account of the megaquake in advance and the catalogue used for forecast generation was incomplete. However, it can also be seen that all models have only limited forecasting power for the period immediately after the quake. Our conclusion does not change when the preliminary JMA catalogue is replaced by the finalized one, implying that the models perform stably over the catalogue replacement and are applicable to operational earthquake forecasting. However, we emphasize the need of further research on model improvement to assure the reliability of forecasts for the days immediately after the main quake. Seismicity is expected to remain high in all parts of Japan over the coming years. Our results present a way to answer the urgent need to promote research on time-dependent earthquake predictability to prepare for subsequent large earthquakes in the near future in Japan.

Key words: Time-series analysis; Probabilistic forecasting; Seismicity and tectonics; Computational seismology; Statistical seismology; Asia.

GJI Seismology

1 INTRODUCTION

The aftershock sequence following the 2011 March 11 magnitude (M) 9.0 Tohoku-Oki earthquake (hereinafter, Tohoku earthquake) is of strong and immediate scientific interest for seismologists and relevant to the critical problem of how seismic hazard changes with time. A better understanding of time-dependent earthquake (aftershock) hazard based on time-dependent forecast models would greatly benefit the public, emergency planners and the media (Jordan *et al.* 2011). In this Letter, we describe a predictability study that is testing existing short-term forecast models against the aftershock sequence. We analysed five different models that create forecasts for the following day on a daily basis (hereinafter 1-day models): ETES, ERS, ETAS, HIST-ETAS5pa and HIST-ETAS7pa (Falcone *et al.* 2010; Ogata 2011; Zhuang 2011). They are currently under test in an earthquake forecast experiment in Japan as part of the international project 'CSEP: Collaboratory for the Study of Earthquake Predictability' (Jordan 2006). CSEP conducts rigorous and strictly prospective forecast experiments for various tectonic environments: California, New Zealand, western Pacific, Japan, Italy and globally. To avoid any possible bias, forecasts are issued for future periods and are tested only against future observations. However, we relax our experimental protocol, using a data set which is often incomplete but is immediately available.

2 METHOD

The models we used fall into the category of space–time clustering models based on the epidemic-type aftershock sequence (ETAS) model (Ogata 1999). The ETES (Falcone *et al.* 2010) is a simpler space–time version of the ETAS model with location-independent parameters. The ERS model (Falcone *et al.* 2010) merges, into a single algorithm, the ETES model and the rate–state constitutive law for seismicity rate (Ruina 1983; Dietrich 1994). The ETAS model (Zhuang 2011) also deals with location-independent model parameters. Into this model, he introduced a way to obtain the occurrence probabilities of forecast earthquakes through simulation. This model has an option called offline parameter estimation that allows it to update its parameters every few days, and this option was selected for the California CSEP experiment. However, as for the Japan CSEP experiment, this feature is not used. Instead, fixed parameters are adopted through all the testing experiment. Contrast to the above models, two variants of the HIST-ETAS model (HIST-ETAS5pa and HIST-ETAS7pa) were developed to deal not only with anisotropic clustering in space but also with location-dependent model parameters (Ogata 2011). The difference between the two is that some parameters associated with time and space components (precisely, the parameters μ, K, α, p and q in eq. (2) of Ogata (2011)) are spatially varying in the HIST-ETAS5pa model while only μ and K are location-dependent in the HIST-ETAS7pa model. More description of the models with equations is given in the respective papers that are open access (for a summary, see also Nanjo *et al.* 2011) whereas Table S1 lists some important model parameter values used in the experiment. In response to quality control of forecasts preliminary generated, the ETAS and HIST-STAS models were altered from their original CSEP versions by the corresponding modellers in advance of this experiment.

All models incorporate the Gutenberg–Richter (GR) relation (Gutenberg & Richter 1944) to relate the forecast frequency of earthquakes to their magnitudes (Figs S1 and S2). The GR *b* values vary among models: spatially varying *b* values (HIST-ETAS models in Figs S1d and e) and location-independent *b* values (ETES, ERS, ETAS models in Figs S1a–c).

Most seismologists had not anticipated an *M*9-class earthquake off the Pacific coast of Tohoku before it actually occurred. Similarly, the national seismic hazard map of Japan also had not included the real possibility of such a large quake; only *M*7–8 class earthquakes were considered realistic. However, all models simply assume the GR relation without tapering off at large magnitudes.

Each forecast model predicts the number of earthquakes in each space–time–magnitude bin of the testing area in Fig. 1. This area is subdivided into $0.1^\circ \times 0.1^\circ$ cells and each cell into magnitude bins in the range $5 \leq M \leq 9$ in steps of 0.1 magnitude units. The forecast periods are consecutive 1-day time windows, each starting at midnight in JST. The first forecast time window was set to start at 00:00 on 2011 March 8. Following the current CSEP experiment, this study design slices any fully time–space correlated non-Poissonian model into supposedly independent forecast and observation space–time–magnitude bins with assumed Poisson

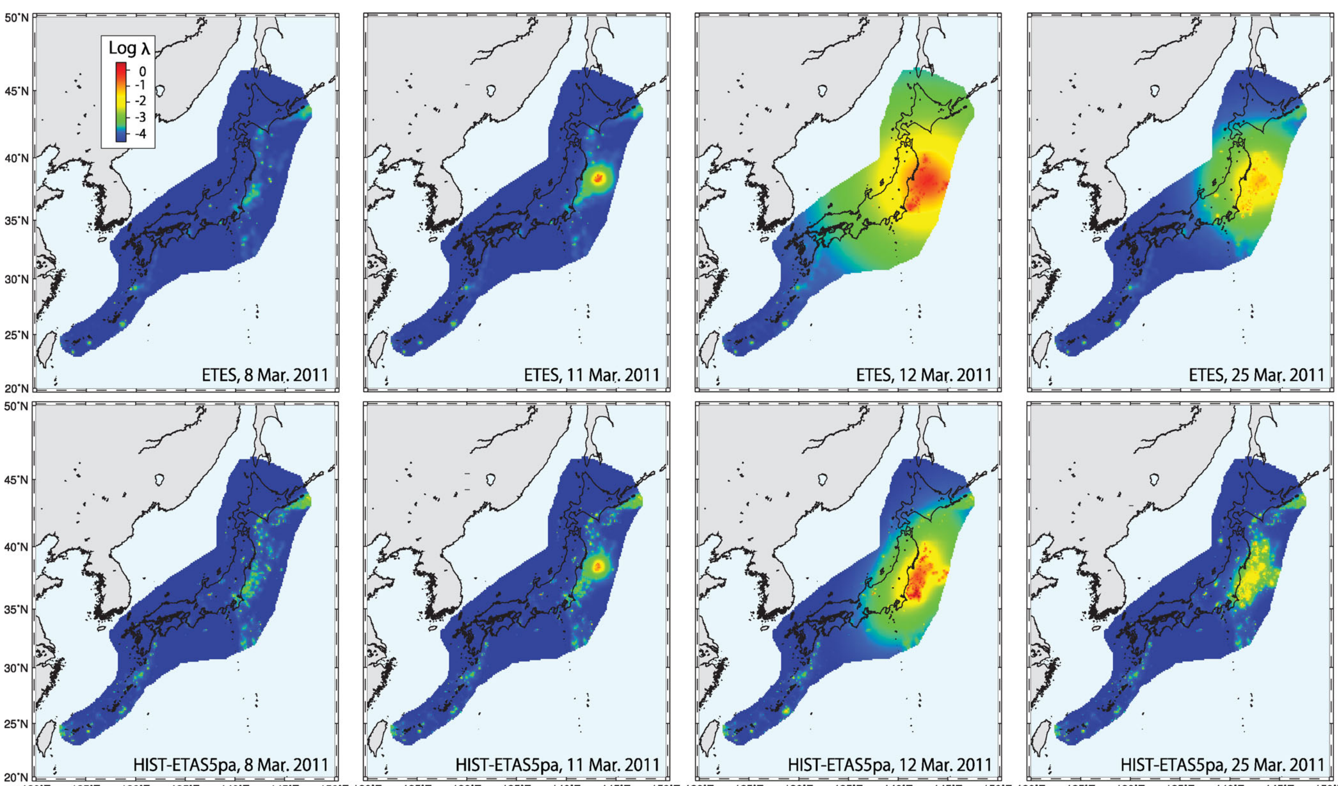

Figure 1. Seismicity forecast maps derived from the ETES (top panels) and HIST-ETAS5pa (bottom panels) models. Colour corresponds to the logarithm of the forecast rate (λ) of all events with $M \geq 4.0$ for cells of $0.1^\circ \times 0.1^\circ$. From left to right, forecasts for 2011 March 8, 11, 12 and 25.

confidence bounds. This assumption introduces a bias into individual tests (Werner & Sornette 2008; Lombardi & Marzocchi 2010; Werner *et al.* 2011). Keeping this bias in mind, some aspects of the test results will be discussed thoroughly.

We used three CSEP tests (Zechar *et al.* 2010a) to evaluate the consistency between observation and forecast in terms of the total number of events (*N*-test), their spatial distribution (*S*-test) and their magnitude distribution (*M*-test). The results are given in terms of quantile scores: δ_1 and δ_2 indices for the *N*-test, ζ index for the *S*-test and κ index for the *M*-test. For the *N*-test, the forecast rate is too low (an underestimation) if δ_1 is very small and it is too high (an overestimation) if δ_2 is very small. A model fails the test, indicating inconsistency between forecast and observation, if its score is below a significance level of 2.5 per cent, the same as used in the original CSEP testing framework. We do not show results of the *L*-test, a consistency test using the joint log-likelihood (Schorlemmer *et al.* 2007), because several studies have found them to be highly dependent on results of the *N*-test (Schorlemmer *et al.* 2010; Zechar *et al.* 2010a; Werner *et al.* 2011; Tsuruoka *et al.* 2012). The *R*-test (Schorlemmer *et al.* 2007) was convincingly shown not to be very useful (e.g. Rhoades *et al.* 2011).

To avoid any bias and to warrant data reliability, the original CSEP testing is designed to be performed after a delay that allows for manual revision of earthquake locations by the seismic network operators. This delay is currently four months as of 2011 March 30 for the catalogue maintained by the Japan Meteorological Agency (JMA). To conduct this specific forecasting experiment in a timely manner to satisfy the interest in its results, we relaxed the original CSEP testing protocol to use the preliminary JMA (P-JMA) catalogue. The P-JMA catalogue, updated and revised daily, often contains erroneous locations and incomplete listings, even for large events. By deciding to use it, however, we have intended to take a first step towards operability-oriented earthquake forecasting in Japan, which may rely on some components that are yet to be finalized scientifically for Japanese earthquakes. We used the combined JMA/P-JMA catalogue both as input data for generating forecasts and as reference data against which those forecasts are tested. The data consist of the JMA's finalized catalogue for earthquakes through 2010 November 30 and the P-JMA catalogue that was available as of 2011 March 30 for earthquakes from 2010 December 1 to 2011 March 27.

The use of the P-JMA catalogue limits the study because it is unclear how well models perform scientifically. Best scientific conclusions are deferred until data reliability is warranted. We replaced the P-JMA catalogue with the JMA finalized catalogue (available as of 2011 December 20) to check the validity of some aspects of forecasts and observations.

3 RESULTS

Seismicity forecasts were considerably higher near the epicentre of the Tohoku earthquake on the day of the mainshock (second column in Fig. 1) relative to the background levels several days before (first column in Fig. 1). The difference can be attributed to the foreshock sequence (Hirose *et al.* 2011). On the day after the mainshock, the entire aftershock region was highlighted remarkably by higher forecast rates (third column in Fig. 1) due to the mainshock and the frequent aftershocks that immediately followed. This increased forecast still remained clearly visible 2 weeks after the mainshock (fourth column in Fig. 1).

Encouragingly, of the total 80 trials (four types of tests for 20 single days), all models failed in the test only in four cases (the δ_1 index for March 9–11 in Fig. 2c and the ζ index for March 11 in Fig. 2e), although the modellers had no opportunity to optimize the models to an *M*9-class earthquake and the input catalogue was incomplete. However, the trials are neither independent forecasts nor independent observations. Generally, the models fail on the most active days. Failures across models tend to happen simultaneously. One could easily achieve a better rate if the study period were extended to include more days with less activity. The important forecasts (right after the mainshock) seem to be most difficult for the models, which is similar to the 1992 Landers case (Woessner *et al.* 2011).

In detail, Figs 2(a), (c) and (d) show that, for March 9–11, all models failed to forecast the abrupt increase in the total number of foreshocks, the mainshock and the frequent aftershocks that followed immediately. After this abrupt increase was over, the ETES model provided consistent predictions for the total numbers on and before March 17 (the overprediction on March 18 was due to the abrupt decrease in the observed number) and then tended to overestimate the numbers. By contrast, the other models tended to underestimate the numbers, although the accuracy improved visibly with time.

The deviation of the spatial distribution of all models from the observed distribution on March 11 (*S*-test in Fig. 2e) was due to an extreme underestimation of the number of earthquakes in the Tohoku earthquake aftershock zone (Figs 2a, c and d). The accuracy of the HIST-ETAS models tended to improve with time, but its timing to start recovering seems to have been delayed, compared with the other models (ETES, ERS and ETAS). Such bad scores of the HIST-ETAS models are again due to the CSEP's 1-d forecasting protocol. A substantial change in the aftershock spatial distribution took place after March 11. On March 11, the aftershocks occurred mainly in the western side of the mainshock towards the coasts, but after that they appear to have extended towards east beyond the trench line including the outer rise area. On the other hand, the aftershock range in latitude remained similar. To confirm, see the figures in Ogata (2012).

One of the major differences between the HIST-ETAS forecasts and the other forecasts is that the former are spatially anisotropic whereas the latter are isotropic. Although the HIST-ETAS models well predicted the spatial range of aftershocks during the first day because the centroids of the HIST-ETAS aftershock forecasts were determined in the first hour after the mainshock (Ogata 2011), these could not predict the seemingly delayed off-fault aftershocks; see the bottom panels of Fig. 1. Because the *S*-test series in Fig. 2(e) was done only for the next days' forecasts, the forecasting in the same day was ignored according to the CSEP protocol. On the other hand, the ETES, ERS and ETAS forecasts are spatially isotropic around a mainshock, and thus these models well forecast the off-fault aftershock area in the next day onwards; see the top panels of Fig. 1.

Incidentally, the forecasts of the immediate secondary aftershocks during the same day were ignored according to the 1-day forecasting protocol. However, if the forecast period is a 1-month time window of 2011 March without daily updating, the *S*-test results of the same five models show a different picture. Ogata *et al.* (2012 revised) evaluated 1-month space forecast throughout 2011 March. The obtained log-likelihoods showed better performance of the HIST-ETAS models than the others. Another major difference between the HIST-ETAS forecasts and the other forecasts is due to the parameter of the aftershock productivity; namely, whether it is

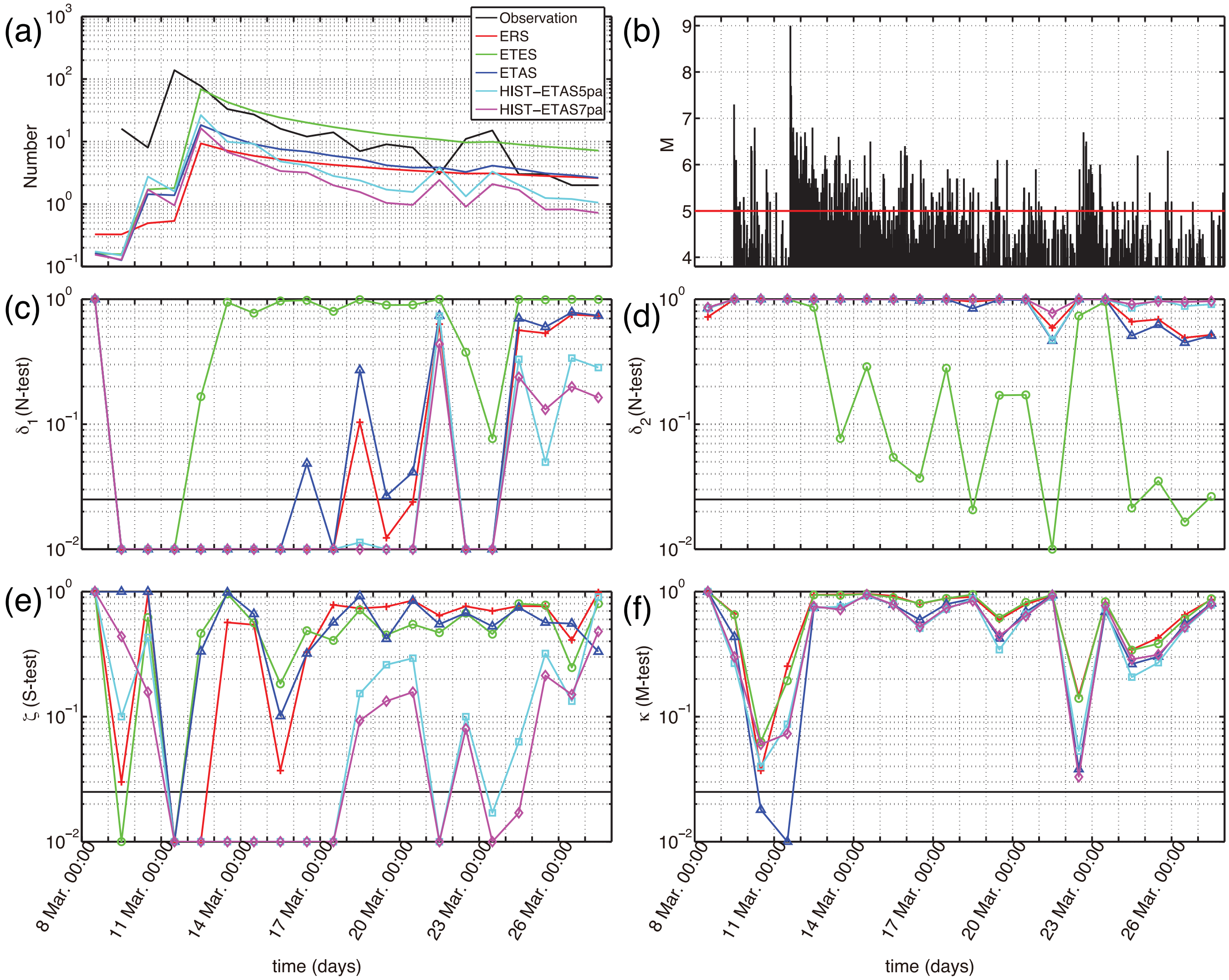

Figure 2. Assessment of 1-day forecast models. (a) Daily numbers of $M \geq 5.0$ events; (b) magnitude–time diagram; (c) N-test δ_1 index; (d) N-test δ_2 index; (e) S-test ζ index; (f) M-test κ index. The significance level is 2.5 per cent (horizontal line). Index values smaller than 0.01 are replaced by 0.01 for display purpose. The Tohoku earthquake and its three M7-class aftershocks at very early times (Hirose *et al.* 2011), which are not missing from the P-JMA catalogue, overlap each other.

spatially varying or constant over Japan. Indeed, the epicentres of the Tohoku aftershocks suggest a non-smooth and highly clustered spatial density.

The κ indices (M-test) were above the significance level for all models, except for the ETAS model that fails in the test on March 10 and 11 (Fig. 2f). It can be seen in Fig. 3 that the failure on March 10 was due to the fact that the ETAS forecast is characterized by a higher b value ($b = 0.8$) than the observation ($b = 0.3$–0.5) and the other forecasts ($b = 0.6$–0.7). We see the same feature for March 11 (Fig. S2). On March 22, all models passed the M-test, but κ indices lowered due to a M6.7 event with its immediate frequent aftershocks.

Catalogue replacement does not show significant change in both forecast and observed distributions (Figs 3 and S2). It also represents that inadequacies in the learning input could be vanishingly responsible for the models' general tendency to underpredict the number of earthquakes. Moreover, visual inspection for individual models shows that sequential forecast maps (such as Fig. 1), generated by using the P-JMA and the JMA catalogues, are quite similar to each other. Thus, the models' performance is not pronouncedly affected by inadequacies in the catalogue (such as incompleteness at low magnitudes).

4 DISCUSSION AND CONCLUSION

We described a predictability study of forecasting models for the aftershocks of the Tohoku earthquake. Our approach is based on testing five different 1-day forecast models used in CSEP. The main difference from the original CSEP experiment is in the use of the temporary P-JMA catalogue to allow for immediate testing after the first weeks of the aftershock sequence that otherwise would be delayed due to the finalization of the regular JMA catalogue.

All 1-day forecasts, except for ETES, seem to have been poor at forecasting the event numbers (Figs 2c and d). It is also seen in Fig. 2(e) that the spatial accuracy of the HIST-ETAS models has been improved in a delayed fashion. These may reveal a problem in the CSEP testing setup. The CSEP protocol for daily forecasts

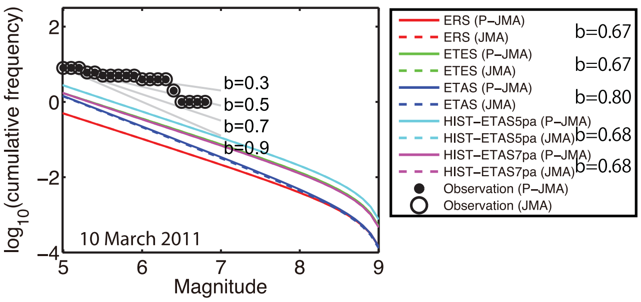

Figure 3. The total frequency–magnitude distributions of earthquakes on 2011 March 10 for the five models. We summed frequencies over spatial bins for each magnitude bin and plotted the cumulative frequency as a function of M. Forecasts based on the P-JMA and JMA catalogues are shown by solid and dashed lines, respectively, but they are overlapping each other. The respective observations are shown by solid and open circles. All models assume the GR relation up to $M9$; truncated curves near $M9$ are attributed to no inclusion of $M > 9$ forecasts into the cumulative frequencies, because the models did not forecast $M > 9$ events according to the experiment rules (See also Fig. S2a and its inset). Reference b values in grey: $b = 0.3$, 0.5, 0.7 and 0.9.

allows each forecast model to take into account the seismicity of the previous day. However, seismicity is changing rapidly over time during the beginning of an aftershock sequence. Therefore, we can expect accuracy to improve if the forecasts can be updated more frequently than at 1-day intervals.

Alternatively, refining a way to simulate immediate secondary aftershocks triggered by aftershocks that supposedly occur in the testing day will lead to constructing short-term models optimized for the current CSEP 1-day protocol. We see the idea introduced by Zhuang (2011) as a first contribution towards an improved ability for CSEP 1-day forecasting. However, we still feel the need of investigating this approach if the test results in Fig. 2 are taken into consideration.

There is also a scientific reason for the mismatch of forecasted numbers of events and the observations. We fit the simple (temporal) ETAS model (Ogata 1999) to the seismicity in the aftershock region for the period from 1997 January 1 through the mainshock occurrence time, and extrapolated the trend until 2011April 30, using the mainshock magnitude of 9.0 as an indispensable piece of information (Fig. 4). The fitting seems to be well acceptable (future research will include the Kolmogorov–Smirnov and RUNS tests to support the fitting statistically). The estimate after the Tohoku earthquake covers the expected cascade of earthquakes including aftershocks of expected aftershocks. The observed cumulative number of aftershocks was about twice as large as forecasted. This does not change when we used a minimum magnitude threshold of $M4$ instead of $M5$. It indicates that the Tohoku earthquake has remarkably high aftershock fertility, even taking into consideration the unusually large mainshock magnitude.

Therefore, an important lesson learned here is that it is an essential framework that allows forecasts to be updated at shorter than daily intervals, optimally every time after an event and to flexibly revise parameter estimates. In all five models tested in this study, the parameters are fixed at the start and cannot be revised to account for changing patterns of seismicity. In this paper, we did not study how the parameters change if they are fit progressively into the aftershock sequence. This approach will lead to intriguing development for more-realistic time-dependent earthquake forecasting.

The ETES is the most basic among the five models whereas the other four models introduced parameters to capture detailed pat-

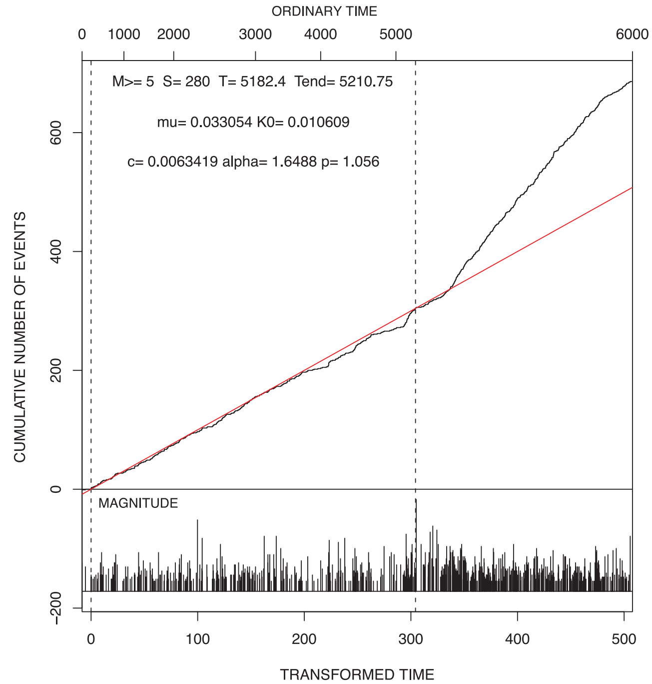

Figure 4. (Top) Fitting of the simple (temporal) ETAS model (red line) to the accumulated number of $M \geq 5.0$ events in the aftershock region (black curve) from 1997 January 1 through the mainshock occurrence time (two vertical dashed lines). The trend was extrapolated until 2011 April 30. The labels on the top denote the actual number of days. The labels on the bottom indicate time transformed in such a way that the expected number is a time-independent constant so that the predicted trend is linear. The observed cumulative number through 2011 April 30 was about 400, compared to about 200 events predicted. The period between two dashed lines is the target period for which the ETAS model parameters (p, c, K_0, μ and α) are computed (Ogata 1999): p, c and K_0 are parameters of the Omori–Utsu formula, μ represents constant-rate background seismicity, and α represents the exponent of the aftershock productivity law. The delay between the time of the mainshock and the upturning of the black line is about 1 hour in actual time. (Bottom) Magnitude–time diagram for $M \geq 5.0$ events.

terns of seismicity. In other words, the models other than the ETES expect to be able to be well-optimized to seismic activity prior to the Tohoku earthquake. Because the Tohoku earthquake drastically changed seismicity, this optimization may face a potential penalty for post-quake forecasting. We suggest that the ETES model is just better suited in the number forecast to this particularly unprecedented aftershock sequence. This point is supported by our further analysis for the individual models, replacing the P-JMA catalogue with the JMA finalized catalogue. It led to little improvement in both forecasts and observations (Figs 3 and S2). This shows that the effect of catalogue replacement on the test results would be practically nought.

Although the existing models may be applicable to operational real-time forecasting in a sense that their performance remains stable over the catalogue replacement, the forecasting ability need to be improved to assure the reliability (agreement with the observed data to be collected over many times), especially for the days immediately after large earthquakes.

Short-term fully operational earthquake forecasting already exists in other regions. The STEP (short-term earthquake probability) forecasts (Gerstenberger *et al.* 2005) of real-time earthquake probabilities in California have been released to the public by the US Geological Survey for some years. Operational earthquake forecasting was also used in a crisis situation in Italy following the 2009 L'Aquila earthquake, where ETAS-based aftershock forecasts were issued daily by the Istituto Nazionale di Geofisica e Vulcanologia (Marzocchi & Lombardi 2009). This work does not focus on fully operational forecasting, but provides a case study of the systematic testing of operability-oriented forecasting in Japan, using the combined JMA/P-JMA catalogue.

The history of earthquakes in Japan gives ample reasons to expect that an enhanced state of seismicity may continue for some time on the order of a decade in all parts of Japan. Because of this fact, operability-oriented (or ideally fully operational) earthquake forecasting needs to be performed as an immediate response to such seismicity, motivated by the expectations, both scientific and social/public, on the scientists. Improving the existing CSEP infrastructure (Schorlemmer & Gerstenberger 2007; Zechar *et al.* 2010b) and learning from similar research activities in the world such as California and Italy (Gerstenberger *et al.* 2005; Marzocchi & Lombardi 2009; Woessner *et al.* 2011) may help to accelerate research in that direction (Jordan *et al.* 2011).

ACKNOWLEDGMENTS

The authors thank the Editor, D. Agnew and two anonymous reviewers for their comments. We also want to thank the open-source community for the Linux operating system and the many programs used in this study. We used the JMA earthquake catalogues and the Generic Mapping Tools. This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. The SCEC contribution number for this paper is 1663. This work was conducted under the auspices of the Special Project for Earthquake Disaster Mitigation in the Tokyo Metropolitan Area.

REFERENCES

Dietrich, J., 1994. A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. geophys. Res.,* **99,** 2601–2618.

Falcone, G., Console, R. & Murru, M., 2010. Short-term and long-term earthquake occurrence models for Italy: ETES, ERS and LTST, *Ann. Geophys.,* **53**(3), 41–50.

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

Gutenberg, B. & Richter, C.H., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34,** 185–188.

Hirose, F., Miyaoka, K., Hayashimoto, N., Yamazaki, T. & Nakamura, M., 2011. Outline of the 2011 off the Pacific coast of Tohoku Earthquake (M_W 9.0)—seismicity: foreshocks, mainshock, aftershocks, and induced activity—, *Earth Planets Space,* **63,** 513–518.

Jordan, T.H., 2006. Earthquake predictability, brick by brick, *Seismol. Res. Lett.,* **77**(1), 3–6.

Jordan, T.H. *et al*., 2011. Implications of recent events for CSEP testing of earthquake forecasts, in *2011 Annual Meeting of Seismological Society of America*, Memphis, TN, 2011 April 13–15, 43pp.

Lombardi, A.M. & Marzocchi, W., 2010. The assumption of Poisson seismic-rate variability in CSEP/RELM experiments, *Bull. seism. Soc. Am.,* **100**(5A), 2293–2300.

Marzocchi, W. & Lombardi, A.M., 2009. Real-time forecasting following a damaging earthquake, *Geophys. Res. Lett.,* **36,** L21302, doi:10.1029/2009GL040233.

Nanjo, K.Z., Tsuruoka, H., Hirata, N. & Jordan, T.H., 2011. Overview of the first earthquake forecast testing experiment in Japan, *Earth Planets Space,* **63**(3), 159–169.

Ogata, Y., 1999. Seismicity analyses through point-process modelling: a review, *Pure appl. Geophys.,* **155**(2–4), 471–507.

Ogata, Y., 2011. Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity, *Earth Planets Space,* **63**(3), 217–229.

Ogata, Y., 2012. Tohoku earthquake aftershock activity and triggered activity near Matsumoto, *Report of the Coordinating Committee for Earthquake Prediction,* **87,** 546–553 (in Japanese with English figure captions).

Ogata, Y., Katsura, K., Falcone, G., Nanjo, K.Z. & Zhuang, J., 2012. Comprehensive and topical evaluations of earthquake forecasts in number, time, space and magnitude, *Bull. seism. Soc. Am.,* submitted.

Rhoades, D.A., Schorlemmer, D., Gerstenberger, M.C., Christophersen, A., Zechar, J.D. & Imoto, M., 2011. Efficient testing of earthquake forecasting models, *Acta Geophys.,* **59,** 728–747.

Ruina, A., 1983. Slip instability and state variable friction laws, *J. geophys. Res.,* **88**(B12), 10 359–10 370.

Schorlemmer, D. & Gerstenberger, M.C., 2007. RELM Testing Center, *Seism. Res. Lett.,* **78**(1), 30–36.

Schorlemmer, D., Gerstenberger, M.C., Wiemer, S., Jackson, D.D. & Rhoades, D.A., 2007. Earthquake likelihood model Testing, *Seism. Res. Lett.,* **78**(1), 17–29.

Schorlemmer, D., Zechar, J.D., Werner, M.J., Field, E.H., Jackson, D.D., Jordan, T.H. & the RELM Working Group, 2010. First results of the Regional Earthquake Likelihood Models experiment, *Pure appl. Geophys.,* **167**(8–9), 859–876.

Tsuruoka, H., Hirata, N., Schorlemmer, D., Euchner, F., Nanjo, K.Z. & Jordan, T.H., 2012. CSEP Testing Center and the first results of the earthquake forecast testing experiment in Japan, *Earth Planets Space,* in press.

Werner, M.J. & Sornette, D., 2008. Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. geophys. Res.,* **113,** JB08302, doi:10.1029/2007JB005427.

Werner, M.J., Helmstetter, A., Jackson, D.D. & Kagan, Y.Y., 2011. High resolution long- and short-term earthquake forecasts for California, *Bull. seism. Soc. Am.,* **101**(4), 1630–1648.

Woessner, J. *et al*., 2011. A retrospective comparative forecast test on the 1992 Landers sequence, *J. geophys. Res.,* **116,** B05305, doi:10.1029/2010JB007846.

Zechar, J.D., Gerstenberger, M.C. & Rhoades, D.A., 2010a. Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. seism. Soc. Am.,* **100**(3), 1184–1195.

Zechar, J.D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J. & Jordan, T.H., 2010b. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency and Computation: Practice and Experience,* **22**(12), 1836–1847, doi:10.1002/cpe.1519.

Zhuang, J., 2011. Next-day earthquake forecasts for the Japan region generated by the ETAS model, *Earth Planets Space,* **63**(3), 207–216.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. The normalized non-cumulative frequency-magnitude distribution for nodes.
Figure S2. Same as Fig. 3 for several days: 2011 March 8, 9, 10, 11, 12, 21, 22 and 25 (see the caption of Fig. 3).
Table S1. Some parameter values used in the experiment.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.