

# ***Nonparametric Aftershock Forecasts Based on Similar Sequences in the Past***

by **Nicholas J. van der Elst and Morgan T. Page**

## **ABSTRACT**

The basic premise behind aftershock forecasting is that sequences in the future will be similar to those in the past. Most forecasts use empirically tuned parametric distributions to approximate the behavior of past sequences and project those distributions into the future. Although parametric models do a good job of capturing the average behavior in a population, they are not explicitly designed to capture the full range of variability between sequences, and sometimes suffer from instabilities or inaccuracies due to overtuning of the model. Here, we present a nonparametric forecast method that cuts out the parametric “middleman” between training data and forecast. The forecast is drawn directly from past outcomes of sequences that appear similar to the target sequence, with similarity defined as the Poisson probability that the event count in a past sequence comes from the same underlying intensity as the event count to-date in the target sequence. The forecast is just the distribution of previously observed event counts, weighted by their similarity. The similarity forecast is only marginally less accurate than the parametric [Reasenberg and Jones \(1989; hereafter RJ89\)](#) method. The rate of severe underpredictions, however, is much lower for the similarity forecast. Although 10% of observed sequences exceed the upper 2.5% range of the RJ89 forecast range, only 3% exceed this range for the similarity forecast. Given an adequate database of past events, the similarity method makes overtuning impossible, minimizes the rate of surprises, and serves as a useful benchmark for more precisely tuned parametric forecasts.

*Electronic Supplement:* An html document for calculating similarity forecasts based on the set of 2307 global aftershock sequences used in this study.

## **INTRODUCTION**

“From what has actually been, we have data for concluding with regard to that which is to happen thereafter” ([Hutton, 1788](#), p. 217).

Aftershock forecasting typically relies on statistical models of past seismic sequences. These models can be more or less sophisticated but rely on empirically tuned parametric distributions

to forecast magnitudes, times, and locations of aftershocks. A typical model might consist of the Gutenberg–Richter relation ([Gutenberg and Richter, 1944](#)) for the magnitude probability distribution, the Omori law for the temporal probability distribution ([Utsu, 1961; Utsu et al., 1995](#)), and perhaps an exponential productivity relation to model the scaling of direct and secondary aftershock productivity with mainshock magnitude ([Utsu, 1971; Ogata, 1988](#)).

The application of parametric models, such as the epidemic-type aftershock sequence (ETAS) model ([Ogata, 1988, 1992](#)), or the simpler Reasenberg and Jones forecasting model ([Reasenberg and Jones, 1989; hereafter, RJ89](#)), allows a forecast to be tuned to a specific sequence, and allows predictions to be made outside the realm of previously observed behavior. These parametric models are designed to capture the typical behavior in a collection of sequences. It is less clear that these models capture the true range of variability between sequences. In standard applications using a central estimate of the forecast model, the variability between sequences is assumed to be a Poisson realization of the average rate model (the Poisson variability is compounded within cascade models such as ETAS, but variability is still fundamentally Poissonian). More sophisticated applications may attempt to propagate some of the epistemic uncertainty surrounding the central parameter estimates ([Omi et al., 2015](#)), but this still may not capture the true distribution of past and future event counts.

Although parametric forecasting models have great utility (both authors are actively engaged in their development), there may be some use for a model that hews as closely as possible to describing past outcomes of similar sequences, without the intermediate step of tuning a parametric probability distribution. Such a model would favor generality over precision and be guaranteed to capture the range of previously observed behavior. A forecast derived purely from the outcomes of past sequences is also immune to runaway instabilities that can arise when projecting epidemic-type parametric models outside the range over which they have been tuned ([Helmstetter and Sornette, 2002](#)).

In this article, we propose a nonparametric forecast method that relies only on two basic assumptions: (1) self-similar triggering—the observation that magnitude-normalized

aftershock sequences tend to look similar; and (2) that aftershock sequences that start out looking similar are likely to continue looking similar in the future. Both of these assumptions are common to parametric forecasts as well; here, we evaluate how far we can get with just this foundation. This approach requires a large catalog of past aftershock sequences to draw from; we apply it here to global seismicity.

### The Self-Similarity Assumption

Mathematically, the statement of self-similarity is as follows. Aftershock productivity increases exponentially with mainshock magnitude  $M_{\text{ms}}$ , with the number of aftershocks greater than some cutoff magnitude  $M_c$  described by

$$N(M_c|M_{\text{ms}}) = 10^{a+\alpha(M_{\text{ms}}-M_c)}, \quad (1)$$

with parameters  $a$  and  $\alpha$  (Ogata, 1988; Helmstetter and Sornette, 2003; Felzer *et al.*, 2004).

It is also widely assumed that the magnitudes of aftershocks follow the Gutenberg–Richter distribution:

$$N(m|M_{\text{ms}}) = N(M_c|M_{\text{ms}})10^{-b(m-M_c)}. \quad (2)$$

Combining equations (1) and (2)

$$N_{\text{as}}(m|M_{\text{ms}}) = 10^{a+\alpha M_{\text{ms}}-bm-(\alpha-b)M_c}. \quad (3)$$

The parameters  $\alpha$  and  $b$  are typically found to be approximately equal (and similar to 1) (Felzer *et al.*, 2004; Helmstetter *et al.*, 2005). The scenario ( $\alpha = b$ ) is referred to as self-similar, the meaning of which becomes clear when we substitute ( $\alpha = b$ ) into equation (3)

$$N(m|M_{\text{ms}}) = 10^{a+b(M_{\text{ms}}-m)} \quad (4a)$$

$$N(\Delta m) = 10^{a-b(\Delta m)}, \quad (4b)$$

in which  $\Delta m = m - M_{\text{ms}}$  is the differential magnitude with respect to the mainshock. Self-similarity is hence the assumption that aftershock sequences of different magnitude mainshocks are statistically equivalent when defined in terms of differential magnitude relative to the mainshock.

Given the assumption of self-similar triggering, we propose a nonparametric approach for statistical aftershock forecasting. Looking at the distribution of aftershock numbers and differential magnitudes in past sequences, we define a subset of sequences that appear similar in a quantifiable sense to the sequence being forecast, and base our forecast directly on the range of observed outcomes of those sequences.

## METHOD

### Defining Sequence Similarity

The core task in generating a similarity forecast is defining the set of similar past sequences. We want the forecast to be as close as possible to model-free, but we need to define a sim-

ilarity metric that includes some finite neighborhood around a specific target sequence. One option would be to choose some arbitrary number range around the target sequence count. Another option is to treat the number of events observed to-date in the target sequence as a realization of a random process with some underlying rate. For each past sequence, its similarity is the probability that its event count was generated from the same random process. This is the approach we follow here.

Consistent with the previous work on aftershock forecasting, we choose a Poisson model to relate observed aftershock counts to an underlying statistical rate. This model has limitations (Kagan, 2010), but is optimally simple, and lends itself well to a parameter-free definition of similarity. Alternatives for quantifying similarity include a boxcar function or a Gaussian smoothing function, but the Poisson model has the advantage of being always positive, having a range that increases with the mean value, and requiring no super parameters.

To define similarity, we take the observed number of events so far  $N_1$  and treat it as a random variable realized from a Poisson distribution with unknown intensity parameter  $\lambda$ . We then seek out all past sequences of  $N_i$  events that could be consistent with  $\lambda$ . Assuming an uninformative (uniform) prior distribution for  $\lambda$ , the probability of obtaining  $N_i$  events in a new trial, given that  $N_1$  events were observed in an initial trial (the ongoing sequence) is

$$P(N_i|N_1) = \int_0^\infty P(N_i|\lambda)P(\lambda|N_1)d\lambda, \quad (5)$$

which has the solution

$$P(N_i|N_1) = 2^{-(N_i+N_1+1)} \frac{(N_i+N_1)!}{N_i!N_1!}. \quad (6)$$

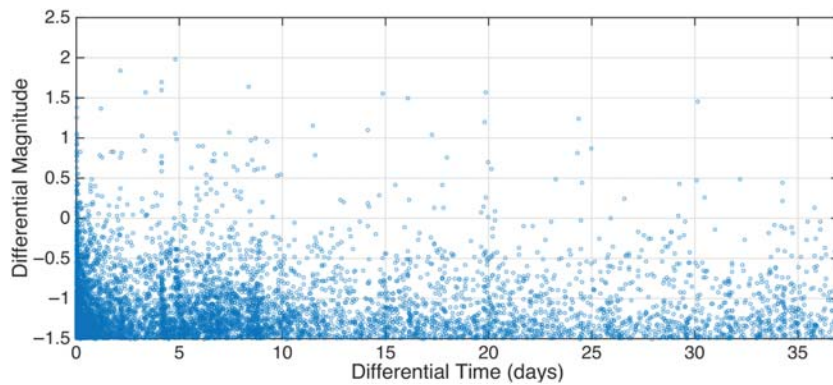
The probabilities  $w_i = P(N_i|N_1)$  are used to weight and combine the observed outcomes of all past sequences into a probability distribution for the outcome of the current sequence. That is, for each past sequence, we count the number of events  $N_2$  within the desired forecast window, and assign that value a weight  $w_i$ . The weighted distribution of  $N_2$  constitutes a probability distribution for the outcome of the ongoing sequence.

### Forecast Metrics

The similarity forecast, like any other forecast, can be defined in terms of the probability of exceeding some differential magnitude threshold (e.g., the magnitude of the mainshock), or by the mean, median, and range of the number of events  $N_2$  in the forecast window. The probability of getting at least one event is given by the Poisson-weighted fraction of similar sequences with  $N_2 \geq 1$ . The mean is defined as the Poisson-weighted mean of  $N_2$ . Likewise, the median and ranges of the  $N_2$  forecast are defined as quantiles of the cumulative distribution of the weights.

### Defining Past Aftershock Sequences

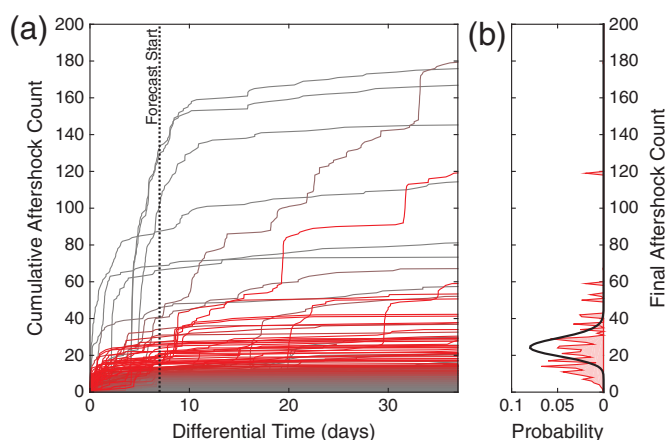
In this study, we define mainshocks as  $M \geq 6$  earthquakes with depths of less than 300 km that are at least five rupture lengths



▲ **Figure 1.** 2307 worldwide aftershock sequences, stacked by differential time and magnitude with respect to the mainshock.

away from all larger events in the previous year. For rupture length, we use the Wells and Coppersmith (1994) relation that averages over all focal mechanisms. Aftershocks of those mainshocks are defined as all subsequent earthquakes with differential magnitude  $\Delta m \geq -1.5$ , occurring within three mainshock rupture lengths. (The rupture length for an **M** 6 is 12 km.) The cutoff on differential magnitude is chosen to be consistent with the mainshock magnitude cutoff of **M** 6, assuming a global average detection cutoff magnitude of **M** 4.5 (Page *et al.*, 2016). Using a lower differential magnitude cutoff may result in more informative forecasts because there is more variability in event counts at lower differential magnitude thresholds. However, there is a direct trade-off between the number of aftershocks gained by lowering the cutoff, and the number of remaining sequences that meet the more stringent completeness level.

We focus on short-term aftershock forecasts, here, with observation windows up to one week and forecast windows up to one month (30 days). We therefore select aftershocks



▲ **Figure 2.** (a) Normalized aftershock sequences, colored by their similarity with a hypothetical target sequence with  $N_1 = 20$  aftershocks after 7 days. (b) Distribution of final event counts, weighted by their similarity with the target sequence. The solid line gives the Poisson forecast with the same mean, for comparison.

that occur within 37 days following a mainshock. We do not restrict aftershocks to be of smaller magnitude than the mainshock. The stacked sequences are shown in Figure 1.

## RESULTS

### Model Output

The similarity metric is illustrated in Figure 2, which shows the complete set of normalized aftershock sequences, colored according to their similarity with a hypothetical sequence with  $N_1 = 20$  aftershocks after 7 days. The similarity forecast distribution (Fig. 2b) has a coefficient of variation (COV) of  $\sim 14$ , much larger than the equivalent Poisson forecast with  $\text{COV} = 1$ .

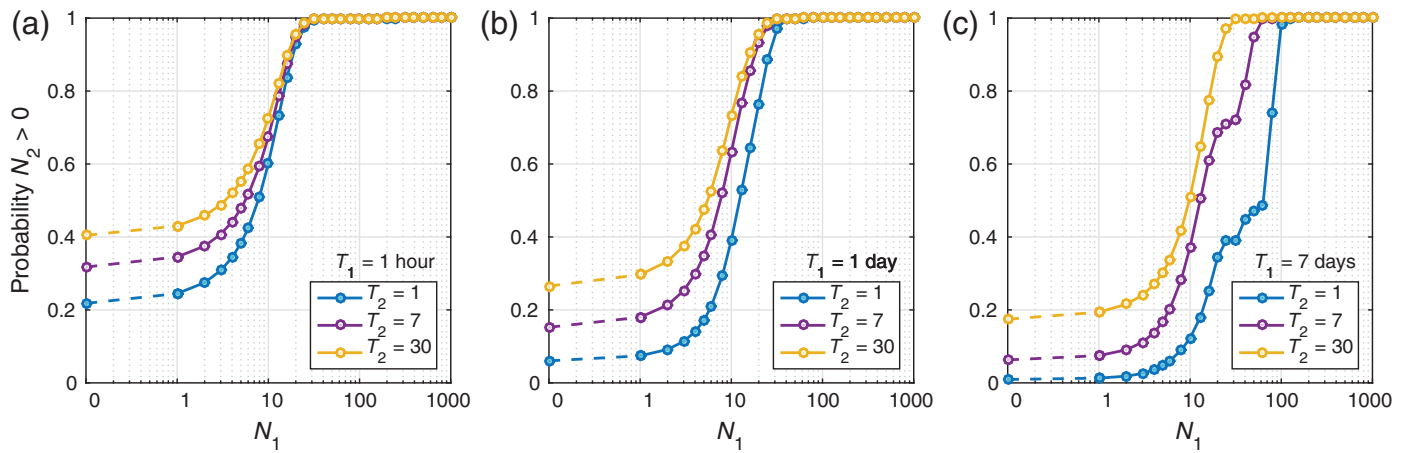
Figure 3 shows the forecast results as a function of the number  $N_1$  of  $M \geq M_{\text{ms}} - 1.5$  aftershocks observed to-date in a hypothetical aftershock sequence, for a range of observation windows (training intervals) and forecast durations. The forecast is the smoothest at early times and low  $N_1$ , where most of the data reside. Conversely, a short duration forecast made after a long observation window (e.g.,  $T_1 = 7$  days,  $T_2 = 1$  day) is not very robust, because the data are more sparse, especially if  $N_1$  is unusually large (Fig. 3c). As a final step, we could regularize the forecast such that the probability of  $N_2 > 0$  increases monotonically or smoothly with  $N_1$ . We do not apply an additional regularization step in this study, so that we can evaluate the maximally simple version of the method.

### Verification and Validation

To validate the similarity forecast, we take each of the 2307 sequences identified in the global catalog and compute a forecast based on its similarity with the remaining 2306 sequences (the leave-one-out method). Because there is no real training of the similarity model, there is no need to isolate a disjoint training and testing set.

We test forecasts for observation intervals ranging from 1 hr to 7 days and forecast durations ranging between 1 and 30 days. The test statistic is the  $\log_{10}$  ratio of predicted to observed events, where the predicted count is taken to be the mean or median of the forecast distribution. Using the median gives an equal number of overpredictions and underpredictions, and approximates the mode of the distribution. To compute the ratio of events for cases where zero events are predicted or observed, we regularize all ratios by again assuming that the counts derive from a Poisson distribution and taking the ratios of the expected Poisson intensities  $\lambda_i$ , rather than the actual observed numbers  $N_i$  (Park *et al.*, 2006). Assuming a noninformative (uniform) prior distribution for  $\lambda_i$ , this regularization amounts to  $\langle \lambda_1 \rangle / \langle \lambda_2 \rangle = (N_1 + 1) / (N_2 + 1)$  (see the Appendix).

The distribution of  $\log_{10}$  ratios for the median-based forecast skews slightly toward the left (Fig. 4). The asymmetry reflects the fact that underprediction errors tend to be larger



▲ **Figure 3.** Empirical probability of observing at least one additional earthquake of magnitude  $\Delta M \geq -1.5$ , given  $N_1$  earthquakes observed so far, for various observation windows  $T_1$  and forecast windows  $T_2$ . (a)  $T_1 = 1$  hr; (b)  $T_1 = 1$  day; (c)  $T_1 = 7$  days.

than overprediction errors because there is a finite limit ( $N = 0$ ) on how far a sequence can fall below the prediction, but there is no such limit on underpredictions. Additionally, the median forecast value is often equal to zero, in which case overprediction is impossible.

### The Rate of Surprises

The main advantage of the similarity forecast is that it precisely captures the full range of variability observed in past sequences,

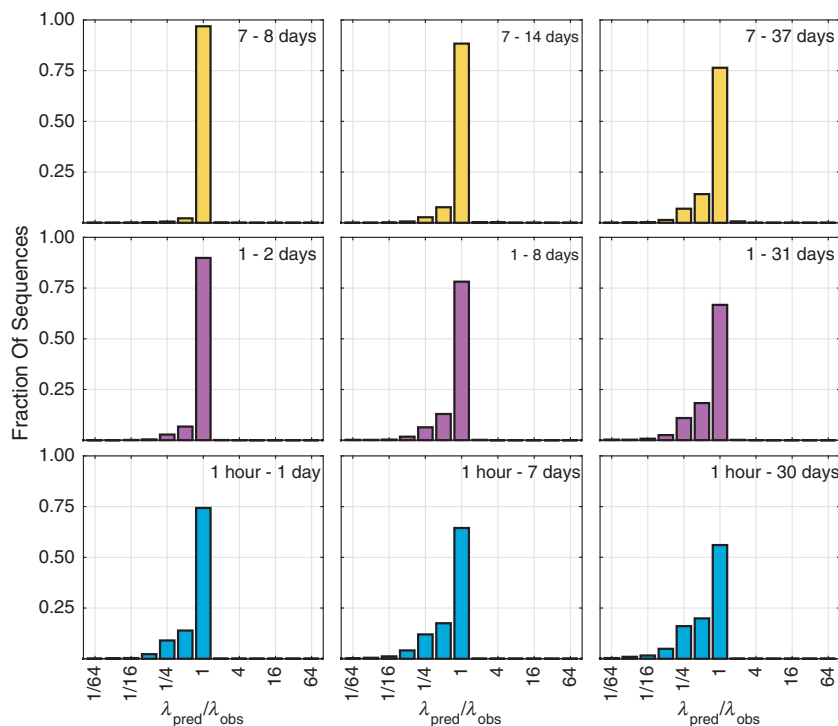
without appealing to any parametric model of aleatory variability. To evaluate the forecast range, we plot the value of the cumulative forecast probability associated with each observation, defining

$$q(n_i) = P_i(N < n_i) + \mu P_i(N = n_i), \quad (7)$$

in which  $n_i$  are the observed event counts,  $P_i(N)$  are the forecast probabilities as a function of event count  $N$ , and  $\mu$  is a uniform random number between 0 and 1 introduced to give a smooth distribution of  $q$  despite the discrete distribution of each forecast  $P_i(N)$ . If the forecast range is perfect, the distribution of  $q(n_i)$  should be uniform. The distribution of  $q(n_i)$  for the similarity forecast is indeed nearly uniform for all observation and forecast windows (Fig. 5).

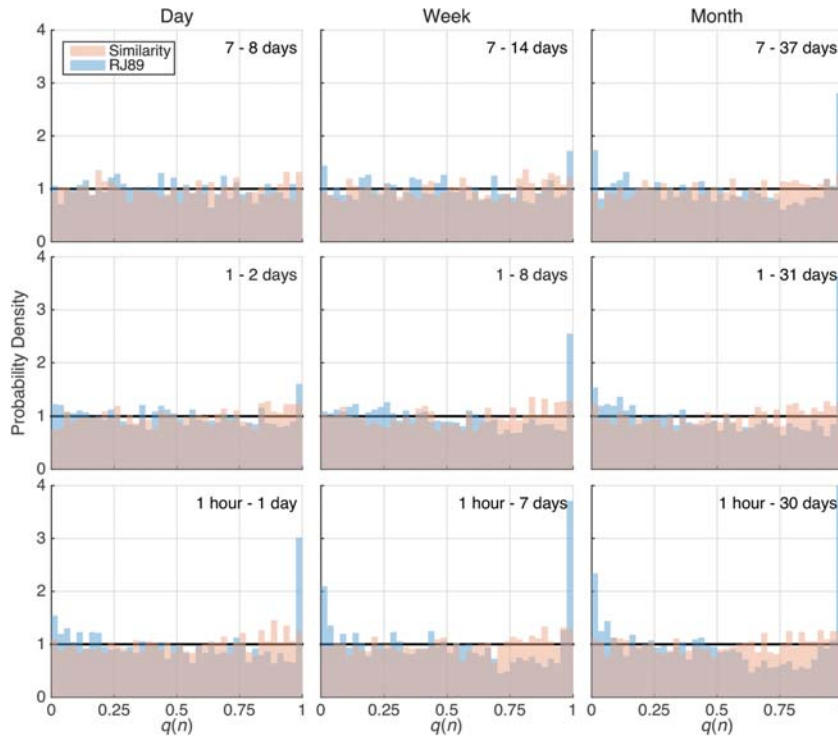
For the purposes of aftershock forecasting, we are most concerned with sequences that fall outside the specified confidence limits of our forecast, which we term surprises. We arbitrarily define a surprising sequence as one with an outcome that exceeds the upper 97.5% range of the forecast. The surprise rate should be approximately equal to 2.5% if the forecast is working perfectly.

The observed surprise rate depends on the observation interval and forecast duration. It is 3.1% for one-month forecasts made after the first hour. For longer observation intervals, the surprise rate drops, with a 2.3% surprise rate for one-month forecasts issued after 1 week. The rate of surprises for anomalously low sequences is essentially nil (the central 95% confidence range almost always includes the value  $N_2 = 0$ ). This is not simply a consequence of the forecast construction, as each target sequence is removed from the distribution of possible outcomes for the purposes of its own forecast.



▲ **Figure 4.** Histograms of prediction errors for the similarity forecast, based on the number of aftershocks above  $M_{ms} - 1.5$ . Different panels correspond to different observation windows and forecast durations, given by the pairs of numbers in each panel. Training intervals increase vertically; forecast intervals increase horizontally to the right.



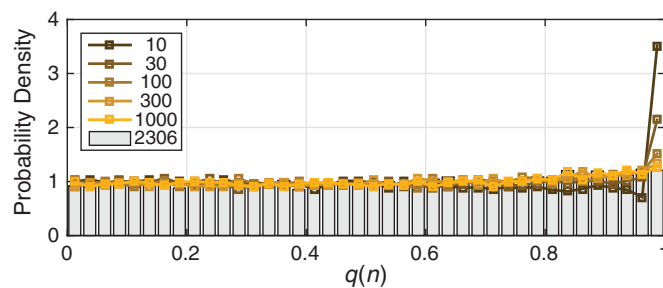


▲ **Figure 5.** Distribution of the cumulative probability of each observed event count with respect to its own forecast (equation 7). An ideal forecast should give a uniform distribution. Pink, the nonparametric similarity forecast; blue, the Reasenberg and Jones (1989; hereafter, RJ89) method assuming Poissonian variability. Forecast start and end times are given in the legend.

The surprise rate also depends on the database of past sequences. If the database is small, the prevalence of never-before-seen sequences will be higher. For catalogs greater than about 100 events, however, the surprise rate converges below 3.5% (Fig. 6).

### Comparison to Reasenberg and Jones (RJ89)

The similarity forecast approach is intended to make overtuning the forecast impossible, capturing all of the real variability



▲ **Figure 6.** Distribution of the cumulative probability of each observed event count with respect to its own forecast (equation 7), as a function of the size of the database of past events. Distributions combine all nine of the observation and forecast widows presented in Figure 5.

between similar past sequences. We therefore expect the similarity forecast to be less precise than model-based approaches such as RJ89, but perhaps with a smaller bias and certainly with a lower rate of surprises.

We here compare the results of the similarity forecast to a sequence-specific RJ89 forecast, using global average Omori parameters as a Bayesian prior for regularizing the sequence-specific estimates (Page *et al.*, 2016). The use of a Bayesian prior means that a generic (average) forecast is issued if zero events are observed in the training/observation interval. Parameters  $a$  and  $p$  are assumed Gaussian with mean  $-2.54$  and  $0.92$ , and standard deviations  $0.71$  and  $0.14$ , respectively. Parameter  $c$  is fixed to  $0.018$  days.

The RJ89 model is based on the Omori law and an exponential productivity relationship, governed by empirical parameters  $a$ ,  $c$ , and  $p$ :

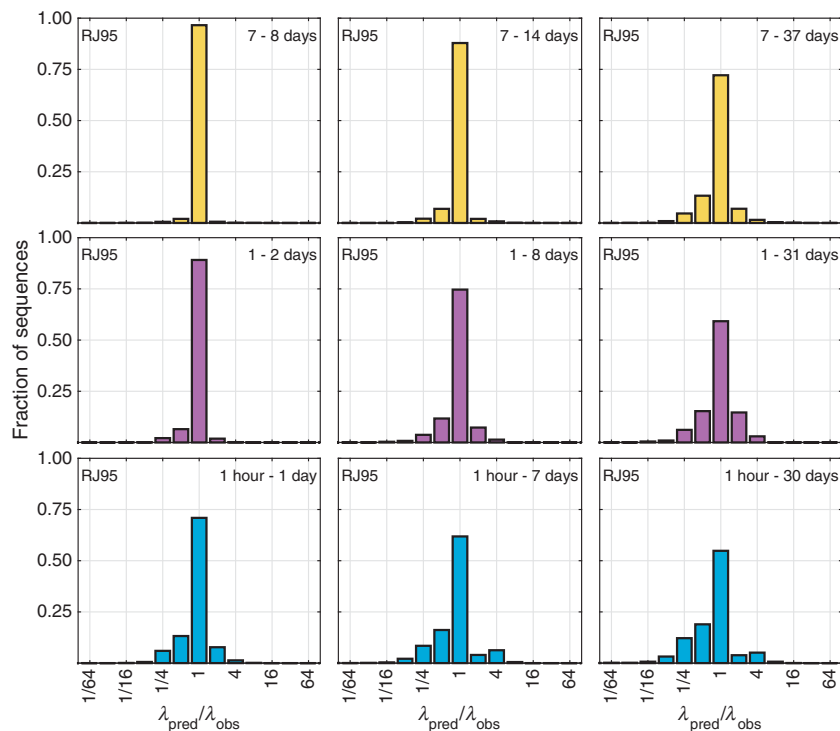
$$r(t) = 10^{a+M_{ms}-M_c}(t+c)^{-p}. \quad (8)$$

The distribution of the forecast number is assumed Poissonian, with intensity parameter equal to the integral of equation (8) over the forecast interval.  $M_c$  is here set to  $M_{ms} - 1.5$ .

For the Bayesian regularization, we assume the parameters  $a$  and  $p$  are independent and Gaussian distributed ( $a_\mu = -2.54$ ,  $a_\sigma = 0.71$ ,  $p_\mu = 0.92$ ,  $p_\sigma = 0.14$ ). The distribution of  $a$  reflects a Gaussian approximation of an inversion result based on the first 10 days of a set of global aftershock sequences very similar to the set used in this study (Page *et al.*, 2016). The mean and standard deviation for  $p$  come from the variability between different tectonic regions found in that same study, whereas parameter  $c = 0.018$  days is fixed to the global average in that study.

The performance of the RJ89 model is comparable to the similarity forecast (Figs. 7 and 8). The main difference between the forecasts is that the misfit between the median forecast value and the observations is much more asymmetrical for the similarity forecast than for RJ89 (Fig. 8). This reflects the fact that the true distribution of observed event counts used in the similarity forecast is skewed toward zero, whereas the Poisson model assumes a more symmetrical distribution around the expectation. Misfits based on the mean forecast value are more symmetrical but result in a larger fraction of overpredictions (Fig. 9), again reflecting the highly skewed distribution of observed event counts.

As expected, the surprise rate is quite a bit larger for the RJ89 forecast than for the similarity forecast. About 8%–14% of sequences exceed the upper 97.5% limit of the one-month RJ89 forecasts, depending on the length of the training interval, compared to  $\sim 3\%$  for the similarity forecast (Fig. 8).



▲ **Figure 7.** Histograms of Reasenberg and Jones forecast errors for the global dataset, based on the number of aftershocks above  $M_{ms} - 1.5$ . Training intervals increase vertically; forecast intervals increase to the right (compare with Fig. 4).

An ETAS-based forecast, which includes sequence variability due to secondary triggering from large aftershocks, would likely prove more accurate than the more simplistic RJ89 approach. We nevertheless focus on RJ89 because it is the closest to being a parametric version of the similarity forecast we have developed here. In future iterations, the similarity forecast could include information about the magnitude distribution—rather than just the number—of aftershocks so far, in which case the ETAS model would be the appropriate model for comparison.

## DISCUSSION

### Advantages and Disadvantages of the Similarity Forecast

The proposed similarity forecast is designed to be as close as possible to model-free. This minimizes the possibility of overfitting and of surprise sequences that exceed the model confidence bounds. The approach is limited in that it cannot predict behavior for sequences outside the realm of previous observations. This shortcoming can be mitigated by supplying a large set of past sequences data. Of course, collecting sequences over a larger area may mask real differences related to the tectonic environment of the mainshock (Page *et al.*, 2016).

We made no attempt to remove possible background earthquakes from the catalog for either the similarity forecast or the RJ89 forecast. Instead, we assume that the background

rate is negligible compared to the aftershock rate in the first 2 months, within three rupture lengths of  $M \geq 6$  mainshocks. Indeed, the measured background rate in the global stack is one earthquake per 230 days in the 1–12 months prior to the mainshocks used in this study (the one-month limit helps avoid foreshocks). This background rate is less than 2% of the average rate during the first week of aftershocks. However, the measured background rate constitutes about 17% of the rate in the 7–37 days postmainshock. Some fraction of the variability in the one-month forecasts may therefore reflect statistics of the background rate, and pushing the similarity forecast method to longer time intervals (say a year) will likely require some consideration of catalog declustering, depending on the intended application.

### Other Similarity Metrics

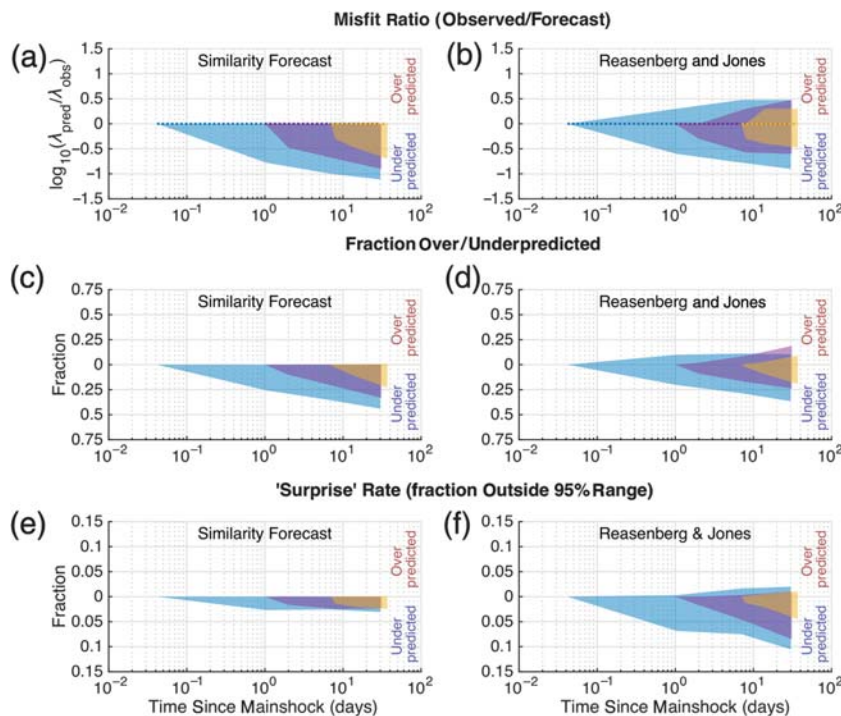
We designed the similarity measure to be particularly simple: it is based only on the number of earthquakes observed so far, above a relatively high differential magnitude threshold ( $\Delta M \geq -1.5$ ). The use of a smaller magnitude threshold would likely improve the forecast performance, but would require a lower magnitude of completeness, which may be hard to guarantee in the early times after a large earthquake.

A simple way to expand the definition of similarity would be to measure the number of earthquakes at additional differential magnitude thresholds and combine the respective similarity forecasts. Other more sophisticated characteristics may also be useful such as moments of the interevent time or differential magnitude distributions, but these would require a more sophisticated machine learning approach to defining sequence similarity and optimizing the weighting of the various characteristics. We leave this for future research.

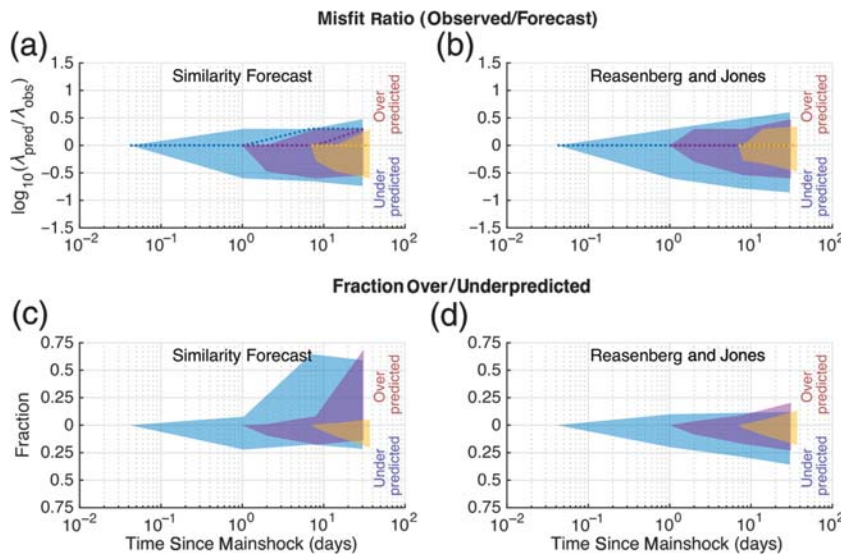
## CONCLUSION

We presented a nonparametric similarity-based aftershock forecasting method, based on the assumption of self-similar triggering, and a Poisson definition of similarity between sequences. The forecast is meant to provide a direct answer to the question: “given  $N$  observed aftershocks so far, how many aftershocks should we expect in some future interval based on similar past sequences?” Although similarity between sequences is defined with respect to a Poisson process, the number forecast is constructed directly from the outcomes of the similar sequences, rather than on any parametric model of aleatory variability.

The forecasting method is as close to model-free as possible, and serves as a reference forecast for comparison with more precisely tuned methods such as Reasenberg and Jones




▲ **Figure 8.** Comparison of model performance, using median forecast values. Colored patches start at different times since mainshock (hour, day, week) and grow with the length of the prediction interval (day, week, month). (a,b) Misfit ( $\log_{10}$  ratio) between predicted and observed numbers. Dotted lines give median, colored patches give central 95% range. (c,d) Fraction of sequences overpredicted and underpredicted by the median forecast value. (e,f) Surprise rate, defined as the fraction of events outside the central 95% forecast range.



▲ **Figure 9.** Comparison of model performance, using mean forecast values. (a,b) Misfit ( $\log_{10}$  ratio) between predicted and observed numbers. Dotted lines give the median, colored patches give the central 95% range. (c,d) Fraction of sequences overpredicted or underpredicted by the mean forecast value.

(1989) or ETAS (Ogata, 1988). Although the method ignores sequence-specific details such as the specific rate of the Omori decay, it is guaranteed to capture the true variability observed in past sequences. In general, the forecast performance is comparable to the parametric Reasenberg and Jones method, but with a much lower rate of surprises that exceed the upper 97.5% range of the forecast.


We envision this forecasting approach as being a useful supplement to existing parametric methods. Work on ensemble forecasts has shown that optimal success is typically obtained by combining multiple models conditioned on the data in different ways (Marzocchi *et al.*, 2012). The similarity forecast fills a methodological gap in the existing suite of parametric approaches.

The similarity forecast method has the advantage that the confidence bounds are representative of true global variability in sequences, and overtuning is impossible, assuming the database is large enough to be representative of past behavior. The simplified model-free approach may prove appealing to emergency managers or nonspecialists who may have little expertise or confidence in parametric models but nonetheless wish to explore the dependencies of the forecast on training and forecast intervals. Readers may generate and explore similarity forecasts using a Javascript code and HTML graphical user interface included in the  electronic supplement to this article.

## DATA AND RESOURCES

Earthquake data were downloaded from the U.S. Geological Survey Comprehensive Catalog (<https://earthquake.usgs.gov/data/comcat/>) going back to 1990 (last accessed July 2017).

## ACKNOWLEDGMENTS

This article benefitted from constructive reviews by Andrew Michael, Susan Hough, and two anonymous reviewers. The authors also acknowledge Karen Felzer, who proposed a similar approach years ago. 

## REFERENCES

- Felzer, K. R., R. E. Abercrombie, and G. Ekstrom (2004). A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.* **94**, no. 1, 88–98.



- Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **4**, 185–188.
- Helmstetter, A., and D. Sornette (2002). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.* **107**, no. B10, 2237, doi: [10.1029/2001JB001580](https://doi.org/10.1029/2001JB001580).
- Helmstetter, A., and D. Sornette (2003). Bath's law derived from the Gutenberg–Richter law and from aftershock properties, *Geophys. Res. Lett.* **30**, no. 20, doi: [10.1029/2003gl018186](https://doi.org/10.1029/2003gl018186).
- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005). Importance of small earthquakes for stress transfers and earthquake triggering, *J. Geophys. Res.* **110**, no. B05S08, doi: [10.1029/2004JB003286](https://doi.org/10.1029/2004JB003286).
- Hutton, J. (1788). Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the globe, *Trans. Roy. Soc. Edinb.* **I**, part II, 209–304.
- Kagan, Y. Y. (2010). Statistical distributions of earthquake numbers: Consequence of branching process, *Geophys. J. Int.* **180**, 1313–1328.
- Marzocchi, W., J. D. Zechar, and T. H. Jordan (2012). Bayesian forecast evaluation and ensemble earthquake forecasting, *Bull. Seismol. Soc. Am.* **102**, no. 6, 2574–2584.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* **83**, no. 401, 9–27.
- Ogata, Y. (1992). Detection of precursory relative quiescence before great earthquakes through a statistical-model, *J. Geophys. Res.* **97**, no. B13, 19,845–19,871.
- Omi, T., Y. Ogata, Y. Hirata, and K. Aihara (2015). Intermediate-term forecast of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches, *J. Geophys. Res.* **120**, 2561–2578, doi: [10.1002/2014JB011456](https://doi.org/10.1002/2014JB011456).
- Page, M. T., N. van der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and inter-sequence variability, *Bull. Seismol. Soc. Am.* **106**, no. 5, 2290–2301, doi: [10.1785/0120160073](https://doi.org/10.1785/0120160073).
- Park, T., V. L. Kashyap, A. Siemiginowska, D. A. van Dyk, A. Zetas, C. Heinke, and B. J. Wargelin (2006). Bayesian estimation of hardness ratios: Modeling and computations, *Astrophys. J.* **652**, 610–628.
- Reasenber, P. A., and L. M. Jones (1989). Earthquake hazard after a mainshock in California, *Science* **243**, no. 4895, 1173–1176.
- Utsu, T. (1961). A statistical study on the occurrence of aftershocks, *Geophys. Mag.* **30**, no. 4, 521–605.
- Utsu, T. (1971). Aftershocks and earthquake statistics (III)—Analyses of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquakes occurrence (1), *J. Fac. Sci. Hokkaido Univ. Ser. VII* **3**, 379–441.
- Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995). The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth* **43**, no. 1, 1–33.
- Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.* **84**, 974–1002.

## APPENDIX

### THE EXPECTATION OF LAMBDA FOR A POISSON PROCESS

Given some observed event count  $N$ , what is the expectation of the Poisson intensity  $\lambda$ ? The maximum-likelihood estimate of  $\lambda$  is just equal to the event count  $N$ , but we are interested in the average value of  $\lambda$  that is consistent with  $N$ , assuming  $\lambda$  can take any value with equal probability. The expectation of  $\lambda$  given  $N$  is defined

$$\langle \lambda | N \rangle \equiv \int_0^\infty \lambda f(\lambda | N) d\lambda, \quad (\text{A1})$$

which for a Poisson distribution is

$$\langle \lambda | N \rangle = \int_0^\infty \lambda \frac{\lambda^N e^{-\lambda}}{N!} d\lambda. \quad (\text{A2})$$

Multiplying the top and bottom by  $(N + 1)$ , we get the convenient result

$$\langle \lambda | N \rangle = (N + 1) \int_0^\infty \frac{\lambda^{N+1} e^{-\lambda}}{(N + 1)!} d\lambda. \quad (\text{A3})$$

Because the term in the integral is just the Poisson likelihood of  $\lambda$  given event count  $N + 1$ , it must integrate to 1, leaving

$$\langle \lambda | N \rangle = (N + 1) \quad (\text{A4})$$

Nicholas J. van der Elst  
Morgan T. Page  
U.S. Geological Survey  
Earthquake Science Center  
525 South Wilson Avenue  
Pasadena, California 91106 U.S.A.  
[nvanderelst@usgs.gov](mailto:nvanderelst@usgs.gov)

Published Online 20 December 2017