Title
Description of earthquake aftershock sequences using prototype point patterns

Authors
Frederic Paik Schoenberg
Katherine E. Tranbarger

Publication Date
2011-10-25

Description of earthquake aftershock sequences using prototype point patterns

Frederic Paik Schoenberg[1] and Katherine E. Tranbarger[1]

Abstract

We introduce the use of prototype point patterns to characterize the behavior of a typical aftershock sequence from the global Harvard earthquake catalog. These prototypes are used not only for data description and summary but also to identify outliers and to classify sequences into groups exhibiting similar aftershock behavior. We find that a typical shallow earthquake of magnitude between 7.5 and 8.0 has five aftershocks of magnitude at least 5.5, and these aftershocks are roughly evenly distributed in log-time between 0.113 days and 2.0 years after the mainshock. The relative magnitudes and distances from the mainshock for the typical aftershock sequence are characterized as well.

Key words: distance metrics, earthquakes, clustering, point processes.

[1] Department of Statistics, 8125 Math-Science Building, University of California, Los Angeles, 90095-1554.

1 Introduction.

A basic question in seismology is the following: what does the typical aftershock sequence look like? That is, after the occurrence of an earthquake of a certain size, what is typically observed in the vicinity of that earthquake, within a given period of time thereafter?

This question has been partially answered by the standard parametric models used to describe earthquake occurrences, such as the widely-used epidemic-type aftershock sequence (ETAS) model (Ogata 1988; Ogata 1998). Such models incorporate well-known parametric forms for the decay in the conditional rate of aftershocks with time elapsed since a triggering event, as well as the overall rate of earthquakes as a function of magnitude. The relation governing the rate of aftershocks over time is known as the modified Omori law:

$$\lambda(t) = \frac{K}{(t+c)^p}, \tag{1}$$

where $\lambda(t)$ is the average rate of earthquakes per unit time occurring around time t since the mainshock, and K, c, and p are parameters (Utsu et al., 1995). The Gutenberg-Richter (G-R) relation characterizing the distribution of earthquake magnitudes can be written in the form:

$$\log_{10}\{1 - F(M)\} = a - bM, \tag{2}$$

where $F(M)$ is the cumulative distribution function at magnitude M, and a and b are parameters. Several different measures of earthquake magnitude exist, and several variants of the G-R law have been proposed, e.g. Kagan (1994), Main (1996), Jackson and Kagan (1999), Kagan (1999), Utsu (1999), Vere-Jones et al. (2001), Kagan and Schoenberg (2001). Alternatives to the modified Omori law have been suggested as well (see Section 12 of Utsu et

al., 1995). In general, however, both of these forms (1) and (2) have been shown to fit rather well to data and are commonly accepted at least as satisfactory first-order approximations for many seismic catalogs.

In addressing the problem of characterizing the typical aftershock sequence, one may refer to the parameter estimates in an ETAS model fitted to an earthquake catalog, for example, and may obtain simulated aftershock sequences using this model. While the parametric approach may be useful for simulating aftershock sequences, it begs the question of how to characterize any one such sequence as *typical*, rather than an outlier. In addition, though such parametric approaches may be useful, in many circumstances it may be desirable to use purely non-parametric methods, and the aim here is the development and application of such non-parametric tools. Further, while one may rather trivially obtain a smoothed average of the point patterns to compute an estimate of the mean number of aftershocks of a given size at a given distance in space and time from the mainshock, an alternative characterization involving actual times, locations, and magnitudes of sample events in such a typical aftershock sequence may be desirable. The effort to obtain such a characterization is the primary motivation of the current paper, and the distance metrics employed are also used to address various other elementary issues, such as the identification of atypical aftershock sequences and the organization of mainshock-aftershock sequences into clusters based on their aftershock properties.

The present work builds on the important contributions of Victor and Purpura (1997), who introduced several useful distance metrics for point patterns. Here, we make use of such distance metrics in introducing the notion of a prototype point pattern. Using the

spike time distance of Victor and Purpura (1997) as an example, we illustrate the use of prototypes for summarizing a catalog of aftershock sequences, for characterizing particular sequences as typical or outliers, for comparing various collections of aftershock sequences, and for organizing such collections into clusters.

In Section 2 we describe the dataset used, which is a collection of aftershock sequences taken from the global Harvard catalog of earthquake occurrences. Section 3 briefly describes the distance metrics for point processes established by Victor and Purpura (1997), before formally introducing prototypes and discussing some of their properties. These prototypes and distance metrics are used in Section 4 for the description of some key features in the global earthquake dataset. Section 4 also addresses sorting the aftershocks into clusters based on their proximity according to the selected Victor-Purpura distance metric, as well as an assessment of the 3-dimensional prototype for simulations of the ETAS model. A discussion of the results and suggestions for further work are presented in Section 5, followed by an Appendix detailing the relationship between the spike time distance proposed by Victor and Purpura (1997) and the integrated absolute difference between cumulative processes.

2 Global Earthquake Aftershock Data

Modern global earthquake catalogs containing detailed information about moderate and large earthquakes have become available since 1977 (see Dziewonski et al. 2000; Frohlich and Davis 1999; Sipkin et al. 2000, and references therein). Among these global catalogs, the dataset that is most extensive and whose completeness properties are best understood is the Harvard catalog (Dziewonski et al. 2000; Kagan 2003).

The Harvard catalog contains estimates of the times, locations, sizes, and orientations of earthquake focal mechanisms, for 19,822 earthquakes occurring between January 1, 1977 and March 1, 2003. Many of these are moderate to large earthquakes occurring at shallow depths and pose major safety risks to humans and structures (Bolt 1993; Lay and Wallace 1995).

An earthquake's location, time and size can be specified in different ways. We focus here on centroid coordinates (the center of gravity for the seismic moment release) and centroid times (estimated time of rupture of the centroid location) as measures of earthquake locations and times. For earthquake sizes, we use the estimated moment magnitude m_w, which is the logarithm of the scalar seismic moment, the latter indicating the total energy released in the earthquake.

One of the difficulties in analyzing and comparing aftershock sequences is the lack of a clear distinction between a mainshock and an aftershock. There are at present no known objective distinctions between the two types of events, other than their temporal ordering, and no definitive way to distinguish "swarms" of earthquake mainshocks from ordinary mainshock-aftershock sequences. In order to avoid subjectivity in our classifications, and in order to address the central question of what one expects to observe following a given mainshock (including both aftershocks and subsequent mainshocks), we define our aftershock sequences simply as all events occurring within a given space-time window of a mainshock, as described in the subsequent two paragraphs. Thus we do not distinguish here between aftershocks and mainshock swarms.

We first define our collection of mainshocks whose aftershock activity we wish to analyze.

In order to ensure that all such mainshocks are essentially comparable, we limit this collection to shallow events only [those with depth less than 70km, following Kagan (2003)], and to those with m_w between 7.5 and 8.0. Further, in order to ensure the similarity of our mainshocks with respect to prior activity, we remove from our collection of mainshocks those earthquakes with any shallow events with either $m_w \geq 7.5$ within 200km and within the preceding 2 years, or with $m_w \geq 8.0$ within 400km and within the preceding 4 years. These spatial-temporal windows were set based on the proposed bounds of Molchan et al. (1997) for ranges containing the bulk of earthquake aftershock activity.

For each of these mainshocks, we define its aftershocks to be all shallow earthquakes greater than a lower cutoff magnitude of m_w 5.5 occurring within 2 years and within a distance of 100km of a mainshock. These windows are again based on the aftershock ranges on p. 1223 of Molchan et al. (1997). The magnitude cutoff of 5.5 is based on the work of Kagan (2003) and Kagan (2004), who showed that for earthquakes of smaller magnitudes the completeness of the Harvard catalog is questionnable. Kagan (2003) showed that there may also be serious misclassification and missing data problems in the first 0.133 days immediately following a major earthquake, so we also restrict our aftershocks to only those occurring at least 0.133 days after its associated mainshock.

Thus defined, there are 49 observed mainshocks in all, each with an average of 5.47 aftershocks and a standard deviation of 4.3 aftershocks. The times of the aftershocks in each sequence are shown in logarithmic scale in Figure 1. One sees that many (13.8%) of the aftershocks occur within the first day after a mainshock, the majority (54.1%) within the first 30 days, and most (86.2%) within the first year.

Figure 2 shows a histogram of the relative times of the aftershocks. The solid curve overlaid is the modified Omori law fitted to the data by maximum likelihood. The fitted parameter estimates are $K = 36.02$ events/day, $c = 0.5721$ days, and $p = 0.9870$.

3 Point Process Distances and Prototypes

Given a collection $\{X_i; i = 1, 2, \ldots, n\}$ of point patterns, one may define its *prototype* as a point pattern Y minimizing the sum

$$\sum_{i=1}^{n} d(X_i, Y), \tag{3}$$

where d is some distance function, that is $d(X, Y)$ is the distance between the two point patterns X and Y. Many choices of d are possible. The spike time distance, used successfully in the description of neuron firings by Victor and Purpura (1997), defines $d(X, Y)$ as the minimal cost needed to transform point pattern X into the pattern Y using a series of elementary operations. These elementary operations include adding a point to X, which is given some cost p_a, deleting a point from X, which is given cost p_d, and moving a point of X by some amount of time Δ, which is given a cost of $p_t\Delta$. Note that p_a must equal p_d in order for d to be a symmetric distance function. Provided $p_a = p_d$ and $p_a, p_t \geq 0$, d is a well-defined distance metric (Victor and Purpura, 1997). This spike time distance extends readily to the case of multi-dimensional point patterns, where in addition to a cost p_t associated with a unit move in time, there may be, for instance, costs p_s and p_m associated with a unit move in space or magnitude, respectively.

This spike time distance metric is not unlike the Earth Movers Distance (EMD) described by Rubin et al. (2000) in that it measures the amount of movement necessary to transform

one random collection into another. Computation of EMD is closely related to the solution of Hitchcock (1941) to the classic transportation problem first discussed by Monge (1781) where of interest was the best way to ship goods from suppliers to consumers while satisfying varying levels of consumer demand. The key distinctions between spike time distance and EMD are that the latter is typically used to determine the distance between histograms of observations rather than the observations themselves, which are typically pixellated values recorded in discrete time rather than points in continuous time, and that the spike time distance allows the addition and deletion of points whereas EMD does not. Indeed, if addition and deletion are not allowed (or equivalently, if the addition and deletion penalties are prohibitively large), then the spike time distance is essentially the same as the integrated difference between the cumulative functions associated with the two temporal point patterns (see Appendix).

Some problems of how to determine the spike time distance between two point patterns and how to find the prototype for a collection of point patterns are discussed in Tranbarger and Schoenberg (2004). We list here two key facts facilitating the solution to these problems. The first is that the spike time distance between two temporal point patterns X and Y is simply given by the sum of the penalties associated with adding or deleting points, plus the sum $\sum_{j=1}^{J} p_t|x_{(j)} - y_{(j)}|$, where $\{x_{(1)}, \ldots, x_{(J)}\}$ and $\{y_{(1)}, \ldots, y_{(J)}\}$ are the *sorted* remaining points of X and Y, respectively. Hence for purely temporal point patterns, the problem of determining which point of X gets moved to which point of Y is not an issue, and finding the distance between X and Y simply amounts to deciding which points are added or deleted and which are moved. For multi-dimensional point patterns, no such obvious shortcut exists, however, but an algorithm that attempts all possible additions, deletions, and pairings of

the points in X and Y is often not prohibitively slow. The second fact worth mentioning is that, for a given collection of multi-dimensional point patterns, using the spike time distance, there exists a prototype for which each coordinate of each point in the prototype is equal to a coordinate of a point in one of the point patterns in the collection. Therefore one way of finding a collection's prototype is by searching over all combinations of coordinates of points in the collection, or by using the step-wise approach suggested in Tranbarger and Schoenberg (2004), which is the method used here.

As noted in Victor and Purpura (1997), one of the main difficulties inherent in metric analyses of point patterns is that of determining the penalties for adding, deleting, and moving points. Note that only the *relative* values of these parameters are relevant for the determination of the prototype. The ratio of the adding/deleting penalty p_a to the moving penalty p_t can have a large impact on the resulting prototypes.

In general, the prototype contains a point t if an aftershock is present near time t in a fraction of at least $p_a/(p_a + p_d)$ of the mainshocks in the dataset. Indeed, if a given point pattern has no aftershock in the interval $(t - p_a/p_t, t + p_a/p_t)$ after the mainshock, then the occurrence of a point in the prototype at time t results in an increase by an amount p_a in the distance from this point pattern to the prototype. On the other hand, for each sequence with an aftershock near time t after the mainshock, the lack of a point near time t in the prototype causes the distance from the point pattern to the prototype to increase by the amount p_d. The prototype will contain a point at t if it is more economical to do so; ignoring moving penalties for the moment this means if $(n - n_t)p_a \leq n_t p_d$, where n_t is the number of sequences with an aftershock near time t after the mainshock. It follows that the prototype

will contain only those points where the fraction (n_t/n) of point patterns in the dataset that contain nearby points is at least $p_a/(p_a + p_d)$.

The fact above can be helpful in determining appropriate values of p_a and p_d. If $p_a = p_d$, then the above ratio is 1/2, but it may be of interest to include points in the prototype even though nearby points may occur in less than half the point patterns in the dataset. If one desires that the prototype Y should represent all aftershocks occurring in at least some proportion p of the aftershock sequences in the dataset, then one may set the ratio of p_a to p_d to $p/(1-p)$. The prototype may still be defined as before, i.e. as the point pattern Y minimizing the sum $\sum_i d(X_i, Y)$; the fact that the function d is no longer symmetric and thus not a proper distance function is immaterial.

Alternatively, if a symmetric distance function is desired, then the prototype can contain no more points than the median number of points in the collection. If one seeks a prototype that is typical in terms of its length, then the moving penalties must be chosen to be miniscule in order to ensure that the prototype's length is close or equal to the median length of the point patterns.

In addition to the determination of prototypes, the (possibly non-symmetric) distance function d can also be used for identification of outliers or clusters. We propose classifying certain individual earthquake sequences as outliers if their distance from the prototype is unusually large. Further, subgroups of earthquake sequences can be classified into clusters based on the principle of minimizing the total distance from the earthquake sequences to their respective cluster prototypes. Note that for determining clusters and distances to the prototype, tiny moving penalties are no longer required and in fact comparably-sized adding

and moving penalties are preferable; these issues are discussed further in Tranbarger and Schoenberg (2004).

4 Results.

4.1 Prototypes of the Harvard Aftershock data

Figure 3 shows the times of events in the prototype for the dataset of aftershock times described in Section 2. One may interpret the prototype in Fig. 3 roughly as representing the times of the collection of aftershocks occurring in the majority of aftershock sequences. For the dataset considered, the times since the mainshock were recorded on a logarithmic scale, as they are shown in Fig. 1. The distance function thus assigns more weight to a move of one hour for a point occurring near the time of the mainshock than for a point occurring long after its mainshock. As noted in the previous Section, only the *relative* values of the penalty parameters matter in determining the prototype, so hereafter we fix the adding penalty p_a to unity. The prototype shown in Fig. 3 was made using a symmetric distance function ($p_a = p_d = 1$) and a small moving penalty p_t of 1/15. Note that the prototype has five points, which is equivalent to the median number of aftershocks in the Harvard dataset. The prototype in Figure 3 may be said to represent a *typical* aftershock sequence in this dataset.

Figure 4 shows a histogram of the distances from each of the point patterns to the prototype in Fig. 3, using moving penalty p_t, adding penalty p_a, and deleting penalty p_d all equal to unity. One sees that there are three outliers with abnormally large distances to

the prototype of over 10, while by contrast, most of the other aftershock sequences have a distance to the prototype of 9 or less. Collectively the mean distance is 5.7 and the standard deviation is approximately 2.4.

Figure 5 shows the five aftershock sequences closest to the prototype, in terms of the spike time distance described in Section 3, with all penalties set to unity. Several of the point patterns are strikingly similar to the prototype. Interestingly, two of the five closest to the prototype lie in the Santa Cruz Islands. Figure 6 shows the five outlier sequences, i.e. those with greatest distance from the prototype, again with unit penalties. Not surprisingly, these include sequences containing many large aftershocks, such as the New Ireland 2000 and Kermadec Island 1986 aftershock sequences. These are the only two point patterns whose distances from the prototype are greater than 12, and these two aftershock sequences are highly atypical in the number and configuration of the sizeable aftershocks they contain.

As noted in Section 3, the distance function used to define a prototype may contain penalties not only for aligning the times of points, but also for aligning their magnitudes and spatial distances from their corresponding mainshocks. Using moving penalties for log-time-since-mainshock, spatial distance from mainshock, and magnitude relative to mainshock magnitude, one obtains a time-space-magnitude prototype for the 49 aftershock sequences; this three-dimensional prototype is depicted in Figure 7. Of note is the prototype's relatively large aftershock around time 77 days after the mainshock. Among the other four aftershocks, there is a slight, gradual decline in the relative moment magnitudes of the aftershocks as the temporal and spatial distances from the mainshock increase.

There is no guarantee that the times of the aftershocks in the one-dimensional prototype

of Figure 3 will be identical to those of the three-dimensional prototype in Figure 7. Indeed, the two sets of times are not equivalent, though the differences are far from dramatic.

4.2 Division of earthquake sequences into clusters

Given a distance function such as the spike time distance with appropriate penalty parameters, one may subdivide the earthquake sequences into clusters, so that within each cluster, each point pattern is optimally close to the prototype of its cluster. That is, one may divide the point patterns into k clusters so that the sum of the distances from the sequences to their cluster prototypes is minimized.

Obviously, when each aftershock sequence is its own cluster this sum is zero, so a parsimonious choice of k is required. Of the 49 aftershock sequences in the Harvard catalog, three contained no aftershocks at all. Figure 8 shows the division of the other 46 aftershock sequences in the Harvard catalog into $k = 2$ clusters. A monothetic agglomerative method (see Späth, 1980) is well suited to the task of dividing the point patterns into groups, as such a method is able to assign point patterns into clusters without computationally intensive intermediate calculations of three-dimensional prototypes. The assignment shown in Figure 8 proceeded through each of the 46 point patterns in order from most aftershock activity to least. At each step, the aftershock sequence of interest was paired with the aftershock sequence minimizing the standardized sum of spike time distances along the time, magnitude, and distance axes. Once two point patterns were matched in this way, they were replaced in the algorithm with their cluster's prototype. This allows any future point pattern of interest to be assigned to a cluster based on its distance to the prototype of a previously formed

cluster, or to an individual aftershock sequence that has not been assigned to a cluster. At each stage, the number of clusters present is reduced by one until all sequences are assigned to one of only two clusters.

As can be seen from Figure 8, there is some geographic clustering present in the results of clustering the aftershock sequences into two groups based on prototypes of aftershock activity. For instance, most (though not all) of the mainshocks in the Southern Indonesian seismic zone have aftershocks belonging to cluster one (circles), whereas the earthquakes in Northern Indonesia and those along the South American plate all tend to fall in cluster two (asterisks). The vast majority of the shallow mainshocks in the Harvard catalog occur in subduction zones along major plate boundaries, but several of the exceptions (e.g. India, Turkey) have aftershock activity placing them in cluster one.

One of the main differences between the aftershock sequences in cluster one (circles) and those in cluster two (asterisks) seems to be that the distances from the mainshock are much more evenly dispersed between 0 and 100 km from the mainshock source for the aftershocks in cluster one, whereas in cluster two the vast majority of the aftershocks occur between 40 and 80 km from the mainshock. Another key distinction is that cluster two tends to have mainshocks with more aftershock activity observed compared with cluster one. Indeed, the difference between the aftershock activities in different regions may be due to differing geological characteristics, or instead may simply be due to differential detection rates of aftershocks in different regions. While the Harvard catalog is perhaps the world's best dataset in terms of completeness, it may nevertheless have serious missing data issues, particularly in regions far from major subduction zones (Kagan, 2003).

4.3 Prototype of the ETAS model

In addition to providing useful summaries of aftershock datasets, prototypes can also be used to summarize models for aftershock sequences. For instance, one may obtain a typical realization of the ETAS model of Ogata (1998) by simulating the model many times, and determining the prototype of these simulated ETAS processes.

Figure 9 shows the prototype applied to data simulated according to the ETAS model, but with mainshocks identical to those in our subset of the Harvard catalog, described in Section 2. That is, for each of the 49 mainshocks, we simulated its aftershocks according to the ETAS model of Ogata (1998), using the fitted parameters in Ogata (1998). The ETAS model specifies the conditional intensity (see Daley and Vere-Jones, 2003) of points around time t and location $\mathbf{x}$:

$$\begin{aligned} \lambda(t,\mathbf{x}) &= \mu(\mathbf{x}) + \int_0^t \int_{\mathbf{x}} \int_{m_0}^{m_1} g(t-t', ||\mathbf{x}-\mathbf{x}'||, m') dN(m', \mathbf{x}', t') \\ &= \mu(x,y) + \sum_{i:t_i<t} g(t-t_i, ||\mathbf{x}-\mathbf{x}_i||, m_i), \end{aligned} \tag{4}$$

with the aftershock triggering density g specified as

$$g(t,x,y,m) = \frac{K_0 exp\{\alpha(m-m_0)\}}{(t+c)^p (x^2+y^2+d)^q}, \tag{5}$$

in agreement with the modified Omori law (1). The parameters used in the simulations are those fitted by Ogata (1998) to Central and Western Honshu Island earthquakes of magnitude at least 5.5 ($\mu = 6.69 \times 10^{-5}$, $\kappa = 3.42 \times 10^{-4}$, $c = 3.19 \times 10^{-3}$, $\alpha = 0.955$, $p = 0.944$, $d = 9.55 \times 10^{-3}$, $q = 1.506$). As with the Harvard and W. Honshu datasets, we limit our attention to simulated aftershocks of magnitude at least 5.5. The ETAS model does not specify the magnitude distribution of the simulated events, as these magnitudes are ordinarily

assumed to be separable with respect to the rest of the process (Ogata, 1998), though this assumption is invalidated in Schoenberg (2003). Here, we simply take the simulated aftershock magnitudes to be sampled with replacement from the empirical distribution of aftershock magnitudes from our subsample of the Harvard dataset described in Section 2.

The ETAS model predicts somewhat fewer aftershocks, on average, than exist in the Harvard catalog, and this difference is easily detected in the comparison of the ETAS prototype in Figure 9, which has only three points, with the prototype of the Harvard aftershocks in Figure 7, which has five points. The difference between the simulated ETAS processes and the observed Harvard aftershocks could well be due to particularities in the Central and Western Honshu aftershocks used to determine the fitted parameters. Indeed, the simulated ETAS processes contained a median of just three points, and the spatial distances of the points in the simulated ETAS prototype from the mainshock are much smaller than the corresponding distances in the prototype for the Harvard catalog. Note that in the simulated ETAS catalog, the slight increase in the magnitudes of the points with time since mainshock is entirely coincidental, since the magnitudes of the simulated points are simply sampled at random from the empirical distribution of aftershock magnitudes. Indeed, we generated other iterations of the simulated ETAS catalog, and in subsequent simulations such a trend did not persist, though the other features of the prototype were the same, including the small number of points.

It is important to note that the simulated ETAS catalog was generated with the parameters specified by Ogata (1998), who estimated these parameters by fitting them by maximum likelihood to the Honshu catalog. The differences between the prototype of the simulated

ETAS catalog and that of the Harvard catalog are most likely attributable to the somewhat atypical features of the Honshu dataset, such as the unusually low occurrance rate of aftershocks, and the fact that these Honshu aftershocks are distributed abnormally tightly, spatially, around the mainshocks.

5 Discussion.

In addressing the motivating question given in Section 1, the prototype in Figure 7 may be described as a typical aftershock sequence for a shallow mainshock of magnitude 7.5 to 8.0. That is, among such mainshocks with no major precursory events of similar magnitude within the previous few years, a typical aftershock sequence contains five aftershocks of magnitude at least 5.5, and these aftershocks are roughly evenly distributed in log-time between 0.113 days and 2.0 years after the mainshock.

However, the selection of this prototype is to some extent dependent upon the rather arbitrary choice of penalties in the multi-dimensional version of the spike time distance function. If the ratios of the moving penalties are changed, then the space-time-magnitude locations of the points in the prototype will change somewhat, and if the penalties are increased sufficiently relative to the adding/deleting penalty p_a, then the prototype will contain fewer points than the median point pattern length of five. It should also be noted that the three-dimensional prototypes we constructed were computed based on the step-wise approximation technique described in Tranbarger and Schoenberg (2004), and hence may not be exact.

In addition, the Harvard catalog has several problems, including time and location errors,

completion problems, and other issues, and although these are relatively minor compared to other earthquake catalogs, they are far from negligible (Dziewonski et al., 2000; Kagan, 2003). In particular, in the period immediately following a mainshock [0 to 0.133 days according to Kagan (2003)], it is extremely difficult for seismometers to distinguish between small aftershocks and repeated shaking due to the seismic waves emitted from the mainshock. We have omitted such seismicity entirely from our analysis, but one may argue that we have neglected important aftershock activity from our prototypes as a result. In particular, differential detection rates in different seismic zones could have a strong influence on the results of grouping the aftershock sequences into clusters. In addition, in this analysis we have merged aftershock sequences from different fault types in determining the overall prototype for the Harvard catalog. If aftershock sequences in these different zones are so disparate that the aftershock sequences are essentially incomparable, then the meaning and interpretability of the prototype may be questioned, and further study may be necessary to determine the prototypes for each relevant grouping of seismic zones.

We have shown how prototypes may be used as summaries not only of point pattern datasets but of models, and have included as an illustration a characterization of the prototype of the ETAS model, which may be compared to the prototype of the Harvard dataset. As noted in the previous Section, however, differences between the two prototypes may be due to unusual aspects of the Honshu dataset to which the parameters in the ETAS model were fitted. The magnitude cutoffs used by Ogata (1998) are essentially the same as those used in our analysis of the Harvard catalog. However, in both cases it should be noted that the lower magnitude cutoff is a critical component of the analysis. If aftershocks below

moment magnitude 5.5 are included, the resulting simulated ETAS processes and Harvard aftershock sequences will have substantially more events, as will the resulting prototypes of the two catalogs.

The spike time distance attributed to Victor and Purpura (1997) is only one of many possible distance metrics for point patterns, and indeed several others are proposed in Victor and Purpura (1997). In introducing the notion of a prototype, we have focused on the spike time distance here because of its simplicity and interpretability. The investigation of results using other types of distance functions and the robustness of these results to choice of distance function are important subjects for future research.

6 Appendix.

Suppose X and Y are two temporal point patterns on $[0, T]$ such that X contains points $x_1, \ldots, x_m$ and Y contains points $y_1, \ldots, y_n$, with $m \leq n$. Viewing $X(t) = \max\{i : x_i \leq t\}$ and $Y(t) = \max\{i : y_i \leq t\}$ as functions, we may investigate conventional distances between the two point patterns as in the following result.

Theorem 1.

$$\int_0^T |X(t) - Y(t)|dt = \sum_{i=1}^{m} |x_i - y_i| + \sum_{i=m+1}^{n} (T - y_i). \tag{6}$$

Proof.

Let $a_i = \max\{x_i, y_i\}$, for $i = 1, \ldots, m$. Let $a_i = T$ for $i = m+1, m+2, \ldots, n$, and $a_0 = 0$.

Similarly let $b_i = \min\{x_i, y_i\}$, for $i = 1, \ldots, m$. Let $b_i = y_i$ for $i = m+1, m+2, \ldots, n$, and $b_0 = 0$.

Let $c_{i,j} = y_j$ if $a_i = x_i$, and $c_{i,j} = x_j$ if $a_i = y_i$. Thus $c_{i,i} = b_i$, for $i = 1, \ldots, m$. (In the case where $x_i = y_i = a_i$, the value of $c_{i,j}$ is irrelevant.)

With this notation, we may write

$$\begin{aligned}
\int_0^T |X(t) - Y(t)|dt &= \sum_{i=0}^{m} \int_{a_i}^{a_{i+1}} |X(t) - Y(t)|dt \\
&= \sum_{i=1}^{n} \sum_{j \geq i : c_{i,j} \leq a_i} (a_i - max\{c_{i,j}, a_{i-1}\}) \\
&= \sum_{i=1}^{m} [(a_i - a_{i-1}) + (a_{i-1} - a_{i-2}) + \ldots + (a_{i-k} - c_{i,i})] + \sum_{i=m+1}^{n} (T - c_{i,i}). \\
&= \sum_{i=1}^{m} (a_i - b_i) + \sum_{i=m+1}^{n} (T - b_i). \qquad \square
\end{aligned}$$

Recall that, assuming the addition penalty p_a is equivalent to the penalty for deletion, the spike time distance between X and Y is $p_a(n + m - 2J) + \sum_{j=1}^{J} p_t|x_{(j)} - y_{(j)}|$, where $\{x_{(1)}, \ldots, x_{(J)}\}$ and $\{y_{(1)}, \ldots, y_{(J)}\}$ are the *sorted* points of X and Y, respectively, that are not added or deleted (Tranbarger and Schoenberg, 2004). Therefore an immediate consequence of the Theorem above is that if the two sequences X and Y have the same length, i.e. if $m = n$, and if $p_a = p_d = \infty$ so that no points may be added or deleted, then the integrated absolute difference between the two processes X and Y is proportional to the total spike time distance between the two point patterns, and the proportionality constant is simply the moving penalty, p_t. Hence if p_t is very small compared to the addition/deletion penalty, so that addition and deletion are essentially not permitted, then the total distance between two point patterns of equal length is simply p_t times the integrated absolute difference between the two processes.

7 References

Bolt, B. (1993). *Earthquakes, 2nd ed.* Freeman, New York.

Daley, D., and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods, 2nd ed.* Springer-Verlag, New York.

Dziewonski, A.M., Ekström, G., and Maternovskaya, N.N. (2000). Centroid-moment tensor solutions for July-September 1999. *Phys. Earth Planet. Inter.* **119**, 311–319.

Frohlich, C., and Davis, S.D. (1999). How well constrained are well-constrained T, B, and P axes in moment tensor catalogs? *J. Geophys. Res.* **104**, 4901–4910.

Hitchcock, F.L. (1941). The distribution of a product from several sources to numerous localities. *J. Math. Phys.*, **20**, 224–230.

Jackson, D.D., and Kagan, Y.Y. (1999). Testable earthquake forecasts for 1999. *Seism. Res. Lett.* **70**(4), 393–403.

Kagan, Y.Y. (1994). Observational evidence for earthquakes as a nonlinear dynamic process. *Physica D* **77**, 160–192.

Kagan, Y.Y. (2003). Accuracy of modern global earthquake catalogs. *Phys. Earth Planet. Inter.* **135**(2-3), 173–209.

Kagan Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bull. Seismol. Soc. Amer.* **94**(4), 1207–1228.

Kagan, Y.Y. and Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *J. Appl. Prob.* **38A**, Supplement: Festscrift for David Vere-Jones, D. Daley, editor, 158–175.

Lay, T., and Wallace, T.C. (1995). *Modern Global Seismology.* Academic Press, San

Diego.

Main, I.G. (1996). Statistical physics, seismogenesis, and seismic hazard. *Rev. Geophys.* **34**, 433–462.

Molchan, G., Kronrod, T., and Panza, G. F. (1997). Multi-scale seismicity model for seismic risk. *Bull. Seismol. Soc. Am.* **87**, 1220–1229.

Monge, G. (1781). *Mémoire sur la Théorie des Déblais et des Remblais.* Histoire de l'Académie Royale des Sciences, Paris.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**(401), 9-27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**(2), 379-402.

Schoenberg, F. (2003). Multi-dimensional residual analysis of point process models for earthquake occurrences. *J. Amer. Statist. Assoc.* **98**(464), 789–795.

Sipkin, S.A., Bufe, C.G., and Zirbes, M.D. (2000). Moment-tensor solutions estimated using optimal filter theory: global seismicity, 1998. *Phys. Earth Planet. Inter.* **118**, 169–179.

Späth, Helmuth (1980). *Cluster Analysis Algorithms for Data Reduction and Classification of Objects.* E. Horwood, Chichester and Halsted Press, New York.

Tranbarger, K., and Schoenberg, F.P. (2004). On the computation and application of prototype point patterns. *UCLA Statistics Preprints* **406**, 1–7.

Utsu, T. (1999). Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure Appl. Geophys.* **155**, 509–535.

Utsu, T., Ogata, Y., and Matsu'ura, R.S. (1995). The centenary of the Omori formula

for a decay law of aftershock activity. *J. Phys. Earth* **43**, 1–33.

Vere-Jones, D., Robinson, R., and Yang, W.Z. (2001). Remarks on the accelerated moment release model: problems of model formulation, simulation, and estimation. *Geophys. J. Int.* **144**, 517–531.

Victor, J. and Purpura, K. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Comput. Neural. Syst.***8**, 127–164.

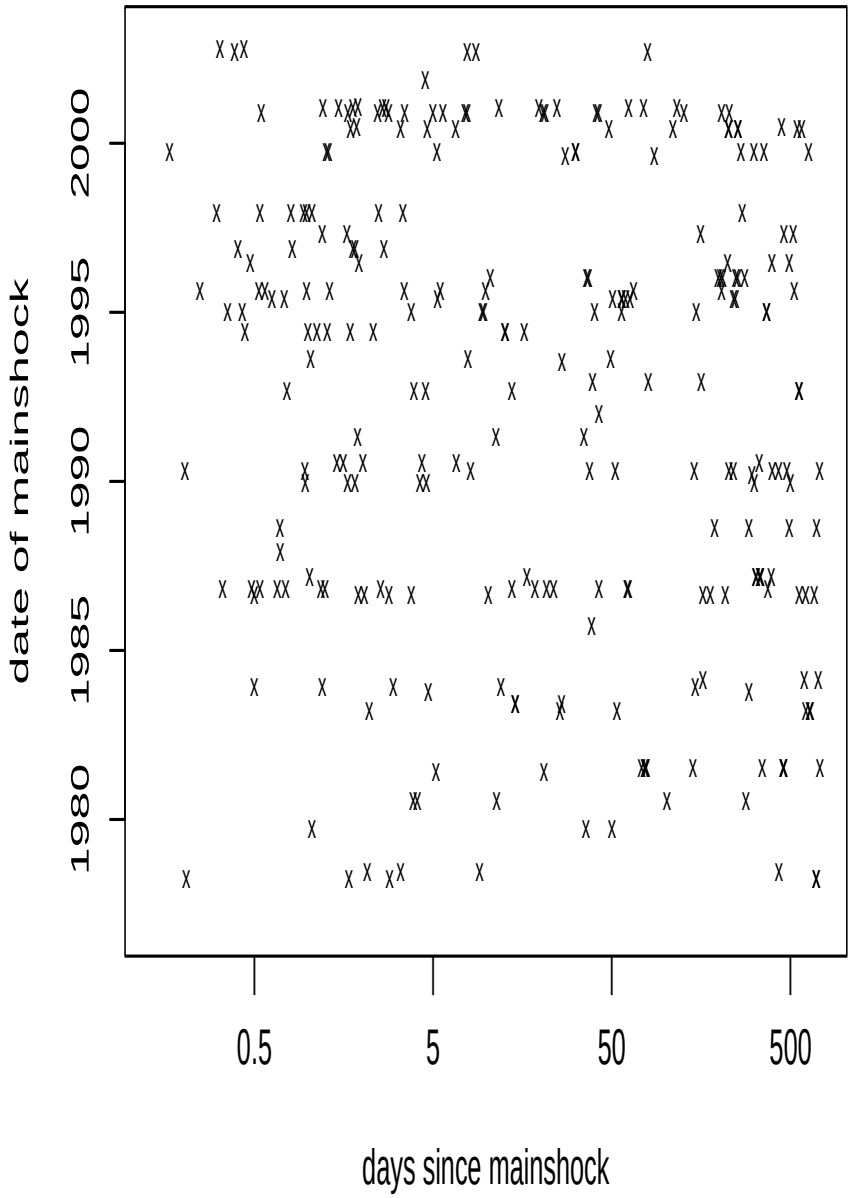

Figure 1: Aftershock times.

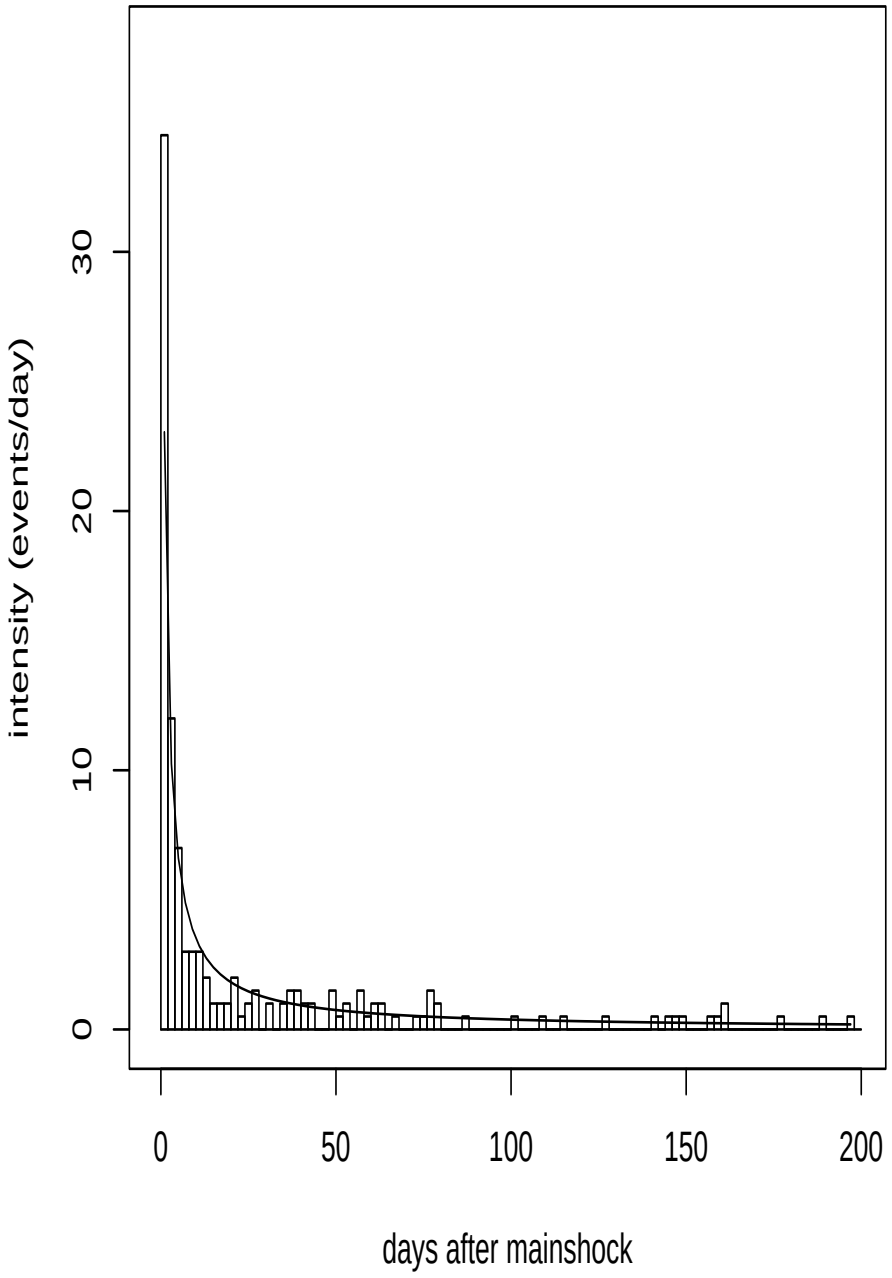

Figure 2: Fit of modified Omori law to global aftershock data.

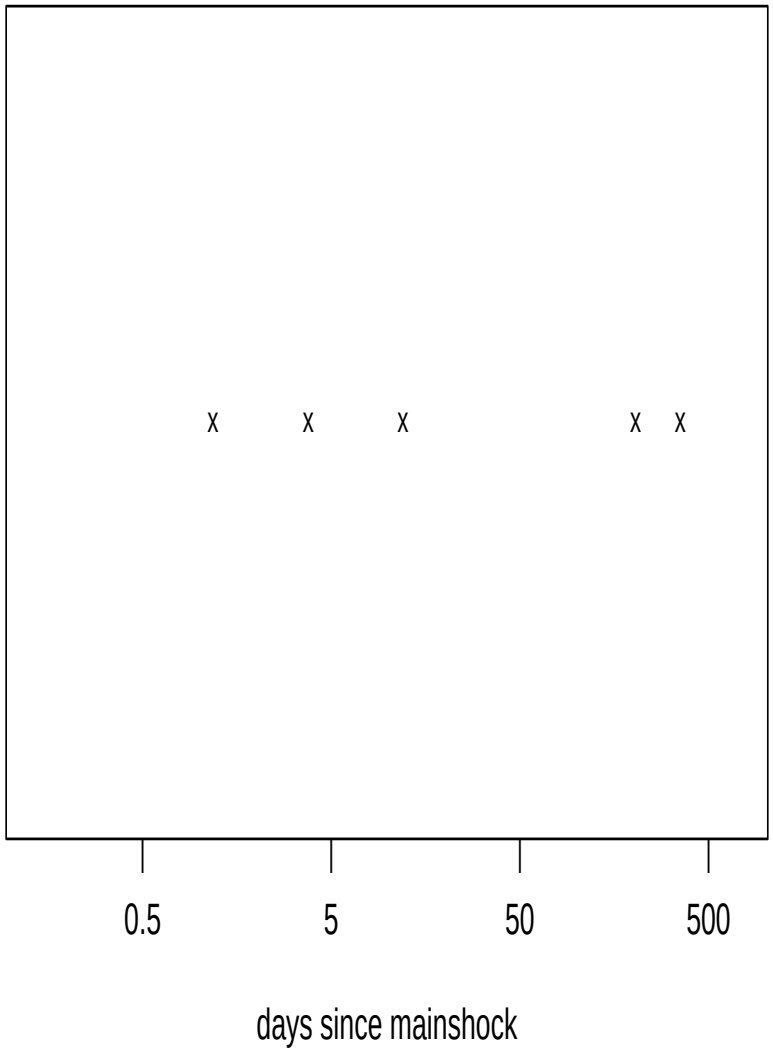

Figure 3: Prototype temporal aftershock sequence.

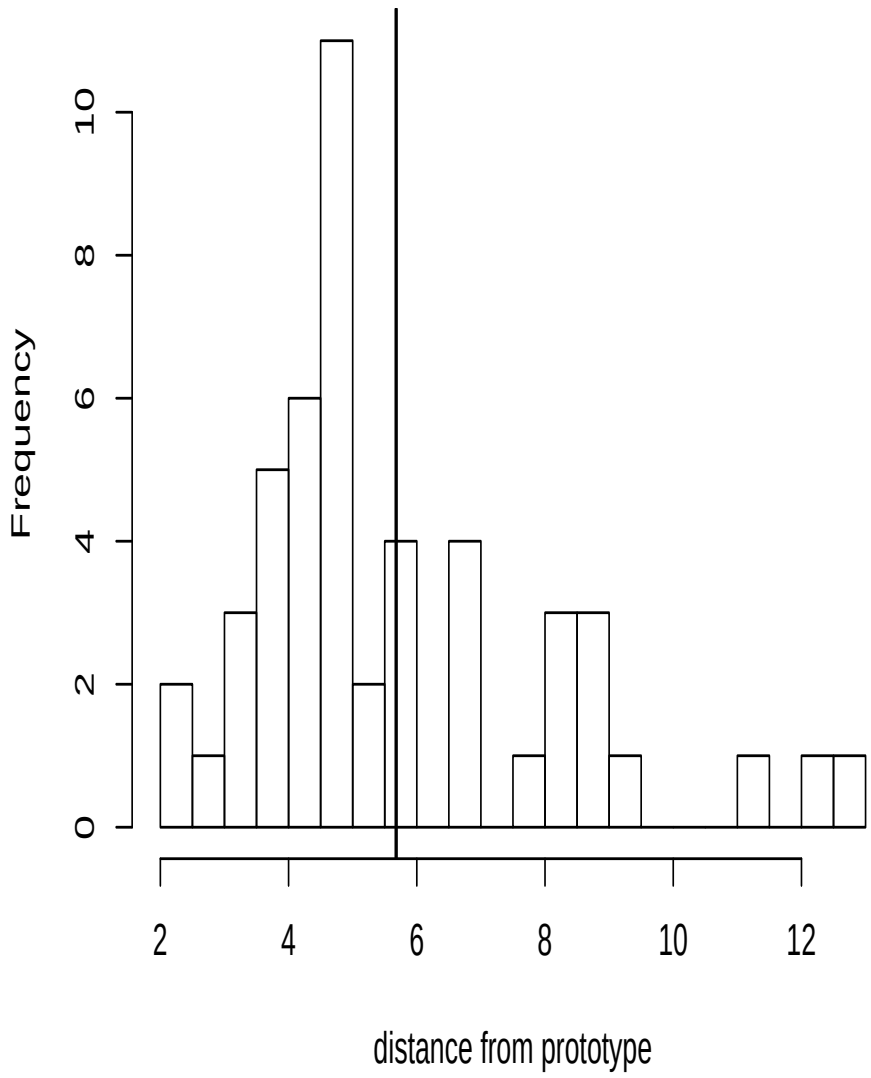

Figure 4: Histogram of distances from prototype.

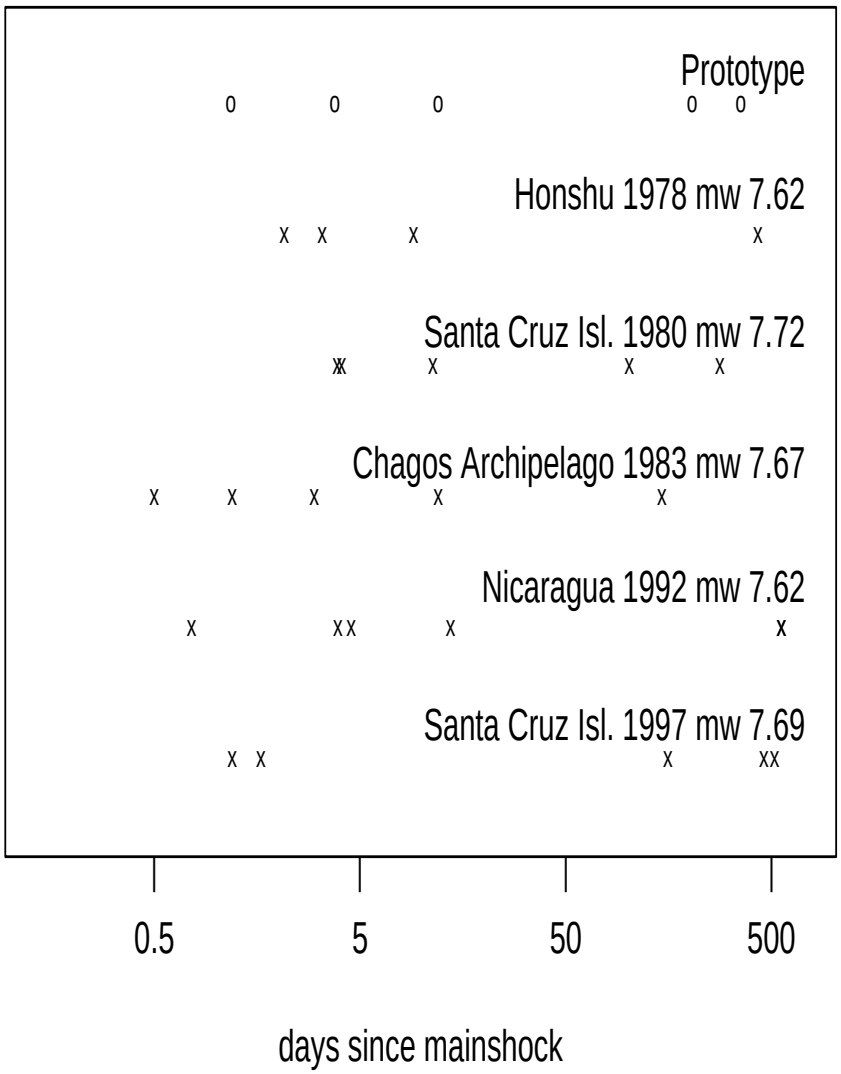

Figure 5: Five aftershock sequences closest to prototype.

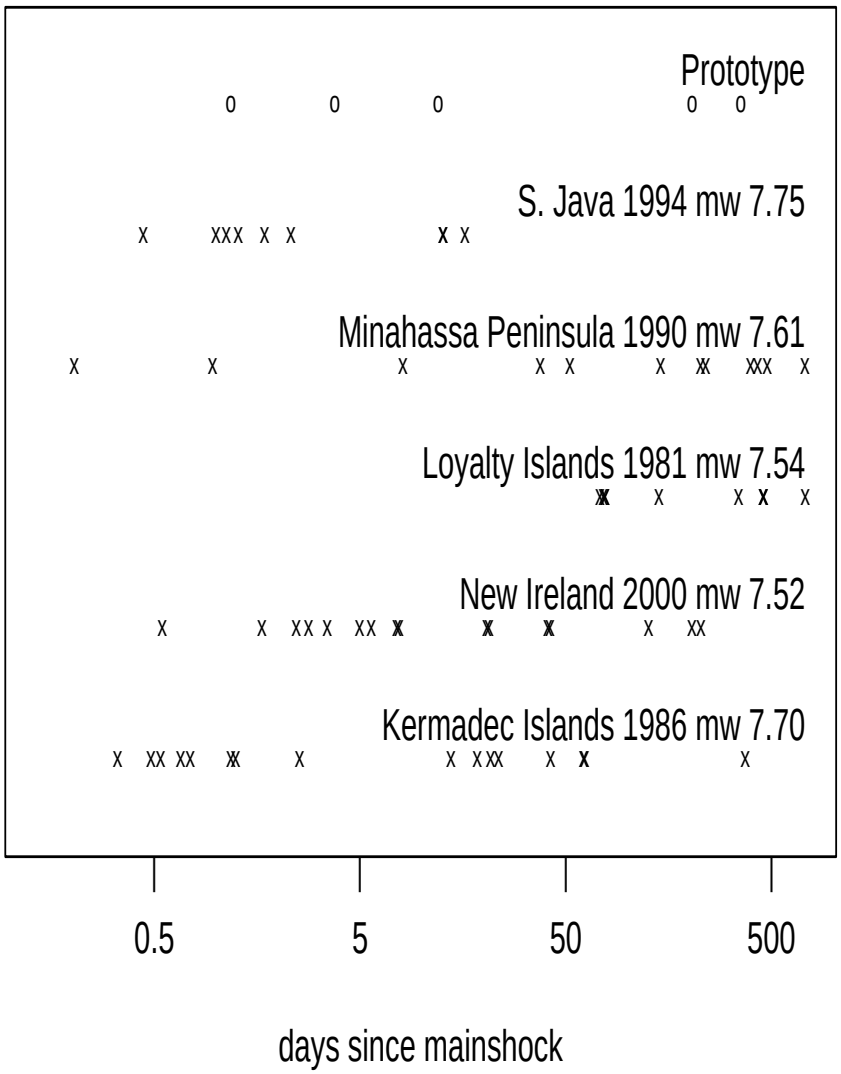

Figure 6: Five aftershock sequences furthest from prototype.

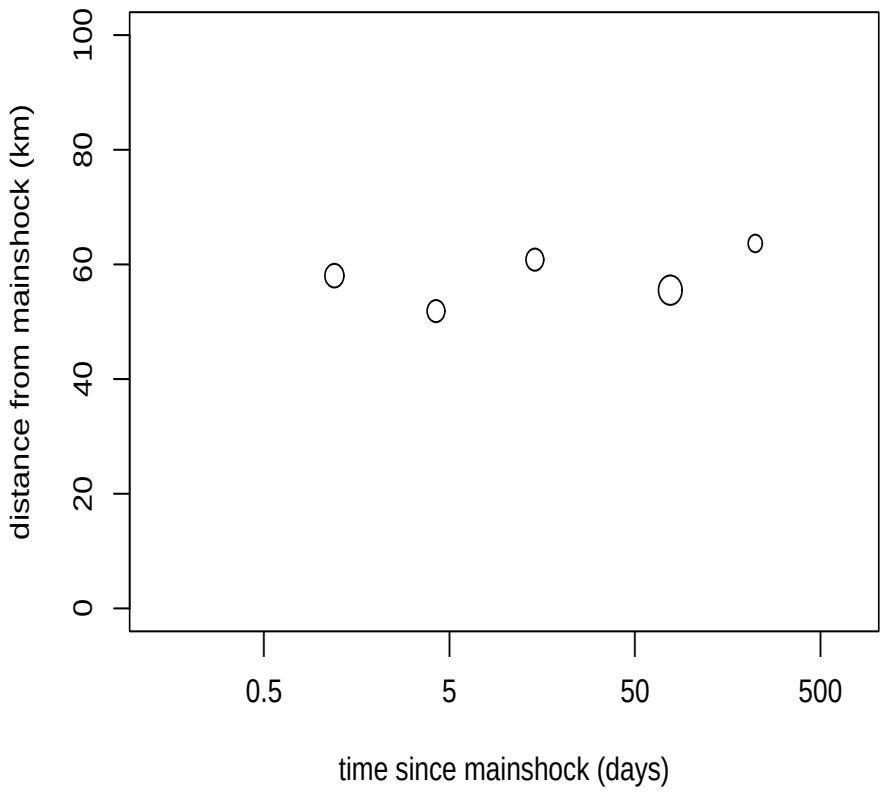

Figure 7: Prototype time-space-magnitude aftershock sequence. The circle sizes indicate the relative size of the aftershock: the relative values of m_w for the prototype points (listed in order of time since mainshock) are 76.5%, 76.0%, 76.0%, 78.5%, and 74.4% of the mainshock magnitude. The adding/deleting penalty used is 1.0, and the moving penalties are 0.01, 0.0003, and 0.3, respectively, for the log-time-since-mainshock, spatial distance from mainshock, and moment magnitude relative to that of the mainshock. These were chosen so that maximal moves in each coordinate are given roughly equivalent weight in the distance function.

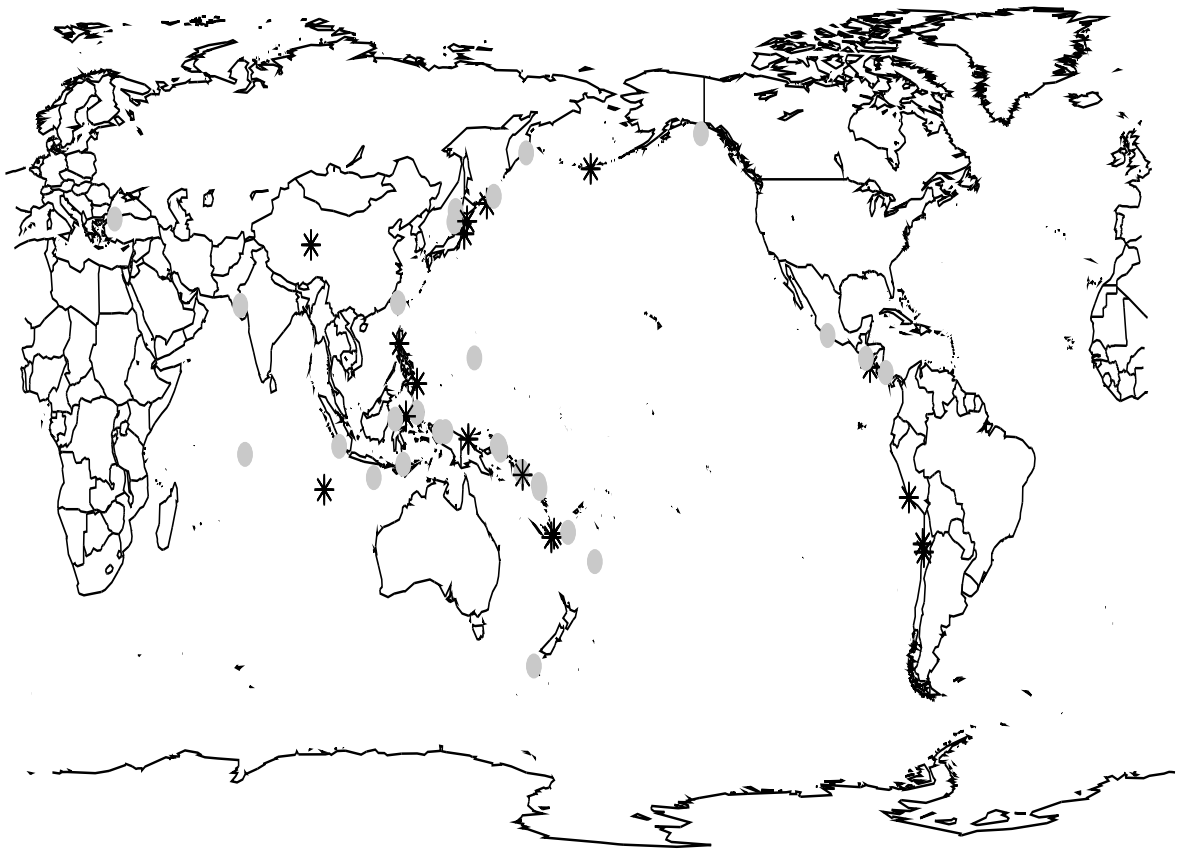

Figure 8: Clusters of mainshocks, based on distances to temporal, spatial, and magnitude prototypes.

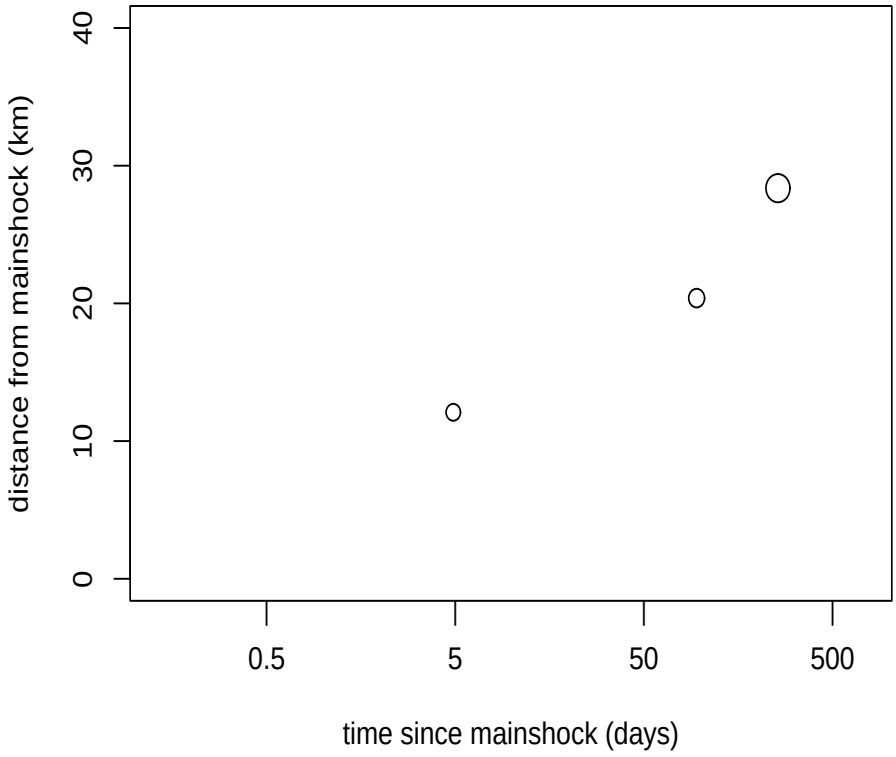

Figure 9: Prototype of simulated ETAS processes, with mainshocks identical to those in the Harvard dataset. Parameters used are those fitted to data in Ogata (1998), and aftershock magnitudes are sampled from the Harvard aftershock data. The adding/deleting and moving penalties in the spike time distance function are the same as those used in the construction of the 3-dimensional prototype for the Harvard catalog ($p_t = 0.01$, $p_s = 0.0003$, and $p_m = 0.3$).