Original Article

Journal of Time Series Analysis

Final version received June 2010

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2011.00734.x

The application of prototype point processes for the summary and description of California wildfires

Kevin Nichols[a], Frederic Paik Schoenberg[a,*,†], Jon E. Keeley[b], Andrew Bray[a] and David Diez[a]

A method for summarizing repeated realizations of a space-time marked point process, known as prototyping, is discussed and applied to catalogues of wildfires in California. Prototype summaries are constructed for varying time intervals using California wildfire data from 1990 to 2006. Previous work on prototypes for temporal and space-time point processes is extended here to include methods for computing prototypes with marks and the incorporation of prototype summaries into hierarchical clustering algorithms, the latter of which is used to delineate fire seasons in California. Other results include summaries of patterns in the spatial-temporal distribution of wildfires within each wildfire season.

Keywords: Prototypes; point processes; marked point processes; spike-time distance; wildfire seasons.

1. INTRODUCTION

Suppose that one observes repeated realizations $N_1, N_2, \ldots, N_m$, of a finite, marked space-time point process. Each realization N_i is a finite collection of points $\{x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}\}$ where n_i is the number of points in realization N_i and any point $x_{i,j} = (t, \mathbf{z}, m)$ occurs in some product space S, the product of temporal, spatial and mark domains. Examples of marked space-time point process data are abundant (for a review, see, e.g. Schoenberg *et al.*, 2002), and include disease outbreaks, births of animals, or the occurrences of other rare events such as earthquakes, lightning strikes, volcanic eruptions, or, as is the focus of our application here, wildfires.

The methods we employ here involve point process prototypes, which in turn rely on distance metrics between two realizations of a point process. Given two such point patterns, a distance between the two can be defined using a variety of different possible distance metrics (Victor and Purpura, 1997; Mateu *et al.*, 2011). Among these possibilities, a particularly simple and important example is the spike-time distance metric introduced by Victor and Purpura (1997) for the description of neuron-cell firings. With the spike-time metric, the distance $D(N_i, N_j)$ between two point patterns N_i and N_j is defined as the minimum total penalty required to transform the points of N_i into those of N_j using three elementary transformations: points may be added to N_i each with a cost or penalty p_a, points may be deleted from N_i each with a penalty p_d and the coordinates of points may be moved with cost proportional to the size of the move. For purely temporal point processes, as described by Victor and Purpura (1997), moving a point t_1 to a new time t_2 would be associated with a penalty $p_m|t_2-t_1|$, and this moving penalty can be extended in obvious ways to the space-time case using $\mathbf{p}_m$, a vector of penalties each corresponding to one coordinate of the domain (see e.g. Schoenberg and Tranbarger, 2008). Figure 1a shows an illustration of the spike-time distance metric. The penalty parameters are typically set by the user and can either be tied to real costs in the applied problem or may be adjusted to achieve a desired number of points in the prototype (Tranbarger and Schoenberg, 2010). In order for the spike-time distance to have the symmetry property requisite of formal distance metrics, p_a must equal p_d. Diez *et al.* (2011) discuss an application using spike-time distance with $p_a \neq p_d$.

Given a collection $\{N_1, \ldots, N_m\}$ of point patterns and cost parameters p_a, p_d and $\mathbf{p}_m$, one may define the *prototype*, P, of the collection as the point pattern minimizing

$$\sum D(N_i, P), \tag{1}$$

as in Schoenberg and Tranbarger (2008). Thus, the prototype is a natural measure of central tendency for collections of point patterns, analogous in many ways to the median of a collection of real numbers. Figure 1b is a representation of fifty one-dimensional point patterns along with the prototype summary for those fifty patterns. The patterns are simulated independent

[a]*Department of Statistics, UCLA*
[b]*U.S. Geological Survey, Western Ecological Research Center and the UCLA Department of Ecology and Evolutionary Biology*
**Correspondence to: Frederic Paik Schoenberg, 8142 Math-Science Building, Department of Statistics, University of California, Los Angeles, CA 90095–1554, USA.*
[†]*E-mail: frederic@stat.ucla.edu*

© 2011 Blackwell Publishing Ltd.

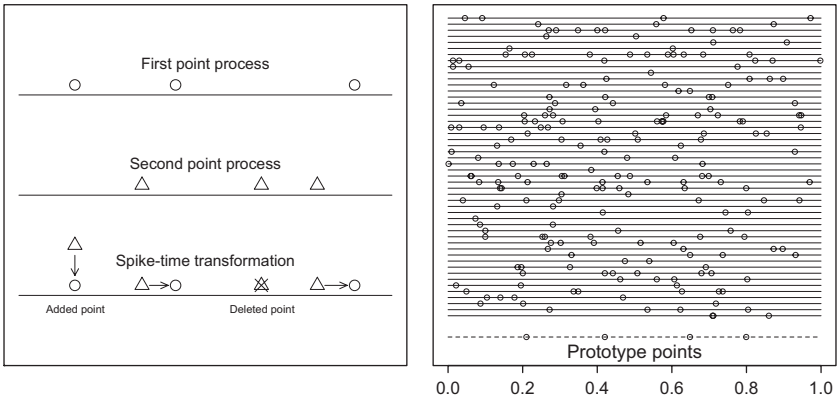

Figure 1. Left: Spike time distance transformation of one point-pattern realization into another. In this example, the second point pattern is transformed into the first. Right: Example of a prototype. The contributing point patterns are 30 simulated independent stationary Poisson processes on the interval [0,1] with rate $\lambda = 4$

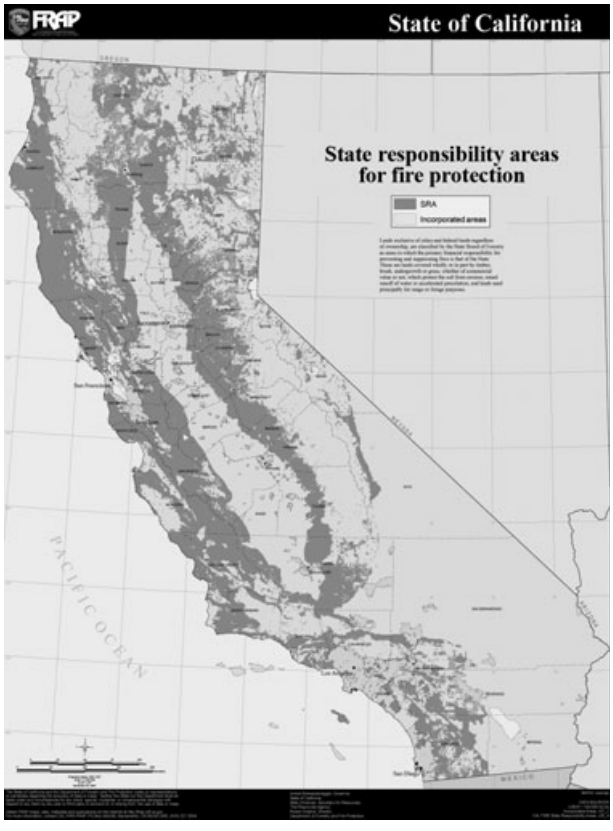

Figure 2. California state responsibility area. The plot is open sourced on the CalFire website, http://frap.cdf.ca.gov

stationary Poisson processes on the interval [0,1] with rate $\lambda = 4$. The resulting prototype contains four points that are nearly equally spaced.

In the case of point processes consisting of the space-time locations where California wildfires occur in a given year, and with wildfire sizes as the marks, prototypes using spike-time distance may provide useful and easily interpretable summaries of central tendency without reliance on a parametric model. Another convenient feature of the prototype is that, unlike other summaries such as kernel smoothings, the prototype is always contained within the spatial support of the data, even when this support is a union of several discrete areas (as in Figure 2). California wildfires naturally present themselves as ideal candidates for multi-dimensional prototyping, because of the lack of agreement on parametric models for their occurrence and due to the high variability of wildfire activity from year to year and season to season, rendering conventional averages of wildfire activity poor descriptions of typical

© 2011 Blackwell Publishing Ltd.

wildfire behaviour. For properties of point process prototypes, including analogies between the prototype and the median for real-valued data, see Schoenberg and Tranbarger (2008) and Diez *et al.* (2011).

Previous summaries of catalogues of wildfire activity have largely been based on primarily temporal methods, including estimating fire rotation intervals or computing fire frequency to size ratios, as well as parametric methods involving the fitting of parametric wildfire models (see, e.g. Johnson and Gutsell, 1994, Keeley *et al.*, 1999, Malamud *et al.*, 2005, Moritz *et al.*, 2005). Unfortunately, temporal summaries that involve the description and/or modelling of each spatial sub-region separately often depend critically on the rather arbitrarily chosen boundaries of the spatial regions, and parametric summaries suffer the further disadvantage of relying on the assumption of a parametric model, which may be questionable. Some alternative non-parametric spatial methods for summarizing catalogues of wildfire activity include simply examining the mean wildfire activity over the given spatial region, or kernel smoothing the data. Such summaries are not easily interpretable and can be quite misleading.

As an illustration, Figure 3 displays wildfires and the kernel smoothing of California wildfire activity on California Department of Forestry and Fire Protection (CalFire) protected areas during the months of April from 1990 to 2006. Kernel smoothings highlight areas of higher or lower wildfire frequency in April, but do not indicate central tendency for the month of April. Prototypes highlight areas of higher wildfire frequency during April, while also providing an estimate of central tendency. As with the mean and kernel smoothing, the prototype is entirely non-parametric and, away from the boundary, the estimates typically do not depend critically on the boundaries of the spatial region. Further, prototypes and spike-time distance are amenable to basic clustering algorithms like K-means (MacQueen, 1967) and can thus be used to classify collections of point patterns or to identify significant changes in the point patterns over space or time.

This article is organized as follows. After briefly describing the California wildfire dataset that we explore in Section 2, extensions of prototype methods are proposed in Section 3. One extension is that of spike-time distance and prototype determination from the case of space-time point processes to that of space-time *marked* point processes. Another is the use of prototypes in clustering algorithms for the purpose of delineating wildfire seasons in California. Results from our prototype analyses of the California wildfire data are presented in Section 4, and a brief discussion is given in Section 5.

2. DATA

Data on wildfire occurrences on CalFire protected areas (see Figure 2) from 1990 through 2006 have been catalogued and provided by the U.S. Geological Survey, Western Ecological Research Center, and were provided to us by CalFire. The dataset is a compilation of several thousand wildfire recordings from various recording agencies. Variables of note for each observed wildfire in the dataset include total area burned, spatial coordinates for the estimated origin location of each wildfire, and the date of origination of each wildfire. Burn maps detailing the precise locations burned in each wildfire are available for many of the fires. A historical analysis of wildfires in these regions was performed by Keeley (1982), who found in particular that 16.2% of wildfires on these lands were caused by lightning, accounting for 13.1% of the total area burned, with the rest of the wildfires caused by humans. Though the database contains many smaller fires, the catalogue of southern California wildfires has been posited to be complete only for wildfires of size at least 0.0405 km^2 (Schoenberg *et al.*, 2003), so only the 6611 recorded wildfires burning at least 0.0405 km^2 are considered for this analysis. In what follows we consider the origin locations, times, and sizes of wildfires in each of the 17 years of data as a separate realization of a spatial-temporal marked point process.

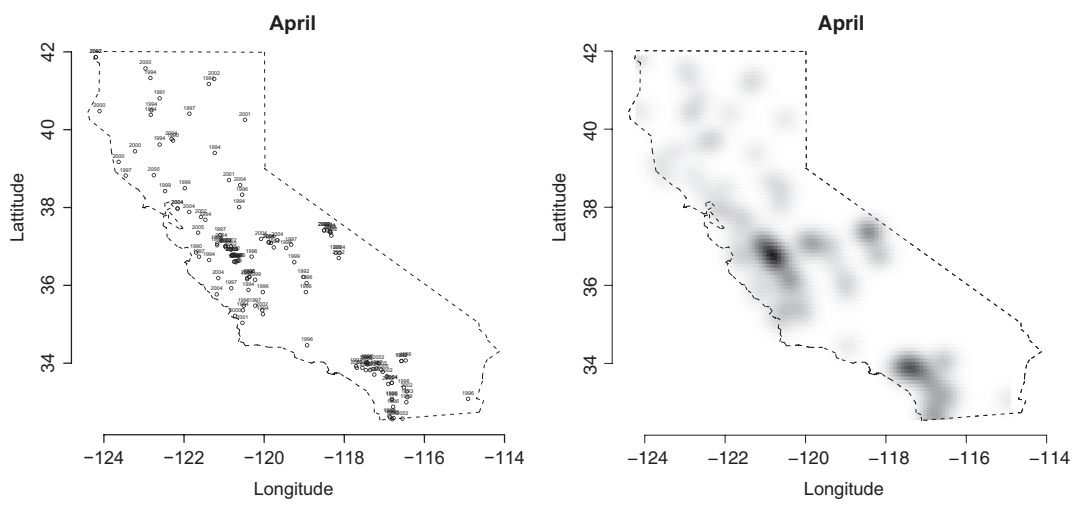

Figure 3. April wildfires that occurred in CalFire protected areas from 1989 to 2006 (left). Numbers indicate years of occurrence and locations indicate the origin locations of the wildfires. Kernel smoothing of April wildfire origin locations (right)

© 2011 Blackwell Publishing Ltd.

Wildfire impact on ecosystems can be measured in various ways, such as by area burned, by fire intensity or the energy output from a fire, by fire severity or the loss of vegetation biomass, and by human impacts such as loss of lives or property damage (Johnson and Miyanishi, 2001; Keeley, 2009). Following the majority of previous studies employing statistical analysis to a catalogue of wildfires, we explore burned area exclusively in this article as a measure of wildfire impact, but it should be noted that most of the methods proposed could alternately function with different choices of metrics.

3. METHODS

This article introduces two new extensions of prototypes of multidimensional point processes. The first is the adaptation of prototypes to *marked* point processes, so that one may not only summarize the spatial or temporal characteristics of a collection of point process realizations, but summarize the marks, or in the case of our applied problem the wildfire sizes as well. It is important to note upfront that prototyping in multiple dimensions is computationally expensive, especially if there are several thousand data points to consider. For this reason, special considerations need to be made, especially when it comes to prototyping the marks of a point process. The second extension is the use of clustering algorithms in conjunction with prototypes to estimate the start and end dates of wildfire seasons for the spatial region considered.

3.1. Prototypes of collections of marked point processes

Given a collection of realizations of a marked point process, one approach to assigning marks to the prototype of the collection is to consider the mark to be an additional dimension of the point process. There are three main problems, however, with treating the marks similarly to how one treats spatial or temporal coordinates. The first is a computational problem and is especially important with large spatial-temporal datasets. Since each coordinate of every point in the prototype of a collection of point process realizations is equal to a coordinate in one of the points in the collection (Tranbarger and Schoenberg, 2010), prototypes are typically determined by searching over large subsets of combinations of observed coordinates (Diez, 2010). Hence, the order of operations in the search increases exponentially with the number of dimensions of the domain. For the wildfire dataset described in Section 2, with two spatial coordinates, one temporal coordinate, and one mark coordinate for each point, searching over all spatial, temporal, and mark coordinates of all points would mean considering $6611^4 = 1.91 \times 10^{15}$ possibilities for each point in the prototype. This exceeds reasonable computing capacity.

A second problem is that in computation of spike-time distance, when considering moving a point of one point pattern to a point of another point pattern, treating differences in marks comparably to differences in spatial-temporal coordinates is problematic. One may choose movement penalties for the temporal and spatial domains objectively, using the criteria suggested by Tranbarger and Schoenberg (2010) or Diez *et al.* (2011). These suggested penalties are set so that in determining $D(X,Y)$, given two homogeneous Poisson processes X and Y with rate equal to the observed average rate of the realizations in the dataset, deleting any point in X or moving it the expected distance to its nearest neighbour in Y are equivalent. For a purely spatial point process, and, for instance, if $p_a = p_d = 1$, the spatial movement cost $p_m(\mathbf{z})$ would simply be $\sqrt{\frac{\pi\bar{n}}{|S|}}$, where $\bar{n}$ is the mean number of points in each spatial-point-process realization and $|S|$ represents the total area of the spatial support. Similarly, for a purely temporal point process, the temporal movement cost $p_m(t)$ would simply be $\frac{\bar{n}}{\Delta t}$, where Δt is the length of the temporal support for the data. Typically, distance for multiple-dimensional spike-time metrics are considered in the L1 norm, which conveniently allows for separate consideration of the spatial and temporal movement penalties for a spatial-temporal prototype. If $p_a \neq p_d$, then one may add the constraint that the prototype should have its number of points equal to the median number over the realizations in the dataset (Diez *et al.*, 2011). The above criteria are shown in Schoenberg and Tranbarger (2008) and Diez *et al.* (2011) to yield useful summaries of central tendency for spatial-temporal point processes. In the case of marked point processes, however, no such convenient criterion for determining the moving penalty for the mark coordinate appears to be readily available.

For application to wildfires, a third complication with treating the mark space in an analogous manner to the spatial and temporal coordinates lies in the fact that most of the wildfires are small in size, but the largest fires are of critical importance in forecasting and in summarizing previous activity. In the computation of spike-time distance, assigning a sizable penalty for moving points in the mark dimension results in many of the points with large marks being deleted rather than moved to points in the prototype. Thus, the prototype is constructed essentially by ignoring many of the largest fires in the dataset. On the other hand, if a very small penalty is used for moving the marks of the points, then wildfires with very different sizes will be given comparable weight in determining the prototype and, since most wildfires are small, the smallest fires will tend to have undue impact in determining the representation of the typical realization in the dataset.

Given a collection of realizations of a space-time (or purely spatial, or purely temporal) marked point process, we propose first to construct an initial prototype P of the realizations ignoring the marks. That is, find P by minimizing

$$\sum D(N_i, P),$$

using spike-time distance in time and space only (or equivalently, setting the moving penalty for changes in mark to zero). If the spatial support may reasonably be approximated by a rectangle, then appropriate values for the moving penalty parameters $\mathbf{p}_m$ governing the spike-time distance metric may be set using the relations

 © 2011 Blackwell Publishing Ltd.

$$p_m(t) = \frac{\bar{n}}{\Delta t},$$

$$p_m(x) = \sqrt{\frac{\pi \bar{n}}{|S_x|}},$$

and

$$p_m(y) = \sqrt{\frac{\pi \bar{n}}{|S_y|}},$$

where $p_m(t)$, $p_m(x)$, and $p_m(t)$ denote the moving penalties along the temporal axis, the x-axis, and the y-axis, respectively, and Δt, $|S_x|$ and $|S_y|$ represent the total duration, longitudinal range, and latitudinal range, respectively, of the spatial-temporal domain. Note that to create a prototype of a purely spatial marked point process, one can either ignore the times associated with the points or simply set the movement penalty in the temporal direction to zero. Similarly, setting the spatial moving penalties to zero results in a prototype of the marginal temporal marked point process.

One may ensure that the prototype contains the median number of points $(\tilde{n})$ of all point-process realizations in the dataset by allowing the addition and deletion penalties to differ. Following Diez *et al.* (2011) and Diez (2010), p_a and p_d are chosen to maximize $p_a p_d$ subject to the constraints that p_a and p_d must be non-negative with $p_a + p_d = 2$ and that the prototype must have $\tilde{n}$ points.

Once values for p_a, p_d, and $\mathbf{p}_m$ are determined, an estimate of the prototype P may readily be computed using the forward search algorithm implemented in the R statistical programming environment package *ppMeasures*. For each point P_j in the prototype, consider the collection of points x_{ij} across all observed point patterns such that the point x_{ij} is moved to P_j in the computation of the spike-time distance $D(N_i, P)$. These points $\{x_{ij} : i = 1, \ldots, m\}$ may be thought of as the points in the data most closely associated with the point P_j. We propose defining the mark associated with the point P_j as the mean of the marks associated with the points $\{x_{ij} : i = 1, 2, \ldots, m\}$. The end result is a prototype P where each point P_j has an assigned value for its mark determined entirely by the observation points that contribute, in terms of spike-time distance, to that point of the prototype.

3.2. Delineating California's fire seasons

We discuss in this section how prototypes may be used in conjunction with clustering algorithms, with an eye toward estimating the start and end dates of California's wildfire seasons. It is well known that wildfires in California occur seasonally, due especially to changes in wind patterns, climate, and land use throughout the year (Pyne *et al.*, 1996; Westerling *et al.*, 2003). For southern California ecosystems, the fire season is marked by large numbers of fires during the months of July and August (Westerling *et al.*, 2003) and by fires with extremely high intensity in October and November (Keeley *et al.*, 2009). A comprehensive understanding of wildfire behaviour in southern California suggests three separate seasons. One season is marked by frequent wildfire occurrence, the next by high wildfire area burned per wildfire, and the third by relative dormancy of fire activity. Indeed, recent efforts at forecasting wildfire activity have been made by separating wildfire activity into three seasons and fitting different parameters in statistical models for each season (Schoenberg *et al.*, 2007). For instance, Schoenberg *et al.* (2009) consider models with different parameters in three different wildfire seasons: a summer season when a high volume of fires occur, an autumn season when wildfire activity is most extreme, and a winter season when wildfire activity is relatively minimal. However, the delineation of these wildfire seasons is quite arbitrary in Schoenberg *et al.* (2009), and in this article we consider objective means for delineating the three main wildfire seasons in California, using prototypes.

Consider the determination of the precise start and end dates of each fire season as a problem in selecting three parameters t_1, t_2 and t_3 so as to optimally organize data into the fire seasons $(t_1, t_2]$, $(t_2, t_3]$ and $(t_3, t_1]$. Given any choice of dates t_1, t_2 and t_3, one may define the centroid of each of the three seasons as the prototype of the wildfire activity within that season, over the 17 years of observations described in Section 2. One may then divide each season into some number K of equally sized sub-intervals and then consider the summed distances from each of the $17K$ spatial-temporal realizations for each season to the corresponding portion of the prototype, using the spike-time distance metric. As noted in Section 3.1, it makes sense to treat the marks (sizes) of the wildfires differently from the spatial and temporal coordinates. Hence, we consider finding a spatial-temporal prototype for each season and, separately, a prototype for the marginal collection of marks within each season. The optimal parameters t_1, t_2 and t_3 may then be estimated by minimizing the sum of the spike-time distances from the corresponding prototypes; that is, minimize

$$\sum_{i=1}^{17} \sum_{j=1}^{3} \left\{ \sum_{k=1}^{K} D(N_{ijk}, P_{jk}) + D(M_{ij}, Q_j) \right\}, \tag{2}$$

where N_{ijk} is the observed spatial-temporal data of wildfire origin times and locations in year i during portion k of season j, P_{jk} is the spatial-temporal prototype for portion k of season j, M_{ij} are the observed wildfire sizes in year i during season j, and Q_j is the mark prototype of the wildfire sizes in season j.

In order for this search for the minimizer of (2) to be meaningful, it is necessary for the spike-time distances in (2) to be comparable and, thus, to have the same spike-time penalty parameters for all three seasons and all years, rather than data-dependent penalty parameters for each season. Hence, a different approach to selecting p_a, p_d and $\mathbf{p}_m$ than the one described in Section 3.1 is required. We propose first setting the addition and deletion penalties to unity and then setting the spatial movement penalty parameters to the inverses of the marginal standard deviations of the data. That is, set

© 2011 Blackwell Publishing Ltd.

$$p_a = p_d = 1, \;\; \mathbf{p}_m = \{1/\sigma_t, 1/\sigma_{z_1}, 1/\sigma_{z_2}\}, \;\; \tilde{p}_m = 1/\sigma_s, \tag{3}$$

where σ_t, σ_{z_1} and σ_{z_2} are the standard deviations of the marginal temporal and marginal spatial occurrences, respectively, of the entire observed dataset of 17 years, $\tilde{p}_m$ is the moving penalty for the marginal mark prototype, and σ_s is the standard deviation of the observed wildfire sizes. The idea behind (3) is that the prototype-construction algorithm will be indifferent, in terms of spike-time distance, between moving a typical observed point one standard deviation versus adding a point.

It is extremely computationally burdensome to consider each of the more than 8.1 million possible combinations of dates for t_1, t_2 and t_3, and to compute all the prototypes and spike-time distances in (2) for each combination. For this reason, we implement a forward-search algorithm to find an approximate minimizer of (2), by first searching over a coarse, limited choice of t_1,t_2 and t_3, and then progressively refining the search.

Specifically, we first limit the search only to the following six possible dates: January 1st, March 1st, May 1st, July 1st, September 1st, November 1st. Next, we refine the search by considering a search over the combinations $\{t_1^* \pm 0$ or 1 month, $t_2^* \pm 0$ or 1 month, $t_3^* \pm 0$ or 1 month$\}$, where (t_1^*, t_2^*, t_3^*) was the optimum set of values found previously in the search over six possible dates. For example, if the optimal starting dates after the first step were $t_1^* =$ January 1st, $t_2^* =$ May 1st and $t_3^* =$ November 1st, then we search from all possible combinations of:

$$t_1 \in \text{(December 1st, January 1st, February 1st)},$$

$$t_2 \in \text{(April 1st, May 1st, June 1st)},$$

$$t_3 \in \text{(October 1st, November 1st, December 1st)}.$$

The iterative search then continues by considering values from the previous step's optimal solution ±15 days then ±8 days, ±4 days, ±2 days and finally ±1 day. The algorithm then selects the minimum of eqn (2) over all values inspected in this forward search; note that the algorithm does not necessarily find the minimum of eqn (2) over all possible choices. of t_1, t_2, and t_3, however.

4. RESULTS

Figures 4 and 5 show prototypes for each of the 12 months of the calendar year, along with the original data from which they were derived. The prototypes created in Figures 4 and 5 are purely spatial prototypes. One can observe in Figures 4 and 5 that the activity is increased in the summer months, especially July and August, and that the sizes of the prototypical fires burning in October are much larger than those in the other months.

The prototypes in Figures 4 and 5 appear to highlight two regions in California as particularly susceptible to wildfire occurrence from May through October. The first is just north of Sacramento in northern California. This region is surrounded by the El Dorado, Mendocino, Plumas and Tahoe National Forests. Of the 6 months from May through October, three of them (June, August, September) have their largest marked prototype points near Chico and Madera, which are cities found cradled between the forests north of Sacramento. The remaining 3 months (May, July, October) have their largest prototype fires near San Diego and Orange Counties in southern California. Another feature well summarized using the prototypes in Figures 4 and 5 is that, from May through September, the typical sizes of large wildfires in the most dangerous regions in northern and southern California are comparable within each month. However, in the month of October, the largest prototype fires in the San Diego/Orange County region are vastly greater than the largest prototype fires in the Chico/Madera region. In fact, for the month of October, the prototype points near Julian, CA, which is just east of San Diego and is surrounded by the Cleveland National Forest, contains prototypical fires that are several times greater in burn area than the largest prototypical fires found in northern California.

After an initialization of the combination of potential dates for the beginning of each of the three wildfire seasons, the algorithm described in Section 3.2 was run. As a result, the start and end dates were found the minimize the cumulative summed difference from all contributing spatial point pattern realizations for each season to a spatial prototype and size prototype.

Table 1 provides the results of this search for the optimal partitioning of the wildfire seasons in California for $K = 4$. Each row of Table 1 provides the optimal partitioning of the data into wildfire seasons, after an iteration of the algorithm. Total summed cost from eqn (2) has also been provided in Table 1 to illustrate how effective each partitioning of the data was at the end of each iteration. The algorithm continues to refine until ultimately considering seasonal partitions to the precision of a single day. One season, which we might call the active or summer wildfire season, is estimated to begin on May 25th and to end on September 26th. The fall season, characterized by fewer but larger wildfires, is estimated to begin on September 27th and to end on November 7th, and the winter wildfire season, characterized by relative dormancy, is estimated to begin on November 8th and to end on May 24th. These results are consistent with, and more precise and objectively estimated than, conventional estimates of wildfire seasons, such as those in Schoenberg *et al.* (2009).

Figure 6 displays the temporal occurrences of the prototype points from the end of April through the beginning of June. Prior to the end of the relatively dormant season (marked by the left vertical line in Figure 6), which incidentally is just slightly before the average start date of Memorial Day weekend (marked by the right vertical line in Figure 6), there are far fewer prototype points than after the start of the active summer wildfire season. Indeed, after the identified starting date for the summer wildfire season, wildfires of at least 0.0405 km^2 occur nearly daily in California.

© 2011 Blackwell Publishing Ltd.

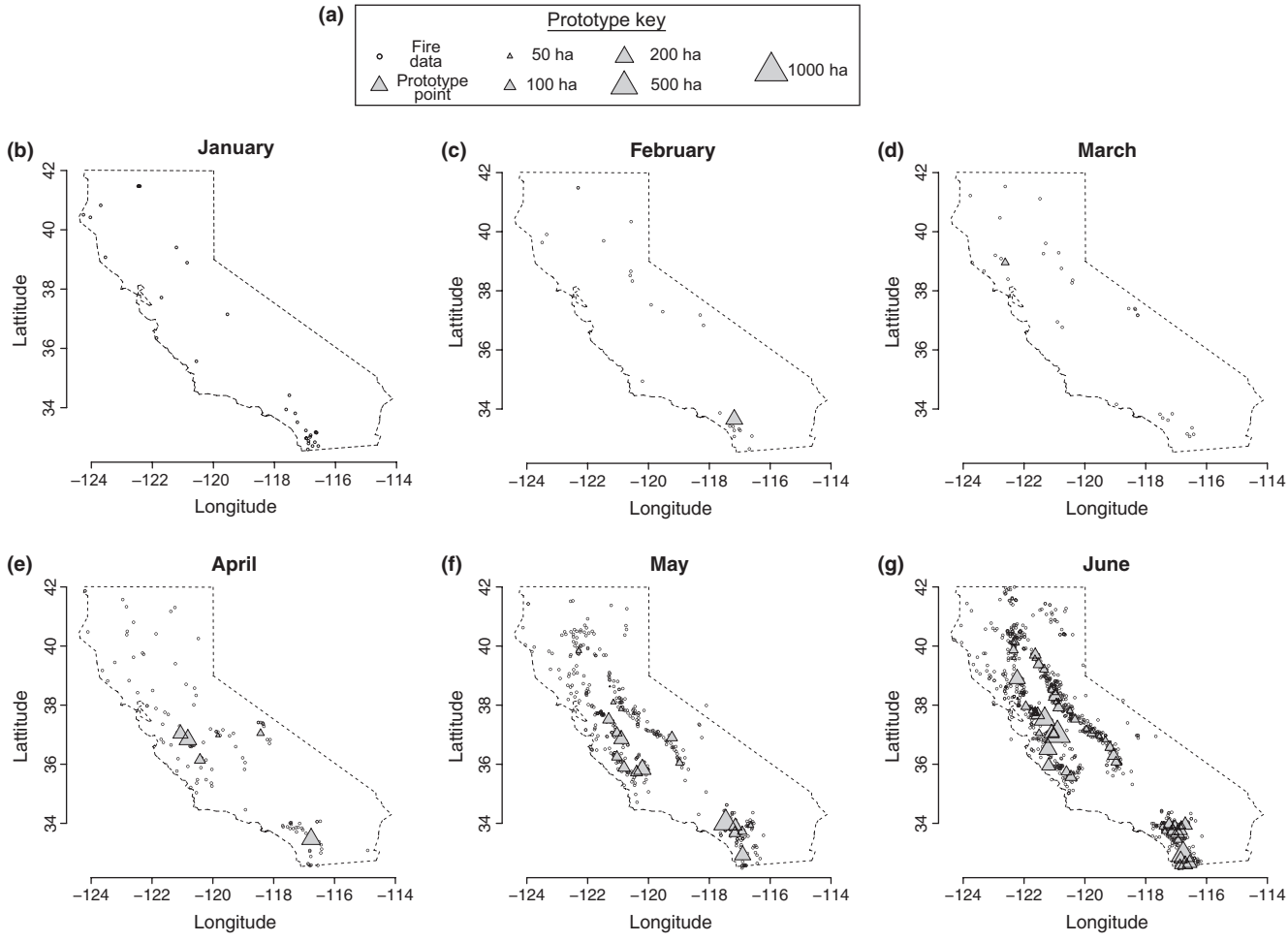

Figure 4. Prototype summaries with mark for January–June. Prototype wildfires are represented by equilateral triangles whose sides are proportional to log areas. Observed wildfire origin locations are represented by circles

The distinction between the seasons is also highlighted in Figure 7, which shows the prototype for each wildfire season. One sees readily from Figure 7 the extremely high frequency of wildfires during the summer season, the predominance of small fires in the winter season, especially in northern California, and the tendency toward large wildfires in the fall season, especially in southern California.

5. DISCUSSION

Prototypes are useful summaries for repeated observations of a point process, can be used for spatial-temporal marked point processes, and can easily be integrated with clustering algorithms so as to partition realizations optimally. One particular advantage of prototypes is that they are entirely non-parametric, and in fact they do not even require that the underlying process be stationary or isotropic. This is a major convenience in the case of California wildfires, which are obviously highly non-stationary in both space and time, with certain locations and months far more fire prone than others, and strongly non-isotropic, since wildfires are far more likely to spread in certain directions, for instance in the direction of prevailing ambient winds and toward higher elevations (Pyne *et al.*, 1996).

It is worth noting that spike-time distance is merely one possible choice for distance between point patterns and, while its simplicity is attractive, for alternative point processes, especially those exhibiting clustering, other choices of distance functions may be used (Victor and Purpura, 1997; Mateu *et al.*, 2011).

The results for California wildfires found here are not surprising, and indeed they are more illustrative of the usefulness of prototypes to summarize the main features in a complex dataset, rather than as means to glean new insights about wildfire activity in California. The observations that most California wildfires occur in the summer and that the largest wildfires tend to occur in the fall were in fact known to the Chumash Indians many centuries ago (Pyne *et al.*, 1996). Nevertheless, summaries based on prototypes may be useful, particularly in the case of quantification or delineation of wildfire seasons. Monthly or annual prototypes may also

© 2011 Blackwell Publishing Ltd.

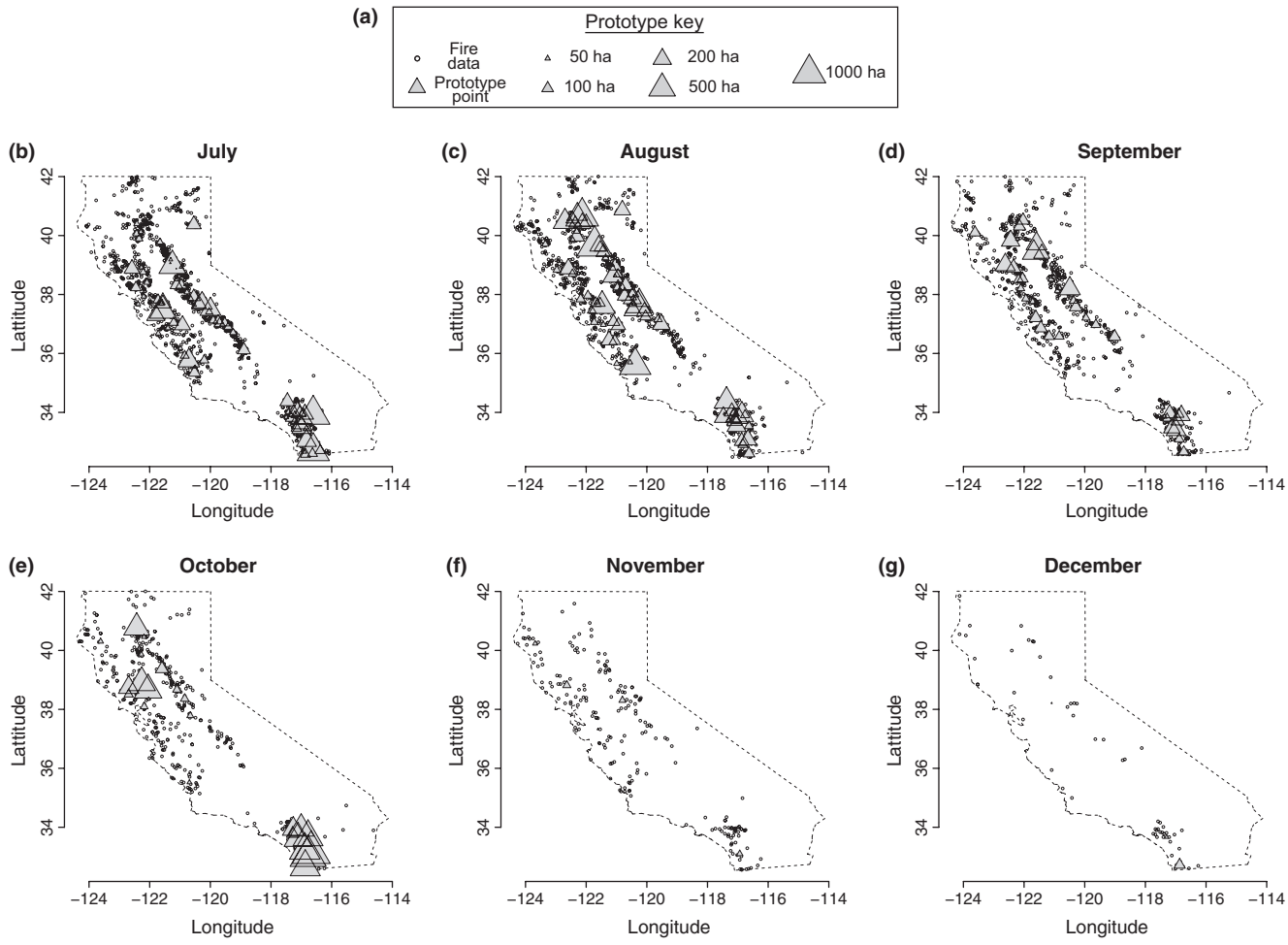

Figure 5. Prototype summaries with mark for July–December. Prototype wildfires are represented by equilateral triangles whose sides are proportional to log areas. Observed wildfire origin locations are represented by circles

Table 1. Optimal temporal clustering of California wildfire data into three wildfire seasons

Iteration i	$t^*_{1,i}$	$t^*_{2,i}$	$t^*_{3,i}$	$\sum$ Cost
1	4/30	10/31	12/31	6993.26
2	531	9/30	11/30	6654.45
3	5/16	9/30	11/16	6577.35
4	5/24	9/30	11/08	6531.42
5	5/24	9/26	11/08	6438.43
6	5/24	9/26	11/08	6438.43
7	5/24	9/26	11/07	6435.84

serve as useful and easily interpretable summaries of expected wildfire activity. Further, for the problems of wildfire forecasting and the estimation of wildfire burn probabilities based on meteorological variables, which are important for urban planning and for the preparation of wildfire suppression and emergency response, a key component is the precise delineation of wildfire seasons, since wildfire hazard models often involve different parameters for different seasons (see e.g. Xu and Schoenberg, 2011). Prototypes may thus be quite directly useful for wildfire forecasting.

The obvious connection between prototypes and medians has been noted previously (Schoenberg and Tranbarger, 2008; Diez, 2010; Tranbarger and Schoenberg, 2010). Like the median as a descriptor of the center of univariate data, the prototype may be a useful summary of the central tendency behaviour of a point process, and it is especially useful when the observations are independent or at least approximately so. In the case of years of wildfire activity, the processes will not be independent, since if a particular location burns one year, then it is perhaps less likely to burn the next year. The data seem to suggest that for California as a whole, the overall dependence between years does not seem especially strong. Indeed, Peng *et al.* (2005) found this dependence for a particular location to be statistically significant but that, when considering regional wildfire activity as a whole, year-to-year dependence was a rather weak factor in forecasting wildfire activity.

© 2011 Blackwell Publishing Ltd.

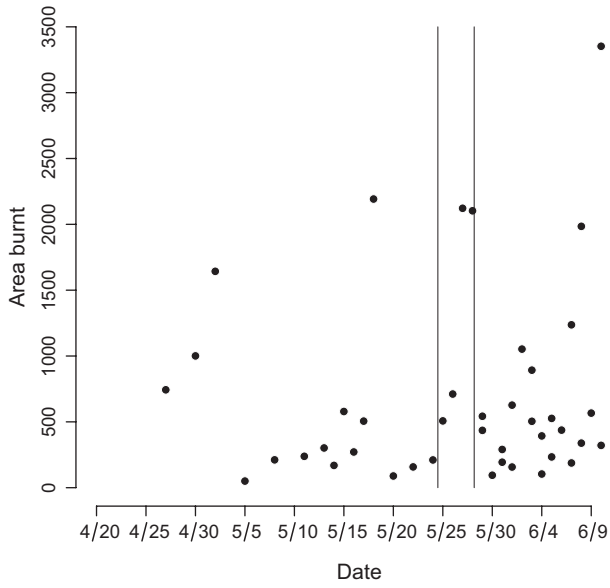

Figure 6. Prototype points from April 20th to June 10th. Vertical lines represent the estimated end date of the winter season (Left) and the average date of Memorial Day (Right) from 1990 to 2006. The vertical axis indicates the mean area of the wildfires associated with each prototype wildfire in the computation of spike-time distance, as described in Section 3.2

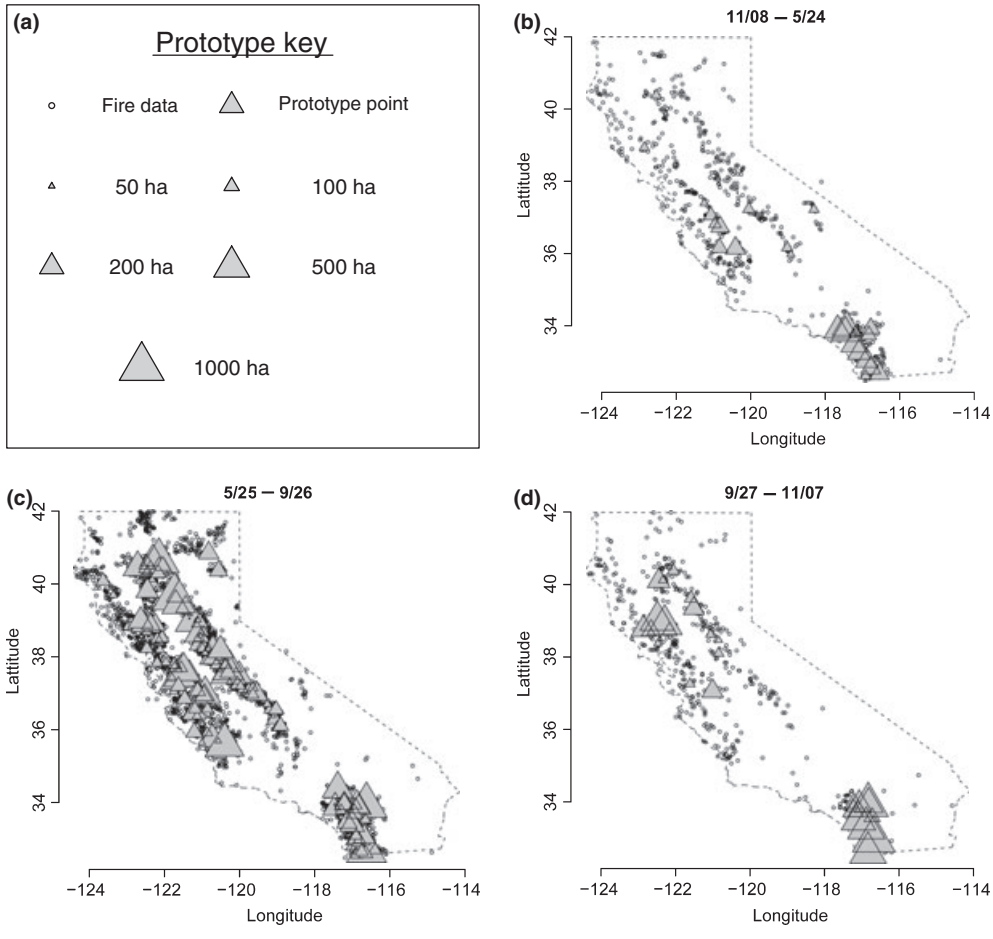

Figure 7. Prototypes for the three delineated fire seasons, using spike-time distance with parameters $p_a = p_d = 1$ and $p_m = 0.3$ determined as described in Section 3.2. Prototype wildfires are represented by equilateral triangles whose sides are proportional to log areas. Observed wildfire origin locations are represented by circles

© 2011 Blackwell Publishing Ltd.

Just as the median may be a useful summary and is commonly used even in the presence of some possible minor dependence in the data, we conjecture that the same may be the case for prototypes for datasets consisting of weakly correlated point-process realizations, and the impact of such dependence on prototypes is an important direction for future research. Other important tasks for further exploration include the assessment of the variability in the delineation of wildfire seasons, perhaps using bootstrap methods, and the study of causal connections between the marks and the times and spatial locations of wildfire occurrences.

Acknowledgements

We thank the California Department of Forestry and Fire Protection for their data. Any use of trade, product or firm names in this publication is for descriptive purposes only and does not imply endorsement by the US Government. Many of the computations were performed using the ppMeasures package in the R statistical programming environment. The Associate Editor and anonymous reviewers provided numerous helpful comments and suggestions.

REFERENCES

Diez, D. (2010) *Extensions of Distance and Prototype Methods for Point Patterns*. PhD Thesis, University of California.

Diez, D., Schoenberg, F. P. and Woody, C. D. (2011) Analysis of neuronal responses to stimuli in cats using point process prototypes. *Biometrics*, in review.

Johnson E. A. and Gutsell S. L. (1994) Fire frequency models, methods and iterpretations. *Advanced Ecology Research* **25**, 239–87.

Johnson, E. A. and Miyanishi, K. (2001) *Forest Fires: Behavior and Ecological Effects*. San Diego: Academic Press.

Keeley, J. E. (1982) Distribution of lightning- and man-caused wildfires in California. In *Proceedings of the Symposium on Dynamics and Management of Mediterranean-type Ecosystems* (eds C. E. Conrad and W. C.Oechel). USDA Forest Service, Pacific Southwest Forest and Range Experiment Station, General Technical Report PSW-58, pp. 431–7.

Keeley, J. E. (2009) Fire intensity, fire severity and burn severity: a brief review and suggested usage. *International Journal of Wildland Fire* **18**, 116–26.

Keeley, J. E., Fotheringham, C. J. and Morais, M. (1999) Reexamining fire suppression impacts on brushland fire regimes. *Science* **284**, 1829–32.

Keeley, J. E., Safford, H., Fotheringham, C. J., Franklin, J. and Moritz, M. (2009) 2007 Southern California wildfires: lessons in complexity. *Journal of Forestry* **107**, 287–96.

MacQueen, J. B. (1967) Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, California, vol. **1**, pp. 281–97.

Malamud, B. D., Millington, J. D. A. and Perry, G. L. W. (2005) Characterizing wildfire regimes in the United States. *Proceedings of the National Academy of Sciences of the USA* **102**(13), 4694–99.

Mateu, J., Schoenberg, F. P. and Diez, D. (2011) On distances between point patterns and their applications. *American Statistician*, in review.

Moritz, M. A., Morais, M. E., Summerell, L. A. and Doyle, J. (2005) Wildfires, complexity, and highly optimized tolerance. *Proceedings of the National Academy of Sciences of the USA* **102**, 17912–7.

Peng, R. D., Schoenberg, F. P. and Woods, J. (2005) A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association* **100**(469), 26–35.

Pyne, S., Andrews, P. and Laren, R. (1996) *Introduction to Wildland Fire*. New York: Wiley.

Schoenberg, F. P. and Tranbarger, K. E. (2008) Description of earthquake aftershock sequences using prototype point processes. *Environmetrics* **19**, 271–86.

Schoenberg, F., Brillinger, D. R., and Guttorp, P. M. (2002) Point processes, spatial-temporal. In *Encyclopedia of Environmentrics*. (eds A.El-Shaarawi and W. Piegorsch). New York: Wiley, Vol. **3**, pp. 1573–7.

Schoenberg, F. P., Peng, R. and Woods, J. (2003) On the distribution of wildfire sizes. *Environmetrics* **14**(6), 583–92.

Schoenberg, F. P., Chang, C., Keeley, J., Pompa, J., Woods, J. and Xu, H. (2007) A critical assessment of the burning index in Los Angeles County, California. *International Journal of Wildland Fire* **16**(4), 473–83.

Schoenberg, F. P., Pompa, J. L. and Chang, C. (2009) A note on non-parametric and semi-parametric modeling of wildfire hazard in Los Angeles County, California. *Environmental and Ecological Statistics* **16**(2), 251–69.

Tranbarger, K. E. and Schoenberg, F. P. (2010) On the computation and application of point process prototypes. *Open Applied Informatics Journal* **4**, 1–9.

Victor, J. and Purpura, K. (1997) Metric-space analysis of spike trains: theory, algorithms and application. *Journal of Neuroscience Methods* **8**, 127–64.

Westerling A. L., Brown, T. J., Gershunuov, D. R., Cayan D. and Dettinger, M. D. (2003) Climate and wildfire in the Western United States. *Bulletin of the American Meteorological Society* **84**, 595–604.

Xu, H. and Schoenberg, F. P. (2011) Point process modeling of wildfire hazard in Los Angeles County, California. *Annals of Applied Statistics*, in press.

© 2011 Blackwell Publishing Ltd.