Journal of Computational and Graphical Statistics

Nonparametric Approach to Analysis of Space-Time Data on Earthquake Occurrences

Edwin Choi [a] & Peter Hall [a]

[a] Centre for Mathematics and its Applications , The Australian
National University , Canberra , ACT 0200 , Australia
Published online: 21 Feb 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Nonparametric Approach to Analysis of Space-Time Data on Earthquake Occurrences

Edwin CHOI and Peter HALL

As an alternative to traditional, parametric approaches, we suggest nonparametric methods for analyzing spatial and temporal data on earthquake occurrences. Nonparametric techniques are particularly adaptive to anomalous behavior in the data and provide a new way of accessing a variety of different types of information about the way in which both intensity and magnitude of events evolve in time. They can be employed to estimate the spatial trajectory of event clusters as a function of time, and to define quiescent and active periods. The latter application suggests new approaches to forecasting high magnitude events. Our methods are founded on multivariate techniques for curve and surface estimation, particularly in contexts where curves or surfaces are unbounded at points or along lines.

Key Words: Bandwidth; Energy; Intensity function; Japan; Kernel; Magnitude; Near neighbor method; Point process; Poisson process; Pole; Quiescent period; Smoothing; Spatial analysis.

1. INTRODUCTION

The problem of developing space-time point process models for earthquake occurrence is of great interest to seismologists. Often, models are applied with the aim of capturing relatively conventional characteristics of earthquake activities. Sometimes, however, they suggest anomalous seismicity patterns in a region. Popular models include the epidemic type aftershock sequence (ETAS) model (Ogata 1988), which is used to describe temporal behavior of an earthquake series; and refinements of Hawkes' (1971) self-exciting point process model, which describe spatial-temporal patterns in a catalog. Ogata (1998) gave detailed discussion of recent extensions of these models.

Standard models applied to seismic data do not always fit those data well, however. In part, this is because parametric models are usually well suited only to a sequence of seismic events that have similar causes. Such problems are reflected in the fact that fitted values of parameters sometimes lie well outside plausible ranges, and in the instability of parameter estimates when comparatively small amounts of new data are added. Moreover, parametric models can be insensitive to anomalous events, since models tend

Edwin Choi is Research Associate, and Peter Hall is Professor, Centre for Mathematics and its Applications, The Australian National University, Canberra, ACT 0200, Australia.

to be formulated through experience of relatively conventional seismic activity. Indeed, anomalies are often the root cause of the aforementioned parameter instability. Since anomalous events are often of at least as much interest as conventional ones (see Ogata 1989), then procedures that tend to conceal anomalies are not necessarily to be preferred.

In this article we suggest nonparametric methods for analyzing seismic data. The methods do not require formulation of structural models, so they do not suffer from the deficiencies noted above. They involve several different approaches to nonparametric estimation of the intensity of a point process that evolves with time. Thus, the intrinsic nonstationarity of the point process is especially targeted. This enables us to quantify the spatial migration of seismic sites with time, to analyze event strength as a function of time, and to study temporal patterns of active and quiescent periods within a region, without constraining these features to have predetermined properties. We argue that, especially as tools for exploratory data analysis, nonparametric methods for the analysis of earthquake data are valuable supplements to more conventional, parametric approaches.

Spatial and temporal data on earthquake occurrences typically involve at least five dimensions—there are three spatial dimensions, one temporal, and one dimension representing strength, often expressed as magnitude on the Richter scale. We use smoothing methods extensively to estimate intensity and strength as functions of space and time. When estimating intensity, a modified near neighbor technique is used to select smoothing parameters in a locally adaptive way. This can be crucial, since the shape of the intensity function varies so greatly, from sharply peaked to almost flat, that global bandwidth choice often fails to adequately reproduce important features of intensity. Indeed, our examples provide evidence that the actual intensity is not square integrable, suggesting that peaks might really be poles or pole lines (i.e., places where the intensity diverges to infinity). This means that conventional methods for bandwidth choice, based for example on least squares cross-validation, may be inappropriate.

Section 2 introduces our nonparametric methods for estimating point process intensity and develops them into techniques for other applications, including trajectory plotting and quiescent period estimation. Section 3 applies the methods to Kanto seismic data.

2. METHODOLOGY

2.1 INTRINSIC DIFFERENCES AMONG COORDINATES

The observed data may be considered to be a realization of a $(d + 2)$-variate point process, where the first d dimensions represent space, the $(d + 1)$st dimension denotes time, the last dimension is strength, and each point represents an event. Alternatively, the point process could be considered as a marked $(d + 1)$-variate process in space and time, where the mark denotes strength.

In the case of earthquake data there are nominally $d = 3$ spatial dimensions, representing latitude, longitude, and depth below the earth's surface. The latter is often not able to be estimated accurately, however, and in fact is allocated to bins that are up to 10 km wide. Reflecting this difficulty, we suggest pooling bins (e.g., confining attention to data at depths of less than 36 km, as in the work in Section 3). After this change, the spatial aspect of our analysis involves only $d = 2$ dimensions. The spatial dimensions

too are recorded with a degree of error, which depends on, among other matters, the spatial distribution of recording stations around the location of the event, and on event depth.

2.2 DEPICTING INTENSITY ESTIMATES

If $d \geq 2$, then the intensity of the evolving space-time process, being a function in $\mathbb{R}^{d+1}$, is not readily representable at a glance. However, its projections are more easily interpreted. Thus, the full intensity may be depicted in d contour or perspective plots of three-dimensional graphs, where in the ith plot, time and the ith spatial dimension are indicated on the horizontal axes, and intensity is on the vertical axis. Note that the appropriateness of this dimension-reduction technique, and the choice of projection axes, depend very much on the application involved. In the context of applications addressed in this article the axes are clear from physical considerations, but this may not be the case in other contexts.

If event strength is a scalar, then it may be depicted by adding color to graphs of estimated intensity, for example in the spectrum from blue (representing weak events) to red (strong events). Thus, strength data can be included on the plots described earlier, without increasing the complexity of the set of projections. For the sake of economy, however, in this article strength will be shown in separate contour plots. Note that, since strength is recorded with error, and only at scattered points in space and time, then it is generally necessary to smooth strength measurements before ascribing a color.

2.3 NONPARAMETRIC ESTIMATION OF INTENSITY

If intensity changes rapidly with respect to either spatial location or time (often it varies sharply with both variables), then it is best estimated using a method that employs a varying bandwidth. We suggest a "coupled" form of near neighbor methods, where spatial and temporal coordinates are treated orthogonally but their smoothing parameters are linked through an adaptive definition of near neighbors in time and space. The reader is referred to Silverman (1986, sec. 5.2) for an account of more conventional near neighbor methods.

Specifically, let X_i (a scalar) denote the projection of the ith d-variate datum onto a specified axis $\mathcal{A}$, and let T_i be the time at which that datum was observed. (Spatial projection is required so as to present variation in space and time in a single diagram.) The space-time intensity λ at (x, t) is defined to equal the expected number of events per unit area at (x, t), and is estimated by

$$\hat{\lambda}(x, t) = \frac{1}{h_1 h_2} \sum_i K_1 \left(\frac{X_i - x}{h_1} \right) K_1 \left(\frac{T_i - t}{h_2} \right), \qquad (2.1)$$

where K_1 is a univariate kernel and h_1, h_2 are bandwidths for spatial and temporal components, respectively. Theoretical properties of $\hat{\lambda}$ are outlined in the appendix. Subsequent to computing $\hat{\lambda}$ for data projected onto $\mathcal{A}$ we calculated its analog when projection was perpendicular to $\mathcal{A}$.

Our "coupled" near neighbor method is based on choosing h_1 and h_2 so as to minimize $h_1 + \alpha h_2$ subject to the box $\mathcal{B} = (x - h_1, x + h_1) \times (t - h_2, t + h_2)$ containing at least k space-time pairs (X_i, T_i). Thus, k (rather than h_1 or h_2) is the principal smoothing parameter. The value of k controls the overall level of smoothing (larger k's give a greater amount of smoothing) while α controls the relative emphasis placed on spatial and temporal coordinates. For example if α is chosen small, then the resulting box $\mathcal{B}$ is stretched along the time axis, and the intensity estimator is localized more in space than in time.

Cross-validation choice of k may be implemented in the usual manner (e.g., Wand and Jones 1995, p. 63f). Given the complexities of the problem involved, however, our choice of k in the numerical work in Section 3 will be based on experimentation.

2.4 ESTIMATING THE SPATIAL TRAJECTORY OF PEAK INTENSITY

The evolution in space of local maxima of the intensity, as time increases, is of particular interest. For the purpose of estimating the trajectory we first estimate spatial intensity as a function of time, in the continuum. To this end we choose a time window of width $2h_2$, and use those data with occurrence times in the interval $\mathcal{T}(t) = [t - h_2, t + h_2]$. As before, we bin the depth component of data vectors, but we no longer project data onto the axis $\mathcal{A}$. Therefore, the data $(Y_i^T, T_i)^T$ that we employ are vectors of length 3, with Y_i, a column vector of length $d = 2$, representing the spatial coordinates of the datum observed at time T_i.

The spatial intensity, $\mu(y \mid t)$, of events that occur at time t is defined to equal the expected number of events that occur per unit area at point y in the plane, when the process is observed at time t. It is estimated by

$$\hat{\mu}(y \mid t) = \frac{1}{h_2 h_3^2} \sum_{i : T_i \in \mathcal{T}(t)} K_1 \left(\frac{T_i - t}{h_2} \right) K_2 \left(\frac{Y_i - y}{h_3} \right), \qquad (2.2)$$

where K_2 is a bivariate kernel and h_3 is a new bandwidth. Here, we have employed a bandwidth matrix $h_3 I_2$ for K_2, where I_2 denotes the 2×2 identity matrix. The notation y will always denote a 2×1 vector in this article. When used in the present context the near neighbor method suggested in Section 2.3 tends to produce closely neighboring, multiple peaks in places where intensity is high, and so is not recommended.

Once $\hat{\mu}(\cdot \mid t)$ has been computed we may readily calculate the spatial location of its maximum, being a bivariate function of t. Using quadratic interpolation from gridded values of t we may estimate the trajectory of the maximum, again as a function of t. Since seismic activity is typically low between relatively short periods of high intensity, it is usually necessary to threshold the maximum intensity at some value τ, say, in order to obtain a trajectory that is representative of the more interesting episodes.

2.5 ESTIMATORS OF STRENGTH

In Section 3.4 we shall discuss three monotonically related measures of strength: magnitude, energy, and stress. Let $g(x, t)$ denote the expected value of strength, measured

in any of these scales, for an event at point (x, t) in space and time. An estimator of g may be defined in the manner of a Nadaraya–Watson estimator in nonparametric regression,

$$\hat{g}(x, t) = \frac{\sum_i K_1\left(\frac{X_i - x}{h_1}\right) K_1\left(\frac{T_i - t}{h_2}\right) G_i}{\sum_i K_1\left(\frac{X_i - x}{h_1}\right) K_1\left(\frac{T_i - t}{h_2}\right)}, \tag{2.3}$$

where G_i denotes the observed strength of the event that occurred at time T_i.

Alternatively, local polynomial methods could be employed. In that case, however, we should consider estimating the logarithm of strength first, and then exponentiating, in order to avoid obtaining negative strength estimates. The disadvantages of Nadaraya–Watson methods in the case of smooth functions (see, e.g., Fan 1993) are not readily apparent here, since $g(x, t)$ behaves so erratically.

From a physical viewpoint, the case where G_i in (2.3) is taken to be energy is preferable, since (unlike magnitude or stress) energy is additive. However, we found that the scale on which energy was recorded was far too long to be effectively accommodated using conventional smoothing methods, and therefore worked instead with its logarithm, which is a linear function of magnitude.

2.6 ESTIMATING ACTIVE AND QUIESCENT PERIODS

Considerable interest centers on detecting changes in seismicity patterns within a region, since they may provide clues for earthquake prediction. Relatively quiet, or quiescent, periods are of particular importance, since they seem often to be precursors to seismic activity. For example, Mogi (1968) investigated a number of strong earthquakes in Japan and showed that the focal regions were relatively calm before the main shocks.

Methods used to investigate quiescent periods are currently parametric; see Ogata (1988, 1989, 1992). Zhuang (1998) fitted the ETAS model to data on earthquake sequences, with the aim of detecting changes in seismicity patterns in a region. He identified a quiescent period before the Cape Palliser earthquake (on October 5 1990, with magnitude 5.3 on the Richter scale) in New Zealand. A drawback of this approach, however, is that it depends very much on the model, and so is subject to the instability and goodness of fit problems noted in Section 1. Using our nonparametric methods we may estimate active and quiescent periods within the whole of a region or within subregions, and estimate the distribution of lengths of these periods, without invoking a structural model.

We suggest the following threshold-based definition of active and quiescent periods. Let m denote a threshold on magnitude, and let $\mu_m(\cdot \mid t)$ be the spatial density of events, with magnitude at least m, that occur at time t. Put $\gamma_m(t) = \sup_y \mu_m(y \mid t)$. Given values $\gamma_1 \leq \gamma_2$, choice of which (like that of m) is up to the experimenter, we say that t is part of a quiescent period if $\gamma_m(t) \leq \gamma_1$, and part of an active period if $\gamma_m(t) \geq \gamma_2$. Quiescent [active] periods are defined to be maximal intervals of values t for which $\gamma_m(t) \leq \gamma_1$ [respectively, $\gamma_m(t) \geq \gamma_2$].

We may estimate μ_m by slightly modifying the estimator of μ at (2.2):

$$\hat{\mu}_m(y \mid t) = \frac{1}{h_2 h_3^2} \sum_{i:T_i \in \mathcal{T}(t)} K_1\left(\frac{T_i - t}{h_2}\right) K_2\left(\frac{Y_i - y}{h_3}\right) I(M_i \geq m), \tag{2.4}$$

where M_i denotes the magnitude of the event that occurred at time T_i. (Alternatively, we could incorporate M_i into the summand at (2.4); see Section 3.5.) Put $\widehat{\gamma}_m(t) = \sup_y \widehat{\mu}_m(y|t)$, and define an estimator of an active or quiescent period to be a maximal interval of values t such that $\widehat{\gamma}_m(t) \leq \gamma_1$ or $\widehat{\gamma}_m(t) \geq \gamma_2$, respectively.

In practice it is not always necessary to threshold on the value of magnitude, since adequate thresholding might have been implicit in the decision about which events to include in the basic intensity estimators suggested in Section 2.3. That is, we might be able to use $\widehat{\mu}$ rather than $\widehat{\mu}_m$. Nevertheless, the relationship of seismic events for which the magnitude is at least a certain amount, to the temporal location and length of quiescent periods, is of particular interest. Therefore, although our estimators of active and quiescent periods might not explicitly involve strength data, those data would often be necessary in order to make inference from quiescent period estimates. See Section 3.5.

3. EXAMPLE: KANTO SEISMIC DATA

3.1 THE DATA

We analyzed Kanto seismic data, compiled by the Center for Disaster Prevention at Tsukuba, Japan, for the years 1980–1993. For the analysis reported here we took $d = 2$, representing spatial location on the surface of the earth. We restricted attention to events whose location was between 138.6° and 139.7° longitude and 34.6° and 35.7° latitude (which we call region $\mathcal{R}$), whose depth was less than 36 km, and whose magnitude was
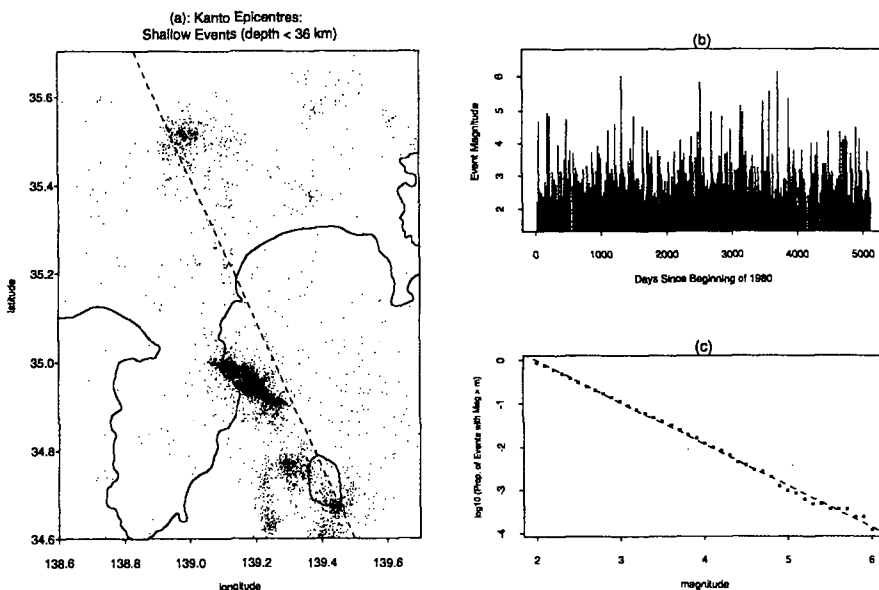

Figure 1. Basic descriptive plots of the seismic data. Panel (a) depicts the geographic locations of epicenters. The dotted line represents an approximation to the volcanic front of the Izu-Bonin Arc. Panel (b) plots magnitude over time, while panel (c) is a frequency-magnitude plot.
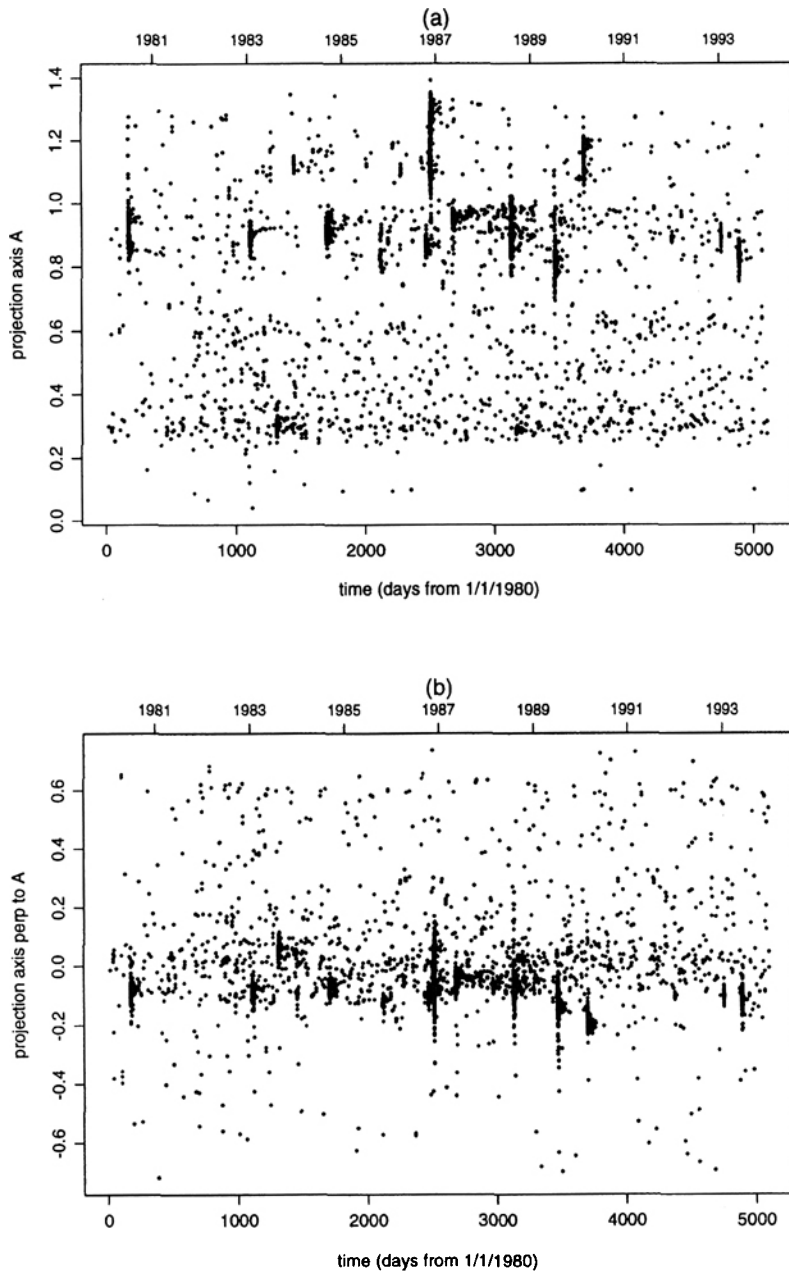
Figure 2. Space-time scatterplots of the data. Panel (a) is a scatterplot of the data binned on depth, projected onto A (so as to remove one spatial coordinate), and with the strength component suppressed. Panel (b) is the counterpart when projection is parallel to A, rather than onto A. To assist in locating position on the axis, we note that in panel (a) the points .1 and 1.3 (or equivalently, the projection axis) of Figures 2 and 3 represent the longitude-latitude pairs (138.85° E, 35.66° N) and (139.48° E, 34.64° N), respectively; while for panel (b), the points −.3 and .3 represent (138.73° E, 35.28° N) and (139.25° E, 35.60° N).
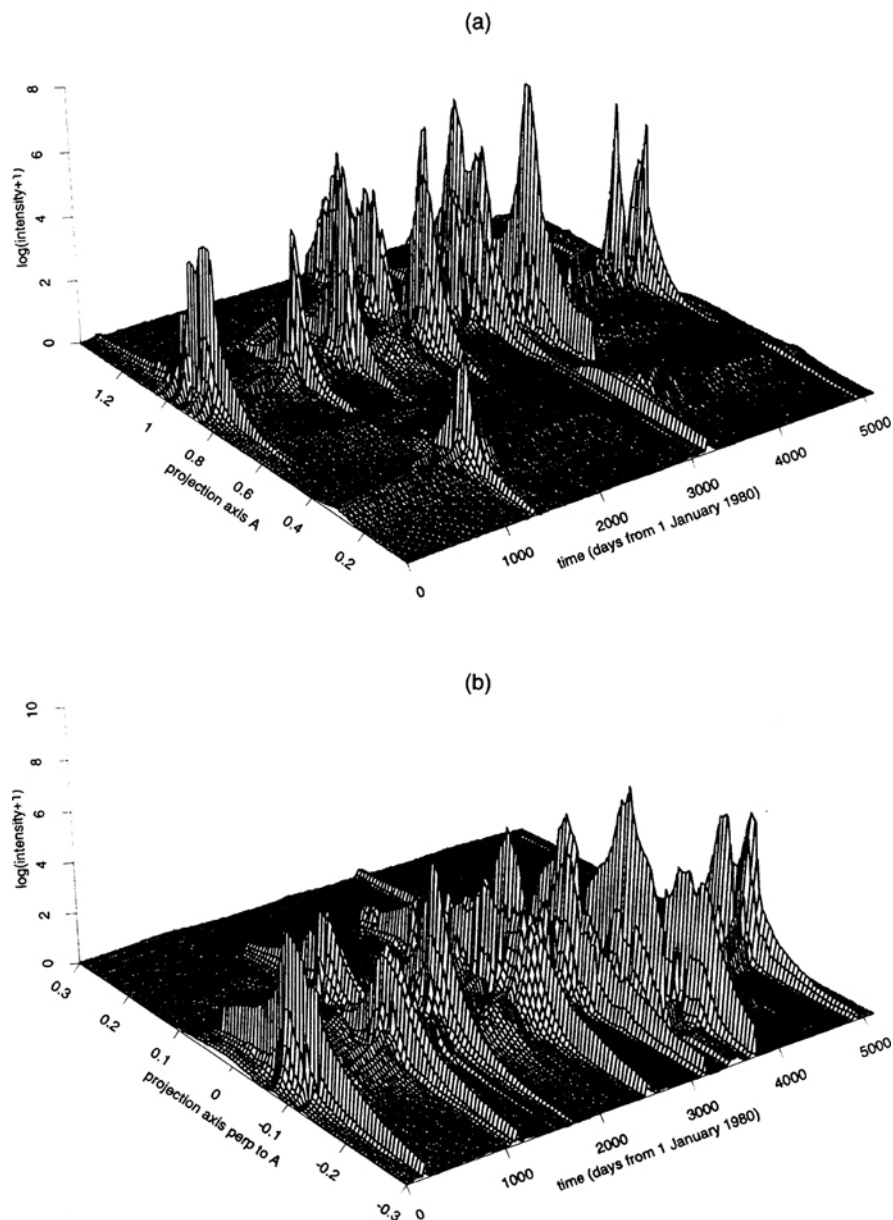
Figure 3. Perspective plots of the estimator $\hat{\lambda}$. Panels (a) and (b) show perspective plots of $\hat{\lambda}$ computed from data in panels (a) and (b), respectively, of Figure 2. (We also employed neighboring data, in order to eliminate edge effects.) See the caption of Figure 2 for coordinate details.

at least 2.0 on the Richter scale. The number of events satisfying these conditions was 8,187. Basic descriptive plots of the data are given in Figure 1. Note in particular the straight line fit in panel (c), which indicates that the Gutenberg–Richter (1944) relation holds for these data. That relation states that the density of events in a given region decreases exponentially with increasing magnitude.

3.2 EXPLORATORY ANALYSIS OF EVOLUTION IN SPACE AND TIME

To aid analyzing and visualizing the intensity of the seismic data, we employed the dimension reduction technique mentioned in Section 2.2. The axis $\mathcal{A}$ was defined to be the line that passes through the longitude-latitude pairs (139.5°E, 34.6°N) and (138.8°E, 35.75°N). It is marked on panel (a) of Figure 1, and represents a linear approximation to the volcanic front of the Izu-Bonin Arc (Koyama 1993). We argue that this is an appropriate choice of projection axis, since this volcanic front is a known source of seismic activity. Time was measured in days, starting on January 1, 1980, and ending on December 31, 1993.

Spatial data were first projected onto $\mathcal{A}$, and then projected parallel to $\mathcal{A}$. Panels (a) and (b) of Figure 2 depict the space-time pairs (X_i, T_i) after projection onto and parallel to $\mathcal{A}$, respectively. Panels (a) and (b) of Figure 3 show the respective versions of $\hat{\lambda}(x, t)$, defined at (2.1). We experimented with a large number of values α in the interval $[.25\alpha_0, 8\alpha_0]$, where $\alpha_0 = (\max_{i,j} |X_i - X_j|)/(\max_{i,j} |T_i - T_j|)$ denotes the ratio of the maximum values of spatial and temporal ranges. Similar intensity plots were obtained for α's in this interval; Figure 3 is drawn for $\alpha = \alpha_0$.

The near neighbor smoothing method captures variations in curvature particularly well, reproducing both sharp peaks that might be theoretically modeled by pole lines, and more subtle variation in regions of low intensity. Even by itself, without reference to data on strength, Figure 3 is revealing. For example, patterns of spatial variation of events that occur close together in time are already evident. Moreover, a small number of shocks that are localized relatively little in time, but are relatively highly localized in space, can be seen to presage a very large number of shocks that are highly localized in time but are not at all localized in space.

Small "buttresses" leading up to the left-hand sides of the parallel "sheets" in Figure 3(b) indicate the spatial locations of the small number of events that are a prelude to a much larger number of events at other places. The buttresses are also present on the opposite side of the sheets, but they generally decrease more gradually there. Hence, while there is a degree of symmetry about the buildup to and decay from periods of intense activity, the decay tends to be more gradual (in time) than the buildup, even though it involves activity at much the same places. Our nonparametric methods offer a particularly simple way of accessing and depicting this information.

Fourteen parallel sheets are identifiable in Figure 3(b), and all have counterparts in Figure 3(a). See Section 3.4 for further analysis of these 14 high intensity clusters of events.

When constructing $\hat{\lambda}(x, t)$ we took K_1 to be the biweight kernel, $K_1(u) = \frac{15}{16}(1 - u^2)^2 I(|u| < 1)$, and estimated intensity on a 201×201 grid. We chose $k = 50$ in our
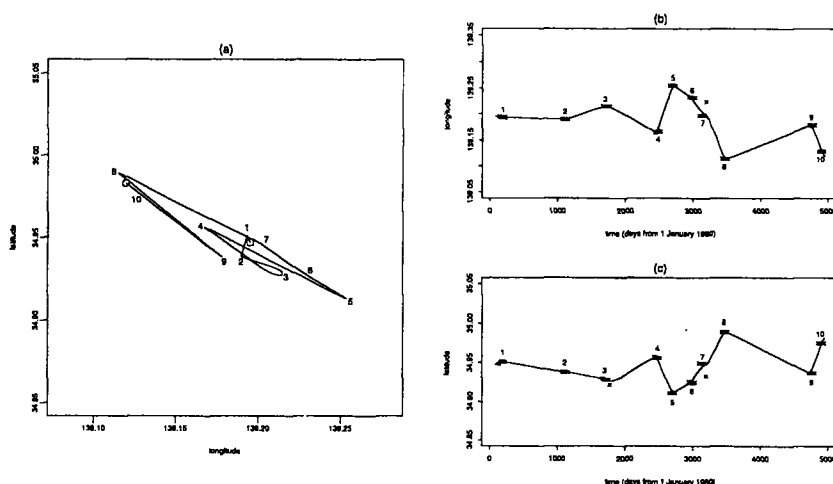
Figure 4. Spatial migration of peak intensity within $\mathcal{R}'$. Panel (a) depicts spatial movement of peak intensity, in the direction indicated by increasing numbers from 1 to 10. Panels (b) and (c) give plots of longitude and latitude components of the peak intensity estimates, respectively.

studies, although our estimators did not change appreciably with k, and similar space-time intensity estimators were obtained for values of k up to 150. If α is increased, then the main effect is to increase the relative amount of smoothing in the spatial coordinate.

3.3 Spatial Migration of Peak Intensity With Time

When estimating the spatial trajectory of peak intensity as a function of time we first computed the estimator $\hat{\mu}(y \mid t)$, a function of $y \in \mathcal{R}$, at 301 equally spaced values t in the interval $[\min_j T_j, \max_j T_j] = [21.48, 5107.81]$. The kernel K_2 at (2.2) was taken to be the bivariate form of the biweight kernel, $K_2(y) = 3\pi^{-1}(1 - y^T y)^2 \, I(y^T y < 1)$, and we used $h_2 = 50$ and and $h_3 = .1$. Broadly similar results were obtained using different kernels and bandwidths.

Let us first consider the region $\mathcal{R}' \subseteq \mathcal{R}$, bounded by 139.0° and 139.4° longitude and 34.8° and 35.1° latitude. Then, $\mathcal{R}'$ contains the majority of the points in the dense cluster in the lower half of panel (a) of Figure 1. The total number of points in $\mathcal{R}'$ is 5,654. Figure 4(a) shows the trajectory of peak intensity in $\mathcal{R}'$, with increasing time corresponding to increasing values of the numbers 1 to 10. Panels (b) and (c) depict the longitude and latitude components, respectively, of the trajectory. The crosses in both panels represent those among the 301 discrete values of t where estimated peak intensity exceeded the threshold $\tau = 50$, and of course the numbers in panels (a)–(c) represent respective points in time. The dates corresponding to greatest magnitude in each of the 10 clusters of crosses are, respectively, June 30, 1980; January 20, 1983; September 5, 1984; October 13, 1986; May 11, 1987; February 20, 1988; August 2, 1988; July 9, 1989; January 10, 1993; and May 31, 1983.

The solid lines were obtained by employing modified local linear smoothing techniques, suggested by Hall and Turlach (1997) for guarding against data sparseness prob-
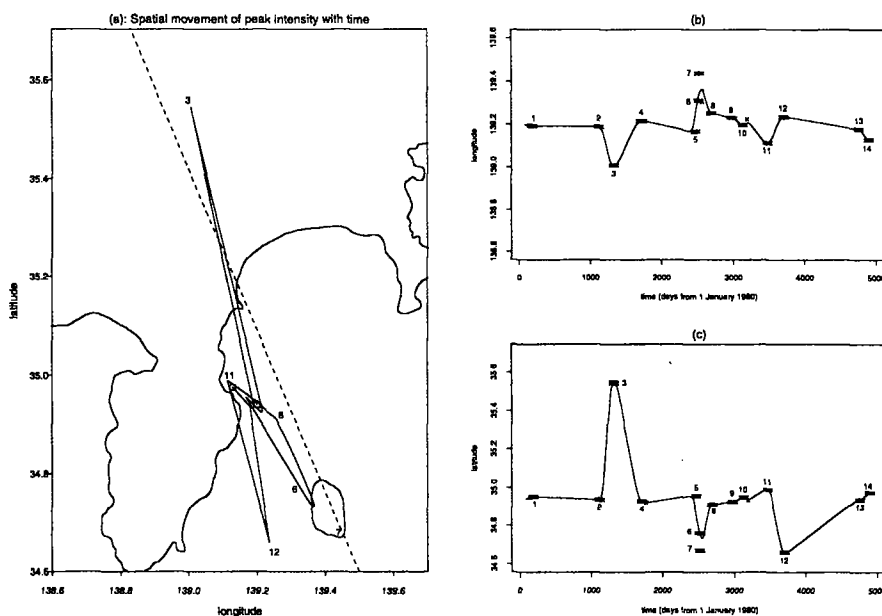
Figure 5. Spatial migration of peak intensity within $\mathcal{R}$. Details are as for Figure 4, except of course that migration within the whole region $\mathcal{R}$ is being addressed, and there are now 14 (rather than 10) clusters of time points where intensity exceeded 50.

lems. We used their $\nu = 4$ interpolation rule, and the biweight kernel, for constructing our local linear estimator. The bandwidth was 200. The continuous plot of the peak intensity shown in panel (a) was obtained from the longitude and latitude components in panels (b) and (c).

Note that peak intensity migrates along an axis which, like the axis $\mathcal{A}$, is inclined to the left, but at a shallower angle. This reflects the shallower inclination of the dense point cloud in Figure 1(a).

Figure 5 is the analog of Figure 4 for the full dataset, represented by all the points in $\mathcal{R}$. (Panel (a) of Figure 4 is in effect reproduced within panel (a) of Figure 5, except that there are excursions from points in $\mathcal{R}'$ to points outside, and vice versa, during the trajectory.) The location of peak intensity is now a highly erratic function. Overall, movement closely parallels the axis $\mathcal{A}$ (and so too the Izu-Bonin Arc), since the off axis activity within region $\mathcal{R}'$ is confined to only a subset of the data. An additional four clusters are now present. The dates that correspond to their maximum magnitudes are August 8, 1983; November 21, 1986; November 22, 1986; and February 20 1990. There is anomalous behavior among 3 of the 14 clusters, owing to spatial movement among different locations at closely spaced times. We have not attempted to represent this on the figure.
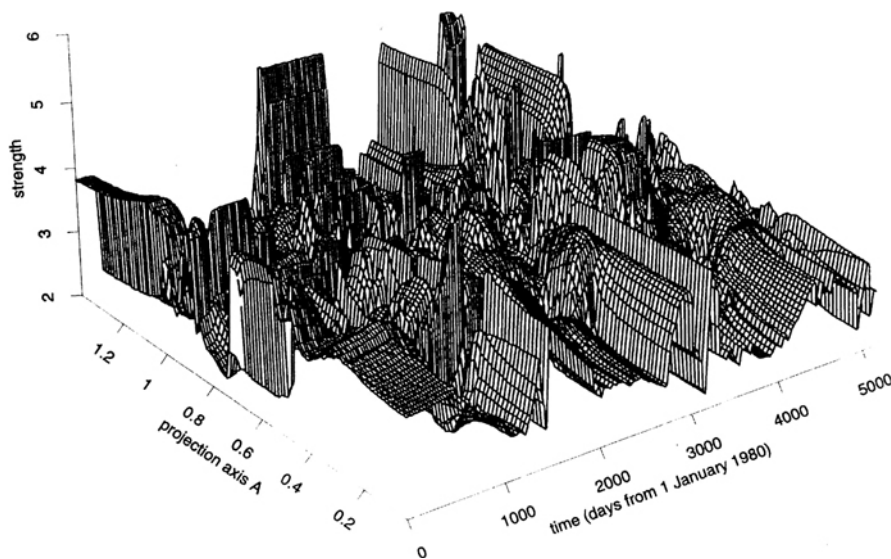
Figure 6.   Perspective plot of strength estimates. The figure shows a plot of $\hat{g}(x,t)$, when strength is defined in terms of magnitude (on the Richter scale), x denotes distance along the axis $\mathcal{A}$, and data were drawn from the region $\mathcal{R}$. Thus, the horizontal axes have the same meanings as those of Figure 3(b).

3.4   STRENGTHS OF EVENTS

Event strength may be defined in a number of different ways. Examples include the total amount of energy or stress released, the total magnitude (on the Richter scale), and the maximum magnitude. The relationship between energy (E) and magnitude (M) has been determined empirically, and is given by

$$\log_{10} E = c_1 + 1.5M,$$

where c_1 is a constant. See, for example, Kasahara (1981, p. 15ff). Since the released stress, S, is proportional to the square root of energy (see Lomnitz 1974, p. 55ff), then stress and magnitude are related by $\log_{10} S = c_2 + .75M$ for a constant c_2. Therefore, if we define energy and stress of a magnitude-M event, *relative* to their values in the case of a prescribed cutoff magnitude M^0, then they are given by $E = 10^{1.5(M-M^0)}$ and $S = 10^{0.75(M-M^0)}$, respectively.

If M_i denotes the magnitude of the event observed at time T_i, then we may write G_i for any one of M_i, $E_i = 10^{1.5(M_i-M^0)}$ and $S_i = 10^{.75(M_i-M^0)}$, and estimate the expected value of strength by substituting into formula (2.3). We suggest computing $\hat{g}$ for the case of magnitude, however, and transforming to any other strength measure in the obvious way.

Figure 6 was computed in this manner. It is a space-time plot of magnitude, and is analogous to the space-time intensity plot in Figure 3(a). In particular, the region used was $\mathcal{R}$, with other data employed only to overcome edge effects. A number of interesting features emerge simply by noting the differences between Figures 3(a) and 6. For example, although the intensity of seismic activity usually declines rapidly (in time)

after an initial spurt, this is not always true of strength, which can either remain at a reasonably high level for a period, or decline as rapidly as intensity.

Moreover, there are occasional events of relatively high strength that are not associated spatially or temporally with places where the intensity of events is high. The top right corner of Figure 6 shows an example of this phenomenon. (The associated events are the result of high magnitude but low intensity volcanic activity near the island of Ō-shima, in the lower right of Figure 1(a).) On the other hand, there are places of high intensity which, while they presage sequences of events of quite high intensity, are not themselves associated with events whose expected strength is high.

Thus, several different classes of events are clearly identifiable by comparing the information in Figures 3 and 6. While this can be deduced in other ways, it is particularly transparent from nonparametric analysis.

The main features of Figure 6 were not altered by even substantial changes to the smoothing parameters k and α (e.g., doubling or halving their values), indicating that the results are unlikely to be artifacts of our method.

3.5   ESTIMATES OF ACTIVE AND QUIESCENT PERIODS

Using data within the region $\mathcal{R}'$ we constructed $\hat{\mu}_m$ as at (2.4), and then computed $\hat{\gamma}_m(t)$ on a fine grid of values t. The pairs $(t, \log[1 + \hat{\gamma}_m(t)])$ were then smoothed using local linear methods. (The logarithmic transform was found convenient for "damping down" the considerable variability in values of $\hat{\gamma}_m(t)$.) We only consider here estimates of quiescent periods.

Figure 7(a) plots the smoothed logarithmic peak intensity estimates $\log[1 + \hat{\gamma}_m(t)]$, when the threshold magnitude m was chosen to be 2.0. We experimented with values of m up to 3.5, and found that our methods were robust against these choices. Quiescent periods were defined as those time intervals for which $\hat{\gamma}_m(t) \leq \gamma_1$, where $\gamma_1 = e^{4.0} - 1$. Although choice of γ_1 is subjective, it is our experience that, for a given value of m, there is normally a range of γ_1 which would produce essentially the same features in a quiescent period plot (see Figure 7(b)). For the values that we employed, we noted that any $\gamma_1 \in [e^{2.0} - 1, e^{5.0} - 1]$ results in a quiescent plot having similar features.

Figure 7(b) depicts estimated quiescent periods, indicated by shaded regions. On the same figure we overlay a time-magnitude plot of events that have magnitude greater than 4.0. Note particularly that each quiescent period contains none of these events, but is quickly followed by one or more such events. This is consistent with the theory that events of magnitude 4 or more generally result from the release of energy built up during quiescent periods. It suggests forecasting, immediately following the end of a quiescent period, that an event of magnitude greater than 4 is imminent. Moreover, using a model for the distribution of the lengths of quiescent periods we may calculate the probability that the next magnitude 4 event will occur within a given time interval. The distribution of lengths of quiescent periods is approximately exponential with "intensity" (inverse of the mean) 1.25. However, Figure 7(b) strongly suggests that lengths of successive quiescent periods are not independent; note that the four shortest periods are consecutive.

Our definition of quiescent periods depends largely on the estimated spatial intensity
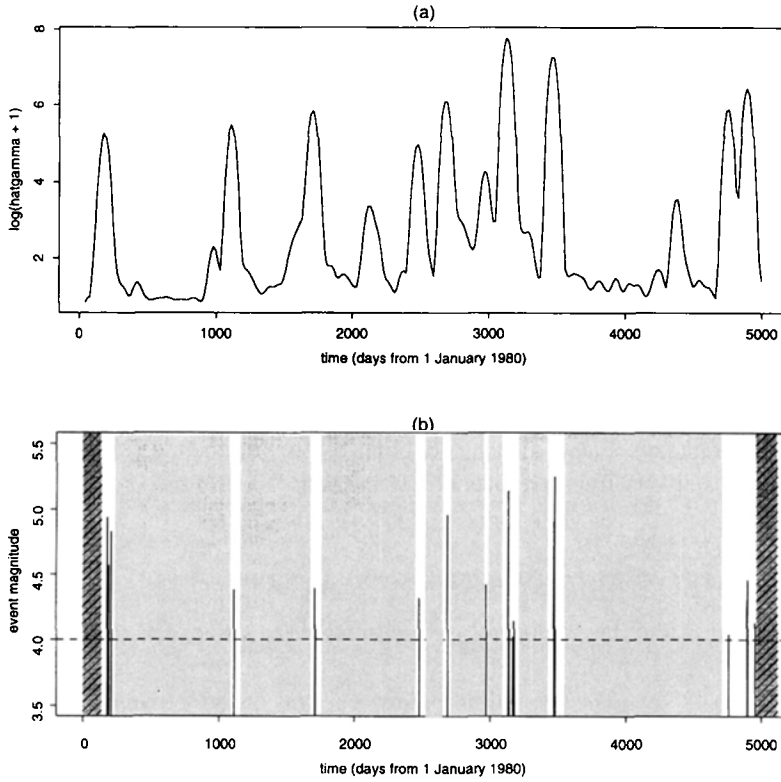
Figure 7. Quiescent periods. Panel (a) plots $\log[1+\widehat{\gamma}_m(t)]$ against time t, where the threshold magnitude was chosen as $m = 2.0$. Estimated quiescent periods are shown by shaded intervals in panel (b). The first and the last estimated quiescent periods are incomplete. The vertical spikes represent events with magnitudes greater than 4.0.

of events defined at (2.4). It is possible, however, to take into account the strengths of individual events. To this end, we suggest two definitions of intensity which can be used in place of (2.4):

$$\hat{\mu}_{m,M}(y|\,t) \;=\; \frac{1}{h_2 h_3^2} \sum_{i:T_i \in \mathcal{T}(t)} K_2\left(\frac{Y_i - y}{h_3}\right) K_1\left(\frac{T_i - t}{h_2}\right) M_i\, I(M_i \geq m)\,.$$

$$\hat{\mu}_{m,E}(y|\,t) \;=\; \frac{1}{h_2 h_3^2} \sum_{i:T_i \in \mathcal{T}(t)} K_1\left(\frac{T_i - t}{h_2}\right) K_2\left(\frac{Y_i - y}{h_3}\right) E_i\, I(M_i \geq m)\,, \quad (3.1)$$

where E_i is defined as $10^{1.5(M_i - 2.0)}$. Our analysis indicated that using $\hat{\mu}_m$, $\hat{\mu}_{m,M}$ or $\hat{\mu}_{m,E}$ to produce Figure 7(b) gives similar results. Nevertheless, $\hat{\mu}_{m,E}$ puts more emphasis on event magnitude, and has the advantage of being sensitive to strong events within a region which may have relatively low intensity aftershocks.

4. CONCLUSIONS

We have introduced nonparametric approaches to analyzing space-time data. Our techniques exploit methods of kernel smoothing, and we have illustrated them using Kanto seismic data.

As we have argued, a major advantage of nonparametric methods is that they do not require formulation of structural models, which are often well suited only to data that have closely related seismic causes. This can require that the spatial and temporal dimensions of a dataset be relatively narrow, or that anomalous events be excluded in other ways. These problems can lead to poorly fitting parametric models, for example where the models fail to reflect observed features of a dataset, or where the fitted parameters lie outside physically plausible ranges. Even if the ultimate analysis will be by parametric methods, nonparametric techniques are invaluable as tools for exploratory or preliminary analysis.

APPENDIX: PROPERTIES OF $\hat{\lambda}$, $\hat{g}$, $\hat{\mu}$, AND $\hat{\mu}_m$

Let us assume that the point process representing the bivariate projection data (X_i, T_i) is a Poisson cluster process in $\mathbb{R}^2$, and that it has intensity $\lambda = \nu\ell$, where ℓ is a fixed intensity function and ν is a scalar, the value of which we shall allow to diverge in our asymptotic analysis. If, in the neighborhood of the point where we are estimating λ, ℓ is bounded and has two bounded derivatives with respect to both coordinates, and if $h_j = h_j(\nu)$ satisfies $h_j \to 0$ and $\nu h_1 h_2 \to \infty$ as $\nu \to \infty$, then the bias of $\hat{\lambda}$ is of size $\nu (h_1^2 + h_2^2)$, and the variance is of size $\nu/h_1 h_2$. Therefore, optimal mean square performance is achieved when h_1, h_2 are both of size $\nu^{-1/6}$.

The near neighbor "coupling" method used to select h_1 and h_2 amounts, in asymptotic terms, to choosing h_1 and h_2 so that $h_1 = \alpha h_2$ and $\nu h_2^2 \alpha\ell(x, t) \approx k$. Thus, noting the optimal sizes of h_1 and h_2 suggested earlier, we see that k should be of size $\nu^{2/3}$ as ν diverges.

The sizes of bias and variance of the estimator $\hat{g}$ may similarly be shown to equal $h_1^2 + h_2^2$ and $(\nu h_1 h_2)^{-1}$, respectively. Again, asymptotic optimality is achieved when h_1, h_2 are both of size $\nu^{-1/6}$. For both $\hat{\mu}$ and $\hat{\mu}_m$ the sizes of bias and variance are respectively $\nu(h_2^2 + h_3^2)$ and $\nu/h_2 h_3^2$, and optimal choices of bandwidth are $\nu^{-1/7}$ (for both h_2 and h_3).

ACKNOWLEDGMENTS

REFERENCES

Fan, J. (1993), "Local Linear Regression Smoothers and Their Minimax Efficiencies," *The Annals of Statistics*, 21, 196–216.

Gutenberg, R., and Richter, C. F. (1944), "Frequency of Earthquakes in California," *Bulletin of the Seismic Society of America*, 34, 185–188.

Hall, P., and Turlach, B. A. (1997), "Interpolation Methods for Adapting to Sparse Design in Nonparametric Regression" (with discussion), *Journal of the American Statistical Association*, 92, 466–476.

Hawkes, A. G. (1971), "Point Spectra of Some Mutually Exciting Point Processes," *Journal of the Royal Statistical Society*, Ser. B., 33, 438–443.

Kasahara, K. (1981), *Earthquake Mechanics* , Cambridge, UK: Cambridge University Press.

Koyama, M. (1993), "Volcanism and Tectonics of the Izu Peninsula, Japan" (in Japanese), *Kagaku (Science)*. 63, 312–321.

Lomnitz, C. (1974), *Global Tectonics and Earthquake Risk*, Amsterdam: Elsevier.

Mogi, K. (1968), "Some Features of Recent Seismic Activity in and near Japan," *Bulletin of the Earthquake Research Institute, Tokyo University*, 46, 1225–1236.

Ogata, Y. (1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9–27.

—— (1989), "Statistical Model for Standard Seismicity and Detection of Anomalies by Residual Analysis," *Tectonophysics*, 169, 159–174.

—— (1992), "Detection of Precursory Relative Quiescence Before Great Earthquakes Through a Statistical Model," *Journal of Geophysical Research*, 97, 19845–19871.

—— (1998), "Space-Time Point-Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402.

Silverman, B. W. (1986), *Density Estimation For Statistics and Data Analysis*, London: Chapman and Hall.

Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.

Zhuang, J. (1998), "Statistical Modelling of Seismicity Patterns Before and After the Cape Palliser Earthquake," Technical Report No. 47, Institute of Statistics and Operations Research, Victoria University of Wellington, New Zealand.