

NONPARAMETRIC ANALYSIS OF EARTHQUAKE POINT-PROCESS DATA

EDWIN CHOI¹ AND PETER HALL¹

Centre for Mathematics and its Applications, ANU

Motivated by multivariate data on epicentres of earthquakes, we suggest nonparametric methods for analysis of point-process data. Our methods are based partly on nonparametric intensity estimation, and involve techniques for dimension reduction and for mapping the trajectory of temporal evolution of high-intensity clusters. They include ways of improving statistical performance by data sharpening, i.e. data pre-processing before substitution into a conventional nonparametric estimator. We argue that the ‘true’ intensity function is often best modelled as a surface with infinite poles or pole lines, and so conventional methods for bandwidth choice can be inappropriate. The relative severity of a cluster of events may be characterised in terms of the rate of asymptotic approach to a pole. The rate is directly connected to the correlation dimension of the point process, and may be estimated nonparametrically or semiparametrically.

AMS subject classifications: Primary 62G05, 62G07; Secondary 63M30.

Keywords and phrases: Bandwidth, bias reduction, correlation dimension, data sharpening, density estimation, epicentre, geophysics, intensity, Japan, Kanto, kernel methods, magnitude, pole, regular variation, smoothing.

1 Introduction

An earthquake-process dataset may often be interpreted as a realisation of a 5-dimensional point process, where the first three, spatial components denote latitude, longitude and depth below the earth’s surface, the fourth represents time, and the fifth is a measure of ‘magnitude’, for example on the Richter scale. Goals of analysis can be very wide-ranging. At one level they may be purely descriptive, perhaps summarising features of the dataset. In this regard, some form of dimension reduction is often critical, putting the information on five dimensions into a form that is more readily accessible and interpretable. At another level the goals may be exploratory, suggesting directions for future analysis, or they may be more explicit and detailed, perhaps with the aim of elucidating properties of subterranean features that played a role in generating the data.

¹ Supported by the Australian Research Council.

In this paper we discuss nonparametric methods for summarising earthquake data, for exploring the main features of the data, and for addressing more structural problems such as the location of poles and pole lines, the way in which those poles migrate with time, and the value of the correlation dimension of clusters of epicentres. (Poles and pole lines are points and line segments, respectively, at which the intensity of the point process asymptotes to infinity.) Many of our arguments are based on kernel-type estimators of intensity, while others employ methods that are parametric in simple cases but are nevertheless valid in contexts which are quite distant from the parametric model. The aim is to develop analytical tools that offer greater diversity, and robustness against departures from structural models, than more traditional parametric approaches. The latter include the popular Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988), which is used to describe temporal behaviour of an earthquake series; and refinements of Hawkes' (1971) self exciting point process model, which describe spatial-temporal patterns in a catalogue. The paper by Ogata (1998) gives detailed discussion of recent extensions of these models.

Disadvantages of parametric models in this setting include their instability when even small amounts of new data are added, and their relative insensitivity to anomalous events, arising from the fact that models tend to be formulated through experience of relatively conventional earthquake activity. Indeed, anomalies are typically 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), procedures that tend to conceal anomalies are not necessarily to be preferred.

Figure 1 depicts spatial components of the type of data that motivate this paper. They are part of the 'Kanto earthquake catalogue', and were compiled by the Centre for Disaster Prevention at Tsukuba, Japan. The points are longitude-latitude pairs representing the locations of earthquakes that occurred in the region of Kanto, Japan, between 1980 and 1993. We have restricted attention here to events whose location was between 138.6° and 139.7° longitude and 34.6° and 35.7° latitude, whose depth was less than 36 km, and whose magnitude was at least 2.0 on the Richter scale. There are 8187 points in the dataset. The diagonal line on the figure is a linear approximation to the location of the volcanic front of the Izu-Bonin Arc (Koyama, 1993), which is a known source of earthquake activity. The region with a dotted boundary defines a smaller subset, near the island of Ōshima, which will also feature in our analysis.

Section 2 describes methods for intensity estimation based on point process data, and outlines applications to which such estimates may be put. Techniques for enhancing multivariate intensity estimates, and for deducing structure from them, are outlined in Section 3. Section 4 introduces methods

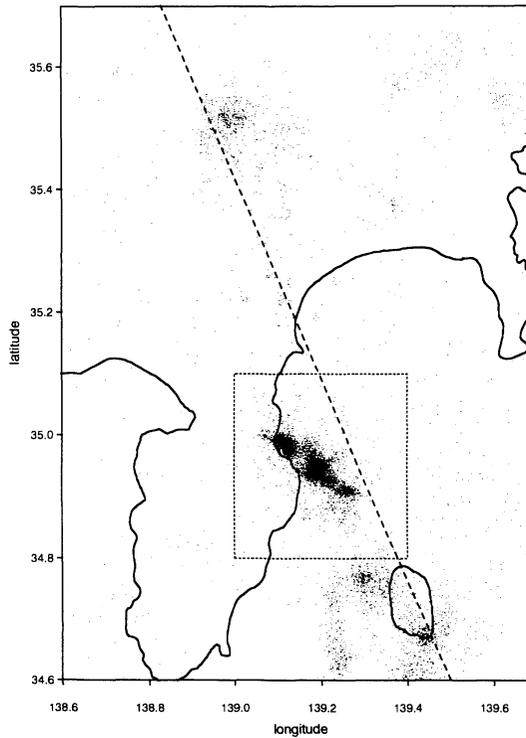


Figure 1. Spatial coordinates of Kanto earthquake data. Data in the smaller region, indicated by the dotted-line boundary, will be used for analysis described in section 2.3. The dashed line diagonally across the figure represents a linear approximation, \mathcal{A} , to the volcanic front of the Izu-Bonin Arc.

for estimating the locations and strengths of poles in intensity functions.

Some discussion of use of the term ‘magnitude’, and of the ‘Richter scale’, is in order. The many different measures of magnitude include those based respectively on energy and on different measures of the amplitudes of shock waves produced by an earthquake. Local Magnitude, more popularly referred to as Richter Magnitude, is of the latter type and is representable in terms of the logarithm of the maximum trace amplitude, measured in micrometers on a standardised seismometer. The magnitude to which we refer in this paper is Local Magnitude, although we shall henceforth call it, and the scale on which it is measured, by its popular name.

2 Data summarisation and exploration

2.1 Dimension reduction

Information about depth in a seismic data vector is often not particularly

accurate, and for example is typically represented in bins up to 10 kilometres wide. Reflecting this difficulty, we suggest pooling bins. The longitude and latitude components too are recorded with varying degrees of error, which depend on, among other matters, the spatial distribution of recording stations around the location of the event, and event depth. We shall not attempt to employ such information in our analysis — it is sometimes explicitly available (see e.g. Jones and Stewart, 1997), or deducible from other measurements — but it can be incorporated.

Even after removing the depth dimension, data vectors can have as many as four components. We suggest looking at the two remaining spatial components separately, by projecting longitude-latitude pairs onto first one axis and then another. An appropriate axis is often clear from physical considerations; see Figure 1. Neglecting the magnitude component for the time being, we now have two bivariate datasets where in each case one component represents time and the other is a spatial coordinate. Each may be used to produce nonparametric estimates of point-process intensity, enabling perspective plots (where the third dimension represents intensity) to be produced.

Of course, contemporary dimension-reduction methods, such as projection pursuit, might also be used to determine projections in the continuum that maximise the ‘interestingness’ of the associated bivariate scatterplots. In their full generality, such approaches can be hard to justify in the present setting, since rotations of axes that are as distinct as time and space are difficult to interpret. Even if dimension reduction is contemplated only for the spatial coordinates, physical interpretation can sometimes be facilitated by using information from outside the dataset (for example, in the case of the Kanto data, the physically-meaningful Izu-Bonin Arc) to determine an appropriate axis.

Magnitude may be depicted by adding colour or a grey shade to graphs of estimated intensity. Therefore, magnitude can be included on the plots described above, without increasing the complexity of the set of projections. It can be shown separately, however, in plots broadly similar to those for intensity. Since magnitude is recorded with error, and only at scattered points in space and time, it is generally necessary to smooth magnitude measurements.

2.2 Kernel estimation of intensity

If space-time data pairs (X_i, T_i) are available after projection of spatial coordinates onto an axis, then the space-time intensity per unit area at (x, t) is estimated by

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

where K_1 is a univariate kernel, usually taken to be a symmetric probability density function, and h_1, h_2 are bandwidths for spatial and temporal components, respectively. A local method for bandwidth choice is critical when estimating the intensity of point processes related to earthquakes, since in many cases there is evidence that the intensity function is neither bounded nor square-integrable. Estimators of the type at (2.1) have been discussed by, for example, Wand and Jones (1995, p. 167f). They were introduced by Ramlau-Hansen (1983) and Diggle (1985). Related methods in a geophysical context have been discussed by Davis and Frohlich (1991).

The panels in Figure 2 are intensity estimates computed using (2.1). We employed the biweight kernel. Bandwidth was chosen using the following ‘coupled’ near-neighbour method. We chose h_1 and h_2 to minimise $h_1 + \alpha h_2$, subject to the rectangle $(x - h_1, x + h_1) \times (t - h_2, t + h_2)$ containing at least k space-time pairs (X_i, T_i) . The value of k controls the overall level of smoothing, with larger k ’s providing a greater amount of smoothing. The relative emphasis placed on spatial and temporal coordinates is governed by α . In particular, if α is chosen small then the resulting rectangle is elongated along the time axis, and so the intensity estimator is localised more in space than in time.

We took α to equal the ratio of the ranges of the spatial and temporal data sets, although substantial changes to the values of k and α (e.g. doubling or halving their values) do not alter the main features of Figure 2. Longitude and latitude components of the original dataset are depicted in Figure 1. The straight line there (denoted here by \mathcal{A}) was the axis onto which spatial components were projected for the first panel. For the second panel, spatial projection was in the perpendicular direction.

Even by itself, without reference to data on magnitude, Figure 2 is revealing. For example, patterns of spatial variation of events that occur close together in time are clearly evident. Similar plots may be derived when event magnitude, rather than intensity, is featured on the vertical scale. A comparison of intensity and magnitude plots can provide a particularly informative description of the way in which a cluster of earthquakes develops in time and space. The relationship between intensity, and magnitude, space-time surfaces is complex, and far from being proportionate. It tends more towards being inversely proportionate, with a high frequency of relatively low-magnitude events being similar in some respects to a low frequency of high-magnitude ones. However, even this is a significant oversimplification. Further details are available from figures in Choi and Hall (1998).

A reader familiar with parametric analysis of earthquake data might query our use of a symmetric kernel in at least the temporal component at (2.1). In particular, it is known that earthquake processes are quite asymmetric in time, with few foreshocks and many aftershocks; see for ex-

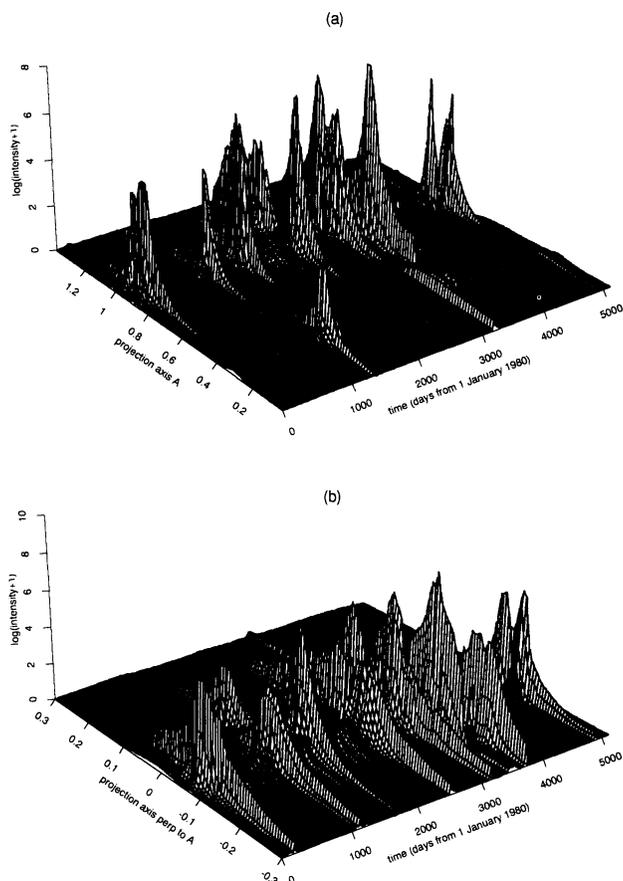


Figure 2. Perspective plots of intensity estimates. Panels (a) and (b) show plots of space-time intensity when spatial components are projected onto and perpendicularly to, respectively, the diagonal line \mathcal{A} in Figure 1.

ample Kagan (1994). A virtue of nonparametric analysis is that it usually adapts well to asymmetries of this type, however. In a related setting, one would seldom have qualms about using a symmetric kernel to estimate an asymmetric probability density, and similar arguments apply here.

2.3 Spatial migration of peak intensity

As Figure 2 suggests, intensity increases rapidly, in both temporal and spatial terms, to a peak that could be interpreted as a pole. The temporal trajectory of the spatial location of the pole is of particular interest. To estimate 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 utilise 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, $\mu(y|t)$, 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

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

where K_2 is a bivariate kernel and h_3 is a new bandwidth. Here, y is a column vector of length 2, and we have in effect employed a bandwidth matrix $h_3 I_2$ for K_2 , where I_2 denotes the 2×2 identity matrix. When used in the present context, near-neighbour methods tend to produce closely neighbouring, multiple peaks in places where intensity is high, and so we do recommend them. We employ instead a global bandwidth.

Once $\hat{\mu}(\cdot|t)$ has been computed we may readily calculate the spatial location of its maximum, being a bivariate function of t ; and thence we may estimate the trajectory of the maximum, again as a function of t . Since earthquake 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.

Panel (a) of Figure 3 depicts spatial migration of peak intensity within the smaller region shown in Figure 1, indicated by the dotted-line boundary there. The great majority of events in that smaller region are the result of high-magnitude but low-intensity volcanic activity near the island of Ōshima, in the lower right of Figure 1. We make this specialisation so as to restrict attention to a region where earthquake activity has a relatively homogeneous cause. Without it, a plot of peak migration is an interwoven mixture of movements representing quite different subterranean activities in different regions; over time, the trajectory jumps from one region to another, often without there being any plausible relationship between adjacent locations.

The total number of data points within the smaller region in Figure 1 is 5654. Intensity was computed using the estimator at (2.2), taking K_1 to be the biweight kernel and K_2 to be its multivariate version, with $h_2 = 50$ and $h_3 = 0.1$. Time was discretised on 301 points. Panels (b) and (c) of Figure 3 depict the longitude and latitude components, respectively, of the trajectory shown in panel (a). The crosses in both panels represent those among the 301 values of t where estimated peak intensity exceeded the threshold $\tau = 50$, and the general locations are numbered 1–10 in respective order. The points corresponding to greatest intensity in each of the 10 clusters of

crosses are respectively 30 June 1980, 20 January 1983, 5 September 1984, 13 October 1986, 11 May 1987, 20 February 1988, 2 August 1988, 9 July 1989, 10 January 1993 and 31 May 1983.

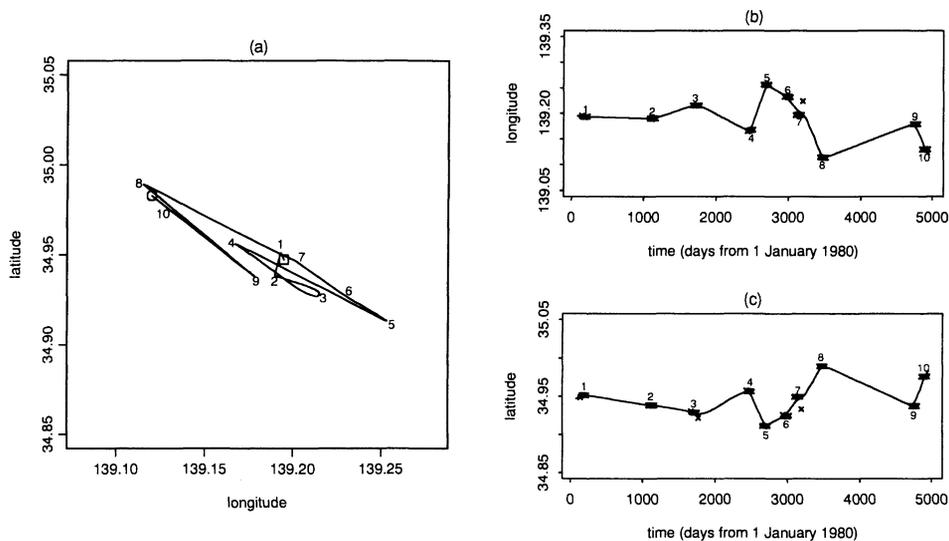


Figure 3. Spatial migration of peak intensity. Panel (a) depicts spatial movement of peak intensity, in the direction indicated by increasing numbers from 1 to 10. Panels (b) and (c) give plots, against time, of longitude and latitude components, respectively, of peak intensity estimates. Attention is confined to data in the region represented by the dotted-line boundary in Figure 1.

3 Data sharpening for intensity estimation

3.1 General principles

The term 'data sharpening' refers to methods for pre-processing data so that, when they are substituted into a conventional estimator, performance is improved relative to what would it be if the raw data were employed. The idea is to enhance performance without degrading the attractive properties that relatively simple estimators enjoy.

For example, in the context of kernel density estimation one can achieve high orders of accuracy by replacing a nonnegative kernel by one with a large number of vanishing moments. But this typically destroys positivity of the estimator, and the high-order advantages are often only available for relatively large sample sizes. By appropriately preprocessing the data before substitution into an estimator with a conventional, positive kernel, it is possible to improve accuracy yet retain the property of positivity (and the property that the estimator integrates to 1).

The methods of intensity estimation discussed in Section 2.2 have much

in common with those for nonparametric density estimation, and in particular share bias properties with that approach. They tend to underestimate intensity at peaks, and overestimate it at troughs. This is a major source of bias — it means that intensity estimators are generally biased downwards at local maxima, and upwards at local minima. If we could move data values closer together near peaks and further apart near troughs, then in large part these problems would vanish, and bias would be reduced, without us having to modify the estimator itself. Possibilities along these lines have been discussed by, for example, M.C. Jones and Signorini (1997) and R.H. Jones and Stewart (1997).

In the context of the method described in Section 2.2, this argument might be made in the following form. Consider the surface \mathcal{S} defined by the true bivariate intensity λ , of which $\hat{\lambda}$ is an estimator. Then \mathcal{S} is the locus of triples (x, t, y) defined by $y = \lambda(x, t)$. If we are at a point $P_i = (X_i, T_i)$ in the plane then, to a good approximation, we shall move across the plane in the direction of the projection of the line of steepest ascent up \mathcal{S} , if we map P_i to the value of the average of all data values in some small neighbourhood of X_i . Such a mapping will tend to move data in precisely the direction needed to overcome biasing problems associated with standard intensity estimators.

Performance of the algorithm is of course influenced by the definition taken for ‘neighbourhood’. This may be given very simply in terms of the neighbourhood used for the intensity estimator itself. For example, suppose we are using a standard kernel estimator such as that at (2.1), with bandwidth $h = h_1 = h_2$ (after rescaling) and a nonnegative, symmetric kernel K_1 , to estimate intensity; and that we wish to sharpen our data in this context. Then a canonical version of our data-sharpening transformation amounts to mapping (X_i, T_i) to the value taken by the standard multivariate Nadaraya-Watson density estimator evaluated at (X_i, T_i) , when bandwidth equals $2^{-1/2}h$ and the data are used as both exogenous and endogenous variables in the regression problem. It may be shown that this procedure reduces bias by two orders of magnitude, from $O(h^2)$, expressed relative to true intensity, to $O(h^4)$.

The key to validity of this assertion is use of the factor $2^{-1/2}$ to adjust bandwidth. Without the factor, bias may be reduced but not by an order of magnitude. The method suggested by Jones and Stewart (1997) does not employ such a factor, although the context of Jones and Stewart’s work is different from that discussed above. These authors are interested in identifying ridges in intensity estimates computed from geophysical datasets, not in intensity estimation *per se*. See Section 3.3 for further discussion.

Derivation of the bandwidth-multiplication factor will be summarised in section 3.2. Further details are available from Choi and Hall (1999b). Suffice it to say here that the factor is unrelated to the bivariate nature of

the problem, but instead depends on the order of the smoothing technique. In particular, if we were estimating an intensity in $p \geq 1$ dimensions, and employing an estimator based on nonnegative, symmetric kernels, then we would use the data-sharpening method suggested two paragraphs above, employing bandwidth $2^{-1/2}h$ in the sharpening step based on the Nadaraya-Watson estimator, regardless of the value of p . Bias would be reduced from $O(h^2)$ to $O(h^4)$. More generally, however, if we were using kernels of order r then we would replace $2^{-1/2}$ by $r^{-1/r}$. The method has simple analogues for other linear estimators, for example those based on orthogonal series or singular integrals.

3.2 Implementation

Rather than reanalyse the data addressed in Section 1 we shall introduce a new data set, this time wholly spatial rather than involving both space and time. It has the advantage of requiring relatively little variation of bandwidth. Panel (a) of Figure 4 depicts epicentres of earthquakes with magnitude at least 2 on the Richter scale, occurring between 100° and 160° longitude and -30° and 30° latitude during the years 1984 to 1995. The number of events which satisfy these criteria is 24471. Data were compiled by NOAA and USGS (1996); see also the printed account in the PDE catalogue (1997). A kernel estimator of intensity at $x = (x^{(1)}, x^{(2)})$, based on data such as these, is

$$(3.1) \quad \hat{\nu}(x) = \frac{1}{h^2} \sum_i K\left(\frac{x - X_i}{h}\right)$$

(compare (2.1)), where now $X_i = (X_i^{(1)}, X_i^{(2)})$ represents a spatial data pair, K is a bivariate kernel and h a bandwidth. The sharpened data are

$$(3.2) \quad \hat{X}_j = \frac{\sum_i X_i K\{2^{1/2}h^{-1}(X_j - X_i)\}}{\sum_i K\{2^{1/2}h^{-1}(X_j - X_i)\}},$$

where K and h are as at (3.1). The ‘diagonal’ terms may be dropped from the numerator and denominator at (3.2), without affecting first-order asymptotic properties. We suggest substituting \hat{X}_i for X_i at (3.1), to compute the sharpened form of $\hat{\nu}$:

$$\hat{\nu}_s(x) = \frac{1}{h^2} \sum_i K\left(\frac{x - \hat{X}_i}{h}\right).$$

Panels (a) and (b) of Figure 5 show $\hat{\nu}$ and $\hat{\nu}_s$, respectively, computed from the data in panel (a) of Figure 4, using bandwidth $h = 2.0$ (chosen subjectively) and taking K to be the product of two univariate biweight

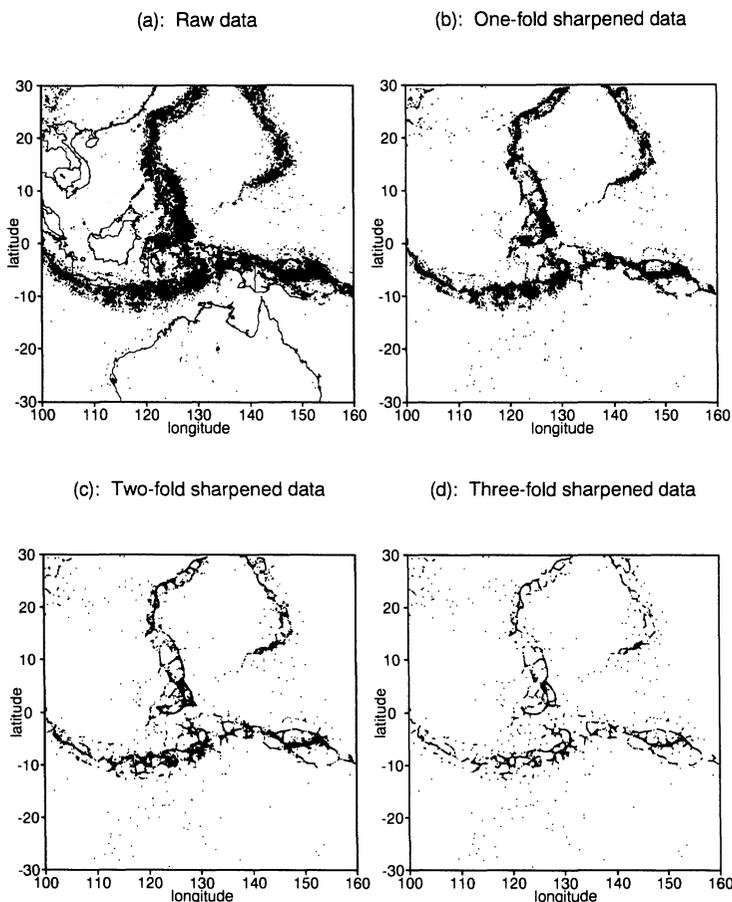


Figure 4: *Scatterplots of sharpened data with decreasing bandwidth.* Panel (a) shows the raw data, and panels (b)–(d) show the effects of applying the sharpening algorithm $m = 1$, 2 or 3 times, respectively. The bandwidth used to compute the sharpened scatterplots decreases in proportion to the inverse of the square root of m .

kernels. Visibly, the intensity peaks are depicted more sharply after sharpening. In particular, the estimated peak intensity at $(146.33^\circ, -6.50^\circ)$ of the data-sharpening intensity estimate exceeds that of the standard kernel estimate by 23%.

Finally we give a short technical argument to demonstrate the appropriateness of the bandwidth-multiplication factor, $2^{-1/2}$. Of course, we assume K is a symmetric bivariate probability density. Suppose the data X_i are p -variate and come from a distribution with density f , and consider the empirical transformation

$$\hat{\psi}(x) = x + \frac{\sum_i (X_i - x) K\{h_1^{-1}(x - X_i)\}}{\sum_i K\{h_1^{-1}(x - X_i)\}},$$

where, for the present, h_1 denotes an arbitrary constant multiple of h . If we

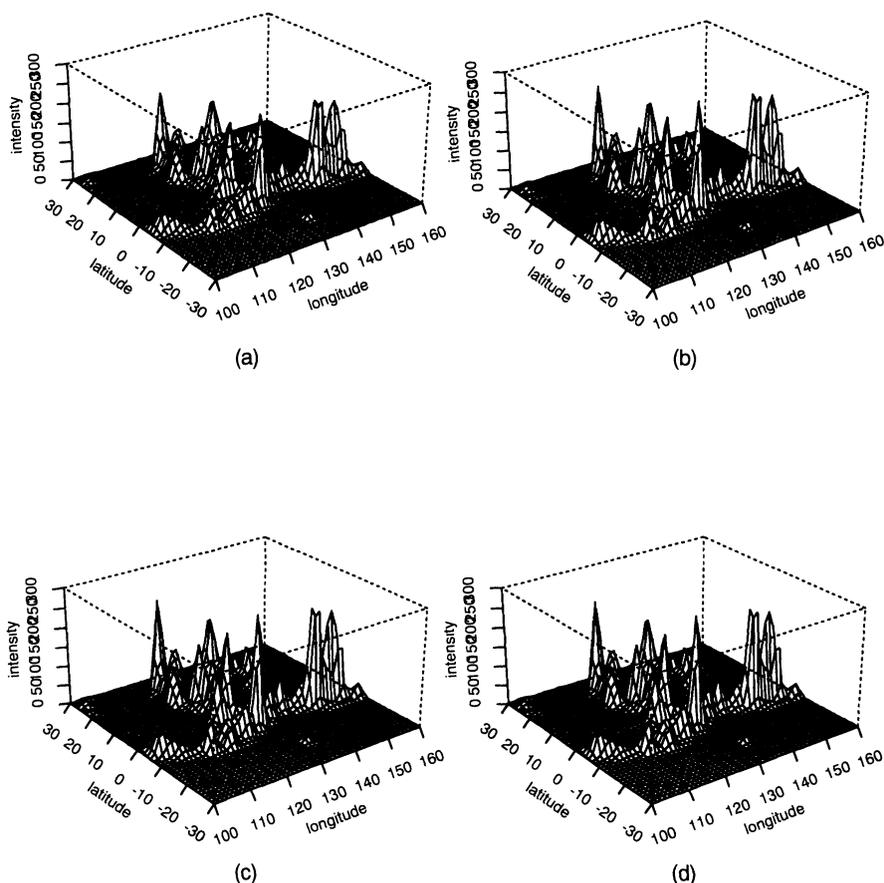


Figure 5: Intensity estimates computed from sharpened data with decreasing bandwidth. Estimates in panels (a)–(d) were computed from the scatterplots in panels (a)–(d), respectively, of Figure 4. The bandwidth used to compute each intensity estimate from the sharpened data was kept fixed at $h = 2.0$, although the bandwidths used to produce successive sharpened scatterplots were decreasing.

take that constant to be $2^{-1/2}$ then $\hat{X}_i = \hat{\psi}(X_i)$. More generally, by Taylor expansion,

$$\begin{aligned}
 E\{\hat{\psi}(x)\} &= x + \frac{h_1}{f(x)} \int u K(u) f(x + h_1 u) du + o(h^2) \\
 &= x + h_1^2 \frac{\kappa \nabla f(x)}{f(x)} + o(h^2),
 \end{aligned}$$

where $\kappa = \int (u^{(1)})^2 K(u) du$, $\nabla f = (f^{(1)}, \dots, f^{(p)})^T$, and $f^{(j)}(x)$ denotes the partial derivative of $f(x)$ with respect to $x^{(j)}$. Put $\nabla^2 f = (f^{(1)})^2 + \dots + (f^{(p)})^2$. Then, using the above expansion of $E(\hat{\psi})$, and the fact that $\hat{\psi}$ has variance of smaller order than the variance of a kernel estimator of f using

bandwidth h , we see that

$$\begin{aligned}
 E\left\{K\left(\frac{x - \widehat{X}_i}{h}\right)\right\} &= \int K\left(\frac{x - y - h_1^2 \kappa \{\nabla f(y)\} f(y)^{-1}}{h}\right) f(y) dy + o(h^2) \\
 &= \int K\left(\frac{x - y}{h}\right) f\left\{y - h_1^2 \kappa \frac{\nabla f(y)}{f(y)}\right\} dy \left\{y - h_1^2 \kappa \frac{\nabla f(y)}{f(y)}\right\} + o(h^2) \\
 &= \int K\left(\frac{x - y}{h}\right) \{f(y) - \kappa h_1^2 \nabla^2 f(y)\} dy + o(h^2) \\
 &= h \left\{f(x) + \frac{1}{2} \kappa h^2 \nabla^2 f(x) \eta - \kappa h_1^2 \nabla^2 f(x) + o(h^2)\right\} = h\{f(x) + o(h^2)\},
 \end{aligned}$$

where the last identity requires $h_1 = 2^{-1/2}h$. This confirms that using the bandwidth-multiplication factor $2^{-1/2}$ reduces bias from $O(h^2)$ to $o(h^2)$. The symmetry of K may be used to prove that bias is actually reduced to $O(h^4)$.

3.3 Iteration

There is no difficulty iterating the sharpening step. To describe the method, let $\widehat{X}_j^{[0]} = X_j$ denote a ‘raw’ data value, and consider using bandwidth h_0 rather than $2^{-1/2}h$ at (3.2). We may generalise the notion of sharpened data by defining the k -fold sharpened form of X_j by

$$(3.3) \quad \widehat{X}_j^{[k]} = \frac{\sum_i X_i^{[k-1]} K\{h_0^{-1}(\widehat{X}_j^{[k-1]} - \widehat{X}_i^{[k-1]})\}}{\sum_i K\{h_0^{-1}(\widehat{X}_j^{[k-1]} - \widehat{X}_i^{[k-1]})\}},$$

for $k \geq 1$. In order to preserve the order of bias reduction, from $O(h^2)$ to $O(h^4)$, it is necessary to change bandwidth from $2^{-1/2}h$ to $h_0 = (2m)^{-1/2}h$ if a total of m iterations is contemplated (i.e. if we employ (3.3) for $1 \leq k \leq m$, and then compute our final estimator using formula (3.1) with $\widehat{X}_i^{[m]}$ replacing X_i). Note that the order of bias reduction does not change with increasing m . More generally, we could employ $h_0 = h_k$, say, in (3.3), where h_1, \dots, h_m satisfy $h_1^2 + \dots + h_m^2 = h^2/2$.

Panels (b)–(d) of Figure 4 are scatterplots of the points $\widehat{X}_i^{[m]}$, for $m = 1, 2, 3$ respectively. They were computed using bandwidth $h_0 = (2m)^{-1/2}h$, with $h = 2.0$, in formula (3.3). The increasing amount of structure observed as m increases is the result of points X_i moving to positions closer to the projections, into the x plane, of ridge lines on the surface \widehat{S} described by the equation $y = \widehat{v}(x)$. As m increases the ridge projections become sharper, and the structure becomes more complex and ‘crinkly’, partly because of the movement of points and partly because bandwidth is decreasing.

However, the structure starts to degenerate as we increase m further. This is starting to become apparent in panel (d) of Figure 4, representing

$m = 3$. By the time $m = 5$ the ridge structure that is clearly visible for $m = 2$ has almost entirely disappeared. It is replaced by a new scatterplot of highly isolated clusters of points. On the scale of Figure 4 it looks similar to panel (a), except that only a few percent of the original data appear to remain. These ‘points’ are actually high-intensity data clusters around several hundred modes, or local maxima, the latter arising because the effective bandwidth has been substantially reduced. This tendency reverses for large m , and in fact the limit, as $m \rightarrow \infty$, of the sharpened scatterplot is the original scatterplot in panel (a) of Figure 4.

Panels (a)–(d) of Figure 5 show versions of the intensity estimate $\hat{\nu}_s$ that correspond to the scatterplots in panels (a)–(d), respectively, of Figure 4. Note that we use the same bandwidth, $h = 2.0$, for all the functions in Figure 5, although the scatterplots in Figure 4 are calculated using successively smaller bandwidths. A key feature of Figure 5 is that intensity estimates are virtually identical in each of panels (b)–(d). While the number of iterations of data sharpening (with steadily reducing bandwidth) has a substantial impact on the point process pattern (see Figure 4), it has little effect on the intensity estimates produced from the patterns, at least for the numbers of iterations employed to generate the last three panels of Figure 5.

In particular, the ratio of the height of the largest peak for a given value of m , to that when $m = 0$ (i.e. for the standard kernel estimator), increases to 1.25 when $m = 2$, and decreases only slightly for $m = 3$ and 4, always remaining above 1.20. There is some evidence of increased variability of intensity estimates for larger values of m , but the increase is substantially less than the changes in the point-process maps (see Figure 4) produced by increasing m . Simulation studies for simpler target intensities show that for fixed sample size, bias also tends to increase if m is increased beyond a certain point. Therefore, employing a high order of iteration to compute the ‘final’ estimate is not recommended.

Results are quite different if we keep bandwidth fixed while iterating the data-sharpening step. Now complexity and ‘crinkliness’ are actually reduced for relatively large numbers of iterations. See Figure 6. However, successively less information is available in parts of the plane where the ‘true’ intensity is not high. As a result, if we compute the intensity estimator $\hat{\nu}$ from datasets that have been successively sharpened in this way, its shape alters rapidly as the number of iterations increases, and overall performance deteriorates.

This is clear from Figure 7. There, the local maxima of the intensity estimate become progressively more pronounced as the number of iterations is increased. Indeed, the growth is so rapid that we have had to truncate height at 300 in order to present panels (c) and (d) of Figure 7; the heights of the largest peaks for those datasets are actually 360 and 430, respectively. In

this setting, where bandwidth is kept fixed in the data-sharpening step, the ratio of the height of the largest peak for a given value of m , to that when $m = 0$ (i.e. the standard kernel estimator), increases monotonically with m , being 1.3, 1.6, 1.9 and 2.0 when $m = 1, 2, 3, 4$ respectively. By way of comparison, when bandwidth is reduced as m increases, the corresponding ratio at first increases and then, for the last two panels, decreases only slightly (and monotonically); see Figure 5.

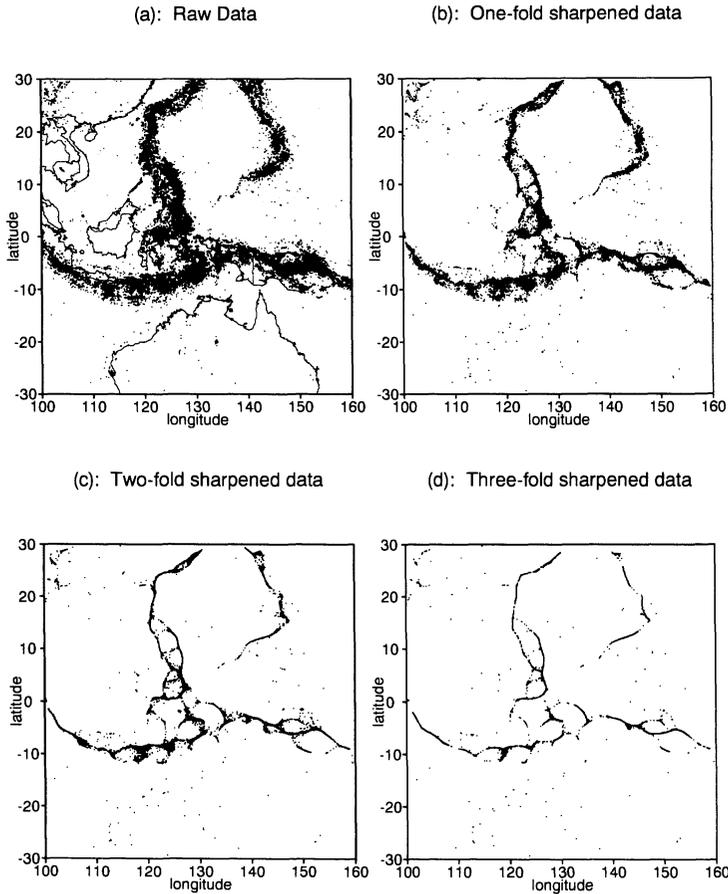


Figure 6: *Scatterplots of sharpened data with constant bandwidth.* Panel (a) shows the raw data, and panels (b)–(d) show the effects of applying the sharpening algorithm $m = 1, 2$ and 3 times, respectively, this time keeping bandwidth fixed at $h = h_0 = 2.0$ in all data sharpening steps.

An algorithm similar to that illustrated by Figure 6 was first suggested by Jones and Stewart (1997). It was designed to elucidate structure in geophysical data of the type depicted in panel (a), not as an aid to intensity estimation. It may be thought of as a device for forcing points to ‘walk’ up empirical approximations to lines of steepest ascent on the surface \hat{S} , rep-

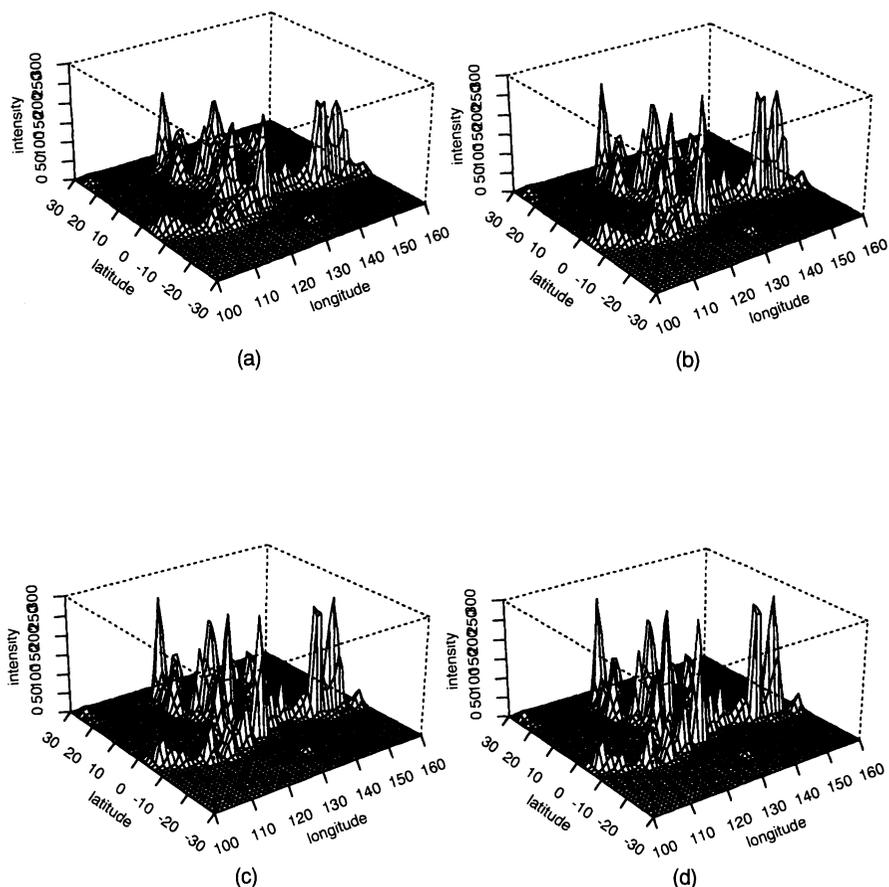


Figure 7: *Intensity estimates computed from sharpened data with constant bandwidth.* Estimates in panels (a)–(d) were computed from the scatterplots in panels (a)–(d), respectively, of Figure 6.

representing a plot of the intensity estimator against spatial location. Initially, in the first few steps, those lines are approximately perpendicular to ridge lines, but after a point has attained a reasonable height on \hat{S} its path starts to turn and, as the number of iterations increases, it moves in a direction that is increasingly parallel to a ridge line.

Increasing the number of iterations beyond this stage tends to reduce performance of the algorithm as a means for elucidating structure. In particular, detail about structure is lost at places where intensity is relatively low; this is already apparent in panel (d) of Figure 6. Jones and Stewart (1997) suggested a stopping rule to help overcome this problem. An alternative approach is to modify the algorithm by forcing the projection in the x -plane of the vector of motion up \hat{S} to be similar to that in early steps. Constraints such as this can substantially improve performance of Jones and Stewart's

algorithm.

4 Estimating the locations and strengths of poles

We have seen in Sections 2 and 3 that the intensity function associated with a point process of earthquake epicentres can rise very steeply from the plane, and give every appearance of having poles or pole lines in places of high intensity. Nonparametric methods may be employed to estimate both the location of a pole or pole line, and the rate at which intensity diverges in its vicinity. For the sake of simplicity we shall confine attention here to poles.

We begin with an idealised model for both the location and ‘strength’ of a pole. Estimators suggested by the model are appropriate very generally, and so the model amounts only to a device for pointing the way to methodology, not to a specific structural assumption. To this end, we assume that in the vicinity of a point v in the plane, the intensity $\nu(x)$ is asymptotic to a constant multiple of $\|x - v\|^{-\alpha}$, where $\alpha > 0$ represents pole strength. In order for the expected number of points in each bounded, nondegenerate region to be finite, we need $\alpha < 2$. Of course, if the point process were Poisson then the actual number of data in a region would be infinite, with probability 1, if the expected number there were infinite. The value of α is related to the correlation dimension, D , of the point process by the formula $D = 2(2 - \alpha)$. See Grassberger and Procaccia (1983).

It may be shown that, even if the data are from a Poisson process with this intensity, maximum likelihood estimation of v is not feasible. Nevertheless, given an estimator \hat{v} of v , a form of maximum likelihood estimation of α is possible. Estimators of v may be based on maximising the number of points within a small region, and have at least two forms, as follows. Let $\mathcal{D} = \mathcal{D}(w, r)$ denote the closed disc of radius r centred at w . Define \hat{v} to be either that value of w which minimises the area of $\mathcal{D}(w, r)$ subject to this disc containing at least a given number, N say, of points; or a value of w which maximises the number of points contained in $\mathcal{D}(w, r)$ for a given value of r . If the points X_i are distributed in the continuum then the former \hat{v} is uniquely defined with probability 1, while the latter is not unique, with the same probability. For this reason we favour the former estimator.

Given \hat{v} we may define an estimator $\hat{\alpha}$ of α to be the minimiser of

$$\ell(\alpha) = \alpha \sum' \log \|X_i - \hat{v}\| + M(\mathcal{D}_2 \setminus \mathcal{D}_1) \log \left(\int_{\mathcal{D}_2 \setminus \mathcal{D}_1} \|x - \hat{v}\|^{-\alpha} dx \right),$$

where $\mathcal{D}_1 \subseteq \mathcal{D}_2$ are concentric discs centred at \hat{v} , \sum' denotes summation over those points X_i that lie in $\mathcal{D}_2 \setminus \mathcal{D}_1$, and $M(\mathcal{D}_2 \setminus \mathcal{D}_1)$ equals the number of such points. The equation $(\partial/\partial\alpha)\ell(\alpha) = 0$ has a unique solution $\hat{\alpha}$. The radii of \mathcal{D} , \mathcal{D}_1 and \mathcal{D}_2 play the roles of smoothing parameters.

It may be proved that $\hat{\alpha}$ and \hat{v} are consistent for α and v , under conditions that are much more general than those asserted by the motivating model $\nu(x) = \nu_0(x) \equiv \|x - v\|^{-\alpha}$. A theory describing rates of convergence may be developed (Choi and Hall, 1997), having points of contact with that for semiparametric estimation of parameters in distributions with regularly varying tails; see e.g. Embrechts, Klüppelberg and Mikosch (1997, Chapter 6). It is straightforward to incorporate effects of noise into the intensity model, for example by assuming that an independent bivariate Gaussian vector with zero mean is added to each point in the plane. Provided the covariance matrices of these vectors are known they do not materially complicate estimation of α .

However, it may be shown that there is negligible information in the data for estimating noise covariance, under intensity models such as $\nu = \nu_0$. Fortunately, information about noise properties is often available from knowledge of the placement of recording stations around the epicentre of an event, and of the nature of the rocks through which shock waves passed on their way to those stations. Indeed, each measurement of longitude, latitude and depth is sometimes accompanied by its own error covariance matrix; see for example Jones and Stewart (1997).

We may estimate v and α for data from temporal clusters, and thereby compute estimators of the spatial trajectory, as a function of time, that are alternative to those considered in Section 2.3. Table 1 provides information about location and pole strength of nine of the ten shallow (i.e. no deeper than 36 km) Kanto event clusters that were the subject of Figure 3. (The cluster numbered 6 in Figure 3, occurring during February 1988 and having relatively low intensity, has been omitted from the present analysis.) Figure 8 depicts the corresponding trajectory, and should be compared with Figure 3. Smoothing parameters were chosen by performing a simulation study involving models that produced realisations approximating the data clusters. Details of the analysis are given by Choi and Hall (1999a). Methods for estimating α in related problems have been discussed by Theiler (1990), Smith (1992), Grassberger and Procaccia (1993), Mikosch and Wang (1995), Harte (1996) and Vere-Jones (1999).

Acknowledgements We are grateful to Dr. D. Harte and Professor D. Vere-Jones for providing the earthquake data analysed in this paper, and to Professor B.L.N. Kennett for helpful comments during the course of our work. The constructive comments of two referees have also been particularly helpful.

Year	Longitude	Latitude	N	\hat{v}_x	\hat{v}_y
1980	139.0,139.3	34.8,35.1	222	139.186	34.966
1983	139.0,139.3	34.8,35.1	238	139.200	34.938
1984	139.1,139.4	34.8,35.1	383	139.217	34.928
1986	139.0,139.3	34.8,35.1	207	139.175	34.948
1987	139.1,139.4	34.8,35.1	489	139.258	34.913
1988	139.05,139.35	34.8,35.1	237	139.195	34.951
1989	138.95,139.25	34.8,35.1	175	139.108	34.986
1993 ¹	139.0,139.3	34.8,35.1	337	139.177	34.937
1993 ²	139.0,139.3	34.8,35.1	614	139.130	34.976

¹: includes only events in January

²: includes only events in May and June

Table 1: Locations and pole strengths for 9 shallow Kanto event clusters.

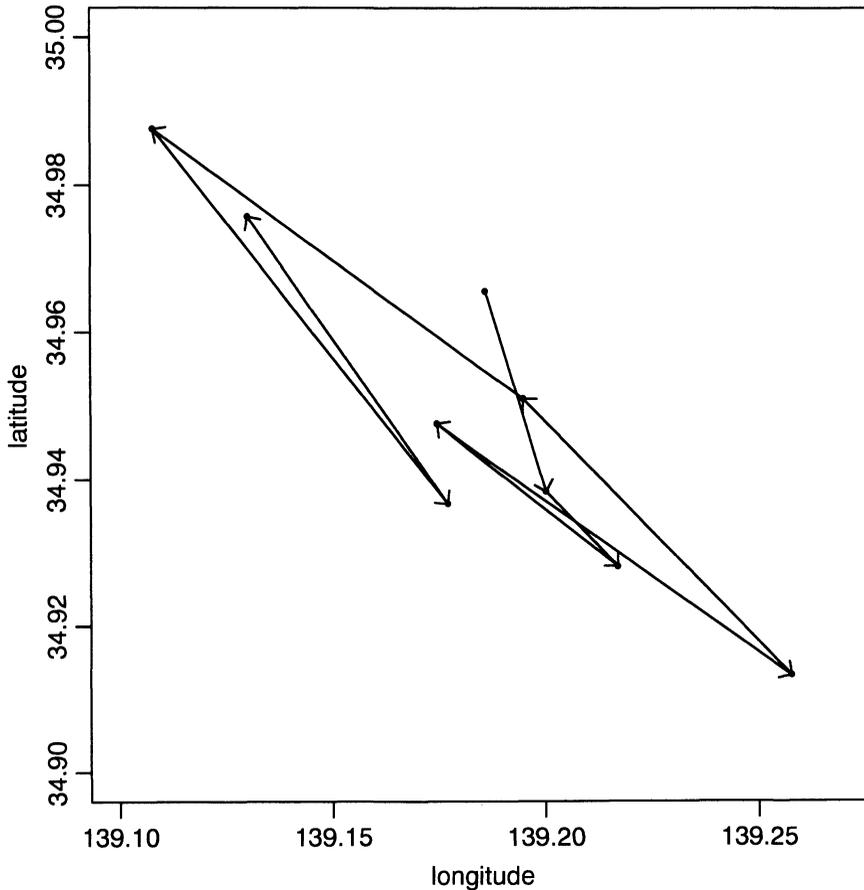


Figure 8. Spatial migration of pole. Data are clusters within the dataset used to compute Figure 3, and represent shallow Kanto events.

REFERENCES

- Choi, E. and Hall, P. (1998). A nonparametric approach to analysis of space-time data on earthquake occurrences. *J. Computat. Graph. Statist.*, under revision.
- Choi, E. and Hall, P. (1999a). On the estimation of poles in intensity functions. *Biometrika*, to appear.
- Choi, E. and Hall, P. (1999b). Data sharpening as a prelude to nonparametric density estimation. *Biometrika*, to appear.
- Davis, S.D. and Frohlich, C. (1991). Single-link cluster analysis of earthquake aftershocks — decay laws and regional variations. *J. Geophys. Res.* **96**, 6335–6350.
- Diggle, P.J. (1985). A kernel method for smoothing point process data. *Appl. Statist.* **34**, 138–147.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.
- Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208.
- Harte, D.S. (1996). *Multifractals— Theory and Applications*. PhD thesis, Victoria University of Wellington.
- Hawkes, A.G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B.* **33** 438–443.
- Jones, M.C. and Signorini, D.F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Assoc.* **92**, 1063–1073.
- Jones, R.H. and Stewart, R.C. (1997). A method for determining significant structures in a cloud of earthquakes. *J. Geophys. Res.* **102**, 8245–8254.
- Kagan, Y.Y. (1994). Observational evidence for earthquakes as a nonlinear dynamic process. *Physica D* **77**, 160–197.
- Koyama, M. (1993). Volcanism and tectonics of the Izu Peninsula, Japan (in Japanese). *Kagaku (Science)* **63**, 312–321.
- Mikosch, T. and Wang, Q. (1995). A Monte-Carlo method for estimating the correlation exponent. *J. Statist. Phys.* **78**, 799–813.
- NOAA and USGS (1996). *Seismicity Catalogs, Volume 2: Global and Regional, 2150 BC–1996 AD* (CD rom). National Oceanic and Atmospheric Administration (NOAA), National Geophysical Data Center, Boulder, Colorado; and U.S. Geological Survey (USGS), National Earthquake Information Center, Denver, Colorado.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9–27.
- Ogata, Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics* **169**, 159–174.

- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402.
- PDE (Preliminary Determination of Epicenters) Catalogue, (1997). Monthly Listings. U.S. Department of the Interior/National Earthquake Information Center, Denver, Colorado.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453–466.
- Smith, R.L. (1992). Estimating dimension in noisy chaotic time series. *J. Roy. Statist. Soc. Ser. B.* **54**, 329–351.
- Theiler, J. (1990) Statistical precision of dimension estimators. *Phys. Rev. A.* **41**, 3038–3051.
- Vere-Jones, D. (1999). On the fractal dimensions of point patterns. *Advances in Applied Probability* **31**, to appear.
- Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall, London.

CENTRE FOR MATHEMATICS AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200, AUSTRALIA
halpstat@fac.anu.edu.au