Ann. Inst. Statist. Math.
Vol. 44, No. 1, 1–11 (1992)

A SPACE-TIME CLUSTERING MODEL FOR HISTORICAL EARTHQUAKES

F. MUSMECI[1] AND D. VERE-JONES[2]

[1] *21 Via Carlo Alberta dalla Chiesa, 00061 Anguillara, Roma, Italy*
[2] *Institute of Statistics and Operations Research, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand*

(Received December 21, 1989; revised November 5, 1990)

Abstract. This paper describes a generalization of Hawkes' self-exciting process in which each event creates a process of "offspring" with conditional intensity governed by a diffusion kernel. The process may be described as a space-time branching process with immigration, the immigration representing a background series of independent events. The model can be fitted by likelihood methods. As an illustration it is fitted to the catalogue of historical Italian earthquakes.

Key words and phrases: Space-time models, cluster point processes, earthquake modelling, self-exciting process, spatial branching process, epidemic models.

1. Introduction

This paper describes an attempt to fit a simple type of space-time cluster process to historical data from Italy. As such it forms one of a series of papers (see Vere-Jones and Ozaki (1982), Ogata and Vere-Jones (1984), Vere-Jones (1988), Vere-Jones and Deng (1988)) which seek to develop models and associated statistical methods for analyzing historical earthquake catalogues.

Neither pure spatial methods nor pure time series methods have proved very revealing in studying patterns of earthquake occurrence. Even such a basic feature as the extreme clustering of smaller events (see, for example, Vere-Jones (1978)) is obscured by projection onto either time or space axes, a process which mixes the clusters and distorts their real characters. Typically, the historical records reveal a superficially random (Poisson) character, with a variety of subtler features—local clusters, possible coincidences, apparent variations or alternations in activity from one part of the region to another—which lie somewhere on the borderline between subjective fancy and statistically verifiable features. Generally speaking, historical consistency (completeness) requires that only the largest shocks (typically $M \geq 6$) be considered, in which case the data often comprises no more than 50–100 events. In such circumstances modelling is clearly more of an art than a science, and more cautious statisticians may well query the value of attempting to fit relatively complex models to such limited and unreliable data. However, the historical catalogues provide a unique, and in most cases the only, record of regional earthquake activity over periods of the order of 1000 years. Bearing this in mind it may be justified to spend a small proportion of the time until the next 1000 years of data have accumulated in speculating how the last 1000 years might be interpreted.

The pioneering papers in the development of space-time models for earthquakes are those of Kagan (1973) and Kagan and Knopoff (1976), which develop and apply a rather complex magnitude-space-time branching process, primarily for use with major earthquakes on a global scale. Their analysis suggests, among other things, a kind of wave of heightened activity spreading out from major events on a global scale. The model outlined in the present paper is an extension of the simpler "self-exciting" processes introduced by Hawkes (1971) and first applied to earthquake data by Hawkes and Adamopoulos (1973). Subsequently a series of papers by Japanese authors (e.g. Ogata *et al.* (1982), Ogata and Katsura (1986, 1988)) have extended the models in several directions, exploiting their simple likelihood structure as a basis for estimation and inference.

The special feature of the present extension is its explicit use of the space-time intensity. The importance of using space-time models, even to explain purely spatial patterns, was emphasized more than a quarter of a century ago by Whittle (1962) in a study of spatial patterns of soil fertility. The point is that the inclusion of the time variable makes possible a causal, evolutionary approach which is simply not available in the purely spatial context. From the point process viewpoint, this means that space-time models can be treated as "marked" point processes, taking time as the key variable and treating the spatial coordinates (latitude and longitude of the earthquake epicentre), and also the magnitude, as marks.

As such the model comes under the general theory for marked point processes developed by Jacod (1975), Karr (1986) and others, in which the point process is uniquely specified by its "compensator" (or time derivative, the "conditional intensity function"). A simple product formula is then available for the likelihood. An introduction to these concepts in the unmarked case is given in Daley and Vere-Jones ((1988), Chapter 13).

We shall not stress the general theoretical aspects in the present paper, but concentrate rather on describing and fitting the model (Sections 2–3 below) with an application to the Italian historical catalogue given in Section 4.

2. Description of the model

We assume the following "self-exciting" form for the space-time conditional intensity $\lambda(t, \boldsymbol{x}, M)$:

$$\lambda(t, \boldsymbol{x}, m)dt\,d\boldsymbol{x}\,dM = f(M)dM\left[\lambda_0(\boldsymbol{x}) + \sum_{i:t_i<t} h(t - t_i, \boldsymbol{x} - \boldsymbol{x}_i, M_i)\right]dt\,d\boldsymbol{x}. \tag{2.1}$$

Here

(i) The left-hand side (LHS) of (2.1) can be interpreted as the conditional probability of an event in the time-space-magnitude window $(t, t+dt) \times (\boldsymbol{x}, \boldsymbol{x} + d\boldsymbol{x}) \times (M, M+dM)$, given the past history of events $\{t_i, \boldsymbol{x}_i, M_i\}$ with $t_i < t$.

(ii) The earthquake magnitudes M_i are supposedly determined, independently of all other aspects of the process, according to a distribution function with density $f(M)$. (In practice $f(M)$ is exponential, corresponding to the so-called "Gutenberg Richter frequency-magnitude relation"; see, for example, Vere-Jones and Smith (1981).)

(iii) $\lambda_0(\boldsymbol{x})$ is the space-time intensity, assumed constant in time, of a Poisson process of independent background events.

(iv) $h(u, \boldsymbol{x}, M)$ is a kernel defining the contribution to the risk (conditional intensity), at time u and location $\boldsymbol{x}$, from an earthquake of magnitude M at the space-time origin. Note the homogeneity assumption that this kernel is a function of the differences $t - t_i$ and $\boldsymbol{x} - \boldsymbol{x}_i$ only, not of the pairs of coordinates (t, t_i) and $(\boldsymbol{x}, \boldsymbol{x}_i)$ separately. In an epidemic-type interpretation, $h(u, \boldsymbol{x}, M)$ describes the risk of a newly infected individual appearing at location $(u, \boldsymbol{x})$ after contact with an infected individual, with "strength of infection" M, at the origin.

The particular feature of the Hawkes' self-exciting model, which is preserved in the above extension, is the representation of the conditional intensity as a sum of contributions from all previous events. As with the original model, the process defined by (2.1) can also be interpreted as a "branching process with immigration" (see Hawkes and Oakes (1974)), the immigration component being described by the constant rate (in time) Poisson process of background events, while the "offspring" from a given "ancestor" with coordinates $(t_0, \boldsymbol{x}_0, M_0)$ form a Poisson process in space-time with intensity $h(t - t_0, \boldsymbol{x} - \boldsymbol{x}_0, M_0)$ $(t > t_0)$.

If either $\lambda_0(\boldsymbol{x})$ is bounded or has finite integral, then it is easy to deduce from general conditions given, e.g. in Daley and Vere-Jones ((1988), Subsections 8.2–8.3) that a sufficient condition for the existence of a stationary (in time) version of the process is

$$\rho \equiv E_M \left[\iint h(u, \boldsymbol{x}, M) du\, d\boldsymbol{x} \right] < 1 \tag{2.2}$$

which simply asserts that the mean number of "offspring" per "ancestor" is less than one, and implies that the expected total number of progeny produced by a single ancestor is finite and bounded in $\boldsymbol{x}$. If the inequality is reversed ($\rho \geq 1$), the process is "explosive", and will typically increase without bound as t increases.

Under the condition $\rho < 1$ the first moment or expectation measure of the stationary process has a density $m(\boldsymbol{x})$ in space which can be found by taking expectations in (2.1). This leads to the integral equation of renewal type

$$m(\boldsymbol{x}) = \lambda_0(\boldsymbol{x}) + \int m(\boldsymbol{x} - \boldsymbol{y}) q(\boldsymbol{y}) d\boldsymbol{y} \tag{2.3}$$

where

$$q(\boldsymbol{y}) = \int_0^\infty E_M[h(t, \boldsymbol{y}, M)] dt.$$

Since $\int q(\boldsymbol{y})d\boldsymbol{y} < 1$ from (2.2), equation (2.3) is readily seen to have the solution

$$m(\boldsymbol{x}) = \lambda_0(\boldsymbol{x}) + \int \lambda_0(\boldsymbol{x} - \boldsymbol{y})R(\boldsymbol{y})d\boldsymbol{y},$$

where the "renewal density" $R(\boldsymbol{y})$ is given by the convergent series

$$R(\boldsymbol{y}) = q(\boldsymbol{y}) + q^{*}q(\boldsymbol{y}) + \cdots,$$

and * denotes convolution in space. Thus the stationary rate (constant in time, varying over space) is, under these conditions, just a smoothed version of the immigration rate. In the non-stationary case ($\rho \geq 1$), the mean rate $m(\boldsymbol{x}, t)$ will generally increase without bound, either linearly ($\rho = 1$) or exponentially ($\rho > 1$), as $t \to \infty$, although more complex types of behaviour can also occur.

In the sequel we shall consider two specific parametric forms for the kernel h. The first, suggested in part by the discussion in Whittle (1962), is a standard diffusion kernel

$$h(u, \boldsymbol{x}, M) = Ae^{\alpha M}e^{-\beta u}\frac{1}{2\pi\sigma_x\sigma_y u}\exp\left\{-\frac{1}{2u}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right\}. \tag{2.4}$$

Here $\boldsymbol{x} = (x, y)^T$, A is an overall constant, $e^{\alpha M}$ describes the dependence of the risk on the magnitude of the exciting event, $e^{-\beta u}$ is an exponential damping factor (energy absorption), and the diffusion constants σ_x, σ_y control the rates of diffusion of risk along the x- and y-directions respectively. For fixed u, the contours of constant risk are ellipses with their axes aligned along the x- and y-axes. For fixed $\boldsymbol{x}$, on the other hand, the risk at time u after an event at the origin decays asymptotically at a rate proportional to $u^{-1}e^{-\beta u}$, which is reminiscent (apart from the exponential term) of the so-called "Omori Law" for the decay in aftershock frequency following a main event.

However, in the context of historical earthquakes the function of the kernel $h(\cdot)$ is not so much to describe the immediate aftershock sequences (which rarely contain more than one or two of the larger events to which the historical catalogue has to be restricted for this type of analysis) as the longer term transfer of stress from one locality to another. Although the direct evidence for a diffusion mechanism for such a process is hardly overwhelming, many historical catalogues, including the Italian catalogue, include examples of pairs or short sequences of large events, separated by distances of 100 km or greater, which occur within time periods which are short by comparison with the overall mean time between events. Similarly the diffusion mechanism may help to account for the impression of waves of heightened activity which move up and down the peninsular in an irregular fashion.

In addition to the Gaussian distributions incorporated in the kernel (2.4) we experimented with a product-Cauchy form

$$h(u, \boldsymbol{x}, M) = Ae^{\alpha M}e^{-\beta u}u^2c_xc_y[\pi^2(x^2 + u^2c_x^2)(y^2 + u^2c_y^2)]^{-1} \tag{2.5}$$

in which the contours for fixed u are closer to diamond shaped, and the decay for fixed $\boldsymbol{x}$ is asymptotically proportional to $u^{-2}e^{-\beta u}$ as $u \to \infty$.

3. Estimation of model parameters

Given the catalogue, the likelihood for a space-time process with conditional intensity of the form (2.1) can be represented in the form

$$(3.1) \qquad \log L = \sum_{i=1}^{N} \log \lambda(t_i, \boldsymbol{x}_i, M_i) - \int_W \lambda(t, \boldsymbol{x}, M) dt\, d\boldsymbol{x}\, dM$$

where W refers to the space-time-magnitude window of the observation zone.

In order to fit model (2.1) to the data, a preliminary estimate of the form of $\lambda_0(\boldsymbol{x})$ is first needed. This can be made, for example, by a simple kernel smoothing of the complete data set. Since the relative weight to be given to the background terms is not known a priori, an additional proportionality factor needs to be introduced into (2.1), by writing it in the form

$$(3.2) \qquad \lambda(t, \boldsymbol{x}, M) = f(M) A \left\{ (1-p)\hat{\lambda}_0(\boldsymbol{x}) + p \sum_{i: t_i \le t} h(t - t_i, \boldsymbol{x} - \boldsymbol{x}_i, M_i) \right\}$$

where $\hat{\lambda}_0(\boldsymbol{x})$ is the estimate of $\lambda_0(\boldsymbol{x})$ and p is constrained to the unit interval $0 \le p \le 1$.

Before entering a numerical optimization routine two further simplifications were made. Already the form (2.1) embodies the assumption that magnitudes are independent of other features of the process. When introduced into (3.1) the integral of $f(M)$ in the second term on the right side of (3.1) equals unity and so drops out of the equation, leaving the likelihood as the product of two terms, the first a product of the $f(M_i)$ and the second independent of the form of $f(M)$. Thus the problem of estimating any parameters in the distribution of magnitudes can be treated as a separate problem in its own right, and the essential task reduces to maximizing the second term, that is the so-called "partial likelihood"

$$(3.3) \qquad \log L_1 = \sum_{1}^{N} \log \lambda_1(t_i, \boldsymbol{x}_i) - \int_{W_1} \lambda_1(t, \boldsymbol{x}) dt\, d\boldsymbol{x}$$

where $\lambda_1(t, \boldsymbol{x})$ is the coefficient of $f(M)$ in (3.2), W_1 is the space-time component of the observation window, and the magnitudes appear in (3.3) only in the form of known (observed) values.

Also the overall constant A in (3.2) can be estimated directly from (3.3), by differentiating with respect to A and solving, which leads to the result

$$\hat{A} = N/I, \qquad \text{where} \quad I = \int_{W_1} \psi(t, \boldsymbol{x}) dt\, d\boldsymbol{x}$$

and we have written $\psi(t, \boldsymbol{x})$ for the term in braces in (3.2). Substituting for $\hat{A}$ back into (3.3) gives the reduced form

$$(3.4) \qquad \log L_1 = N \log N - N \log I - N + \sum^{N} \log \psi(t_i, \boldsymbol{x}_i).$$

A similar calculation can be performed for the special case of a constant-rate Poisson process, which leads to the likelihood

$$\log L_0 = N \log N - N \log |W| - N \tag{3.5}$$

where $|W|$ is the space-time "volume" of W. Hence we can replace the maximization of (3.3) by the maximization of the log likelihood ratio $\log(L_1/L_0)$, in the form

$$\log(L_1/L_0) = \sum_1^N \log \psi(t_i, \boldsymbol{x}_i) - N \log(I/|W|). \tag{3.6}$$

This manipulation reduces the number of unknown parameters by one and gives the solution directly in terms of a likelihood ratio which is independent of the choice of scales in the time and space dimensions, and is the quantity of immediate practical interest.

The estimation of the remaining parameters can then be completed by selecting one or other of the parametric forms (2.4), (2.5) for $h(\cdot)$ and applying a standard optimization routine, such as those available in the NAG or IMSL subroutine libraries.

The initial attempts to produce a satisfactory solution failed, apparently as a result of the mixture form of (2.1) and the existence in the catalogue of events with identical spatial coordinates $\boldsymbol{x}_i$, a combination which produced spurious maxima with $\sigma_x \to 0$, $\sigma_y \to 0$. This difficulty was overcome by constraining σ_x and σ_y to lie above small positive constants ε_x and ε_y, paying careful attention to overflow procedures when the arguments of the exponentials in (2.4) become too large, and using the double precision version of the optimization routine. With these additional precautions satisfactory convergence to an interior point of the parameter space was obtained.

4. Application to the Italian historical catalogue

As an illustration of the modelling and estimation procedures, we describe their application to major events in the Italian historical catalogue. The application was restricted to events with a maximum Mercalli intensity of 9 or greater, and to the time period 1000–2000 AD. A comb diagram for the data is shown in Fig. 1, and a plot of the epicentres in Fig. 2.

One problem with this particular catalogue is that the size of events is indicated only by the maximum Mercalli intensity and not by the magnitude. Although intensities are approximately linearly correlated with magnitudes, they are a measure of surface damage rather than of energy release, and their use as a surrogate for magnitudes in the analysis represents one step further away from a direct link with the geophysical reality. A revised catalogue with magnitudes is in preparation but was not available for the analysis.

Other problems, typical for historical catalogues, concern doubts about the completeness of the data set, even for these large events (an increase of frequency with time is visible even by eye in Fig. 1) and a tendency for events to be

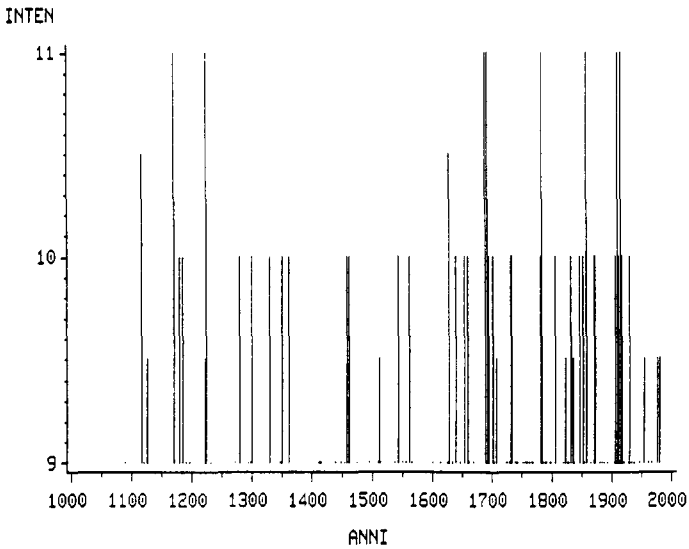

Fig. 1. Comb diagram for Italian historical data AD 1100–1980.

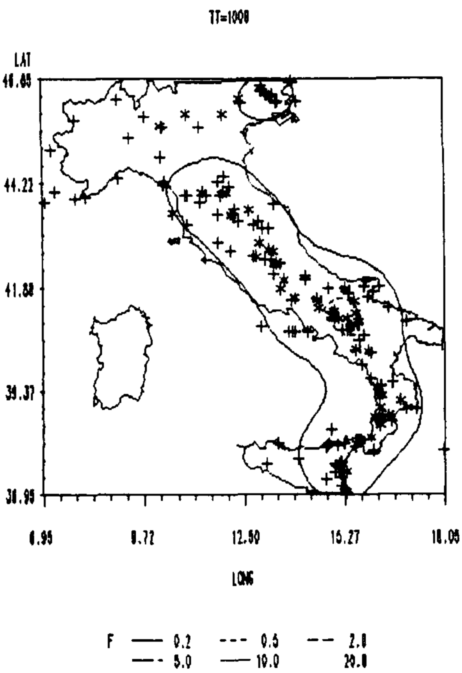

Fig. 2. Background intensity $\lambda_0(x)$ with superimposed events AD 1100–1980.

given identical epicentres when they affect the same localities. See, for example, Margottini and Serva (1988) for further illustrations of the problems of dealing with historical catalogues.

Despite these problems the model was applied directly to the catalogue data, as a test of the feasibility of the methodology and to gain a preliminary idea of the results that might be expected from such an analysis.

The results of the estimation procedure are shown in Table 1, which lists the parameter values for four evaluations, three with the Cauchy kernel using different versions of the background intensity $\lambda_0(\boldsymbol{x})$, and one with the diffusion kernel. The values are in artificial units, after rescaling all dimensions to the interval $(-1, 1)$. In real units, the values of β represents a half life for the exponential decay in the range 400–2000 years, and the values of c_x, c_y represent spreading velocities of the order of 5–50 km per century. The values of H in the first column represent the values in kilometres of the radii of a preliminary smoothing kernel applied to the total data set to estimate $\lambda_0(\boldsymbol{x})$.

Table 1. Parameter values for cluster models.

(a) Cauchy-type clusters

Smoothing parameter	p (proportion)	α (magnitude)	β (time decay)	c_x (lat)	c_y (long)	$\Delta \log L$
$H = 50$	.992	$-.00292$	1.1264	.00922	.00017	39.5
20	.987	$-.00064$	.4684	.02789	.01931	62.8
10	.980	$-.00029$	.3900	.04208	.03272	84.1

(b) Diffusion clusters

	p	α	β	σ_x	σ_y
$H = 20$	.9856	.0000	.4590	.0079	.0040

Once the parameters have been estimated, interpolated values of the risk at any time t and location $\boldsymbol{x}$ can be obtained by substitution in the intensity formula (2.1) with h given by (2.4) or (2.5) as appropriate. One way of illustrating the results is as a sequence of contour plots for the intensity at different times t. Ideally, the sequence of plots should be displayed in movie-style, and would then illustrate graphically the rise and fall of high risk regions in response to the development of various earthquake clusters. An indication of the results to be expected is given in the panel of results in Fig. 3, which shows the risk contours for 1620, 1660, 1700 and 1740. They illustrate, for example, the sudden appearance and slower delay of a high risk region near the Adriatic coast in 1660, and the relative stability of high risk regions near Friule in the North East, and in Sicily.

A few general comments concerning the interpretation of the parameter values may be made.

(i) In no cases was significant dependence on the magnitudes noted: for practical purposes one might as well take $\alpha = 0$ in either form (2.4) or (2.5) for $h(\cdot)$.

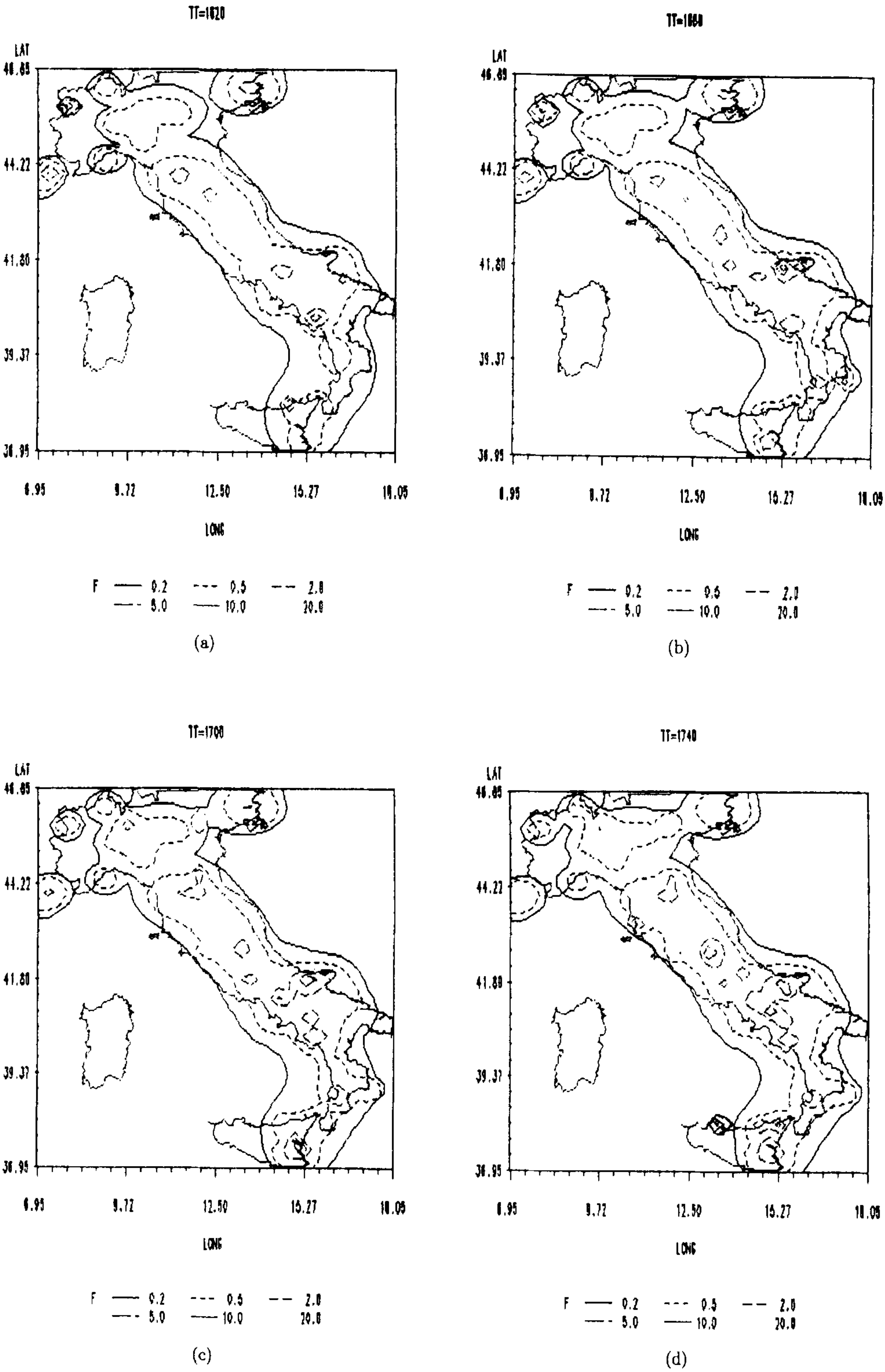

Fig. 3. Examples of evolution of space-time clusters: AD 1620–1740.

The result is counterintuitive and tends to suggest that the model is not tracking any physically meaningful feature.

(ii) The decay parameter β and the diffusion terms c_x, c_y all indicate a long-sustained, highly spatially concentrated risk. This suggests that the main modelling effect may just be to reproduce the near coincidence of historical epicentres from different epochs, a feature of the catalogue already referred to.

(iii) The estimated number of offspring per ancestor is given in terms of the parameters of (3.2) as

$$\hat{\rho} = \frac{Ap}{\beta} \frac{b}{b-a}$$

where b is the coefficient in the frequency magnitude (here intensity-magnitude) distribution

$$f(M) = be^{-bm}.$$

If the estimated coefficients are substituted into this formula, the values obtained are of the order of 10–25, and indicate a system in the early stages of supercritical growth. This estimate is very rough, and does not allow for the finiteness of the region, but is almost certainly related to the completeness problem with the catalogue, and the need to reproduce an increasing overall frequency of events with time. This increasing tendency is evident even in the relatively short time span covered by the diagrams in Fig. 3.

(iv) Given the parameters, a simulated catalogue can be developed following essentially the simulation procedures for point processes outlined by Ogata (1981) for the 1-dimensional case. The points in a realization are simulated in sequence; for each point, first the time coordinate, and then the space and magnitude values, are obtained, starting from the form of the risk function at the time of the preceding event, and recalculating the risk after the addition of the current point. The results of such realizations were not so tightly linked to the geographical boundaries of Italy as the actual data indicating that geographical constraints on the possible locations of events are not properly taken into account in the model. The simulation model can also be used to forecast the risk into the future.

(v) As the degree of smoothing in the estimate of $\lambda_0(\boldsymbol{x})$ is decreased, allowing tighter contours round the data, the overall fit, as well as the relative importance of the immigration term, increases. This is hardly surprising, but raises the question of how the one feature should be traded off against the other. One possibility here is to replace the non-parametric smoothing procedure by a parametric smoothing. Then the tightness of fit is controlled by the number of terms in the parametric smoothing model, and a model-selection criterion such as AIC can be used to judge the optimal stopping point.

Overall, the results of this preliminary analysis are rather disappointing, and not such that any great geophysical significance could be planned on them. Rather, they are dominated by features which seem to be artefacts of the data. A more careful and thorough study would be required, testing out the sensitivity of the different features to variations in the data set, checking the consistency of the estimation procedures from simulated data, etc., before one could be confident in the physical interpretation of the results.

Acknowledgements

This paper is based on work undertaken for a research contract between ENEA in Rome and the Institute of Statistics and Operations Research, VUW, Wellington, on the development of statistical methods for use with historical earthquake catalogues. The authors would like to acknowledge the support and encouragement of Dr. Mauro Basili, Head of the Laboratorio Ingegneria Dei Siti, ENEA, and to colleagues in both institutions for their help and advice.

REFERENCES

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*, Springer, New York.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, **58**, 83–90.

Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes—regional comparisons, *Bull. Internat. Statist. Inst.*, **45**, 454–461.

Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process, *J. Appl. Probab.*, **11**, 493–503.

Jacod, J. (1975). Multivariate point processes: predictable projections, Radon-Nikodym derivatives, representation of martingales, *Zeitschrift für Wahrscheinlichkeitstheorie und ihrer Grenzgebiete*, **31**, 235–253.

Kagan, Y. (1973). On a probabilistic description of the seismic regime, *Fizika Zemli*, **4**, 110–123.

Kagan, Y. and Knopoff, L. (1976). Statistical search for non-random features of the seismicity of strong earthquakes, *Physics of the Earth on Planetary Interiors*, **12**, 291–318.

Karr, A. F. (1986). *Point Processes and Their Statistical Inference*, Dekker, New York.

Margottini, C. and Serva, L. (eds.) (1988). *Workshop on Historical Seismicity of Central-Eastern Mediterranean Region*, ENEA and IAEA, Rome.

Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Trans. Inform. Theory*, **IT-27**, 23–31.

Ogata, Y. and Katsura, K. (1986). Point process models with linearly parameterized intensity for the application to earthquake catalogue, *J. Appl. Probab.*, **34**, 231–240.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial homogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, **40**, 29–39.

Ogata, Y. and Vere-Jones, D. (1984). Inference for earthquake models: a self-correcting model, *Stochastic Process. Appl.*, **17**, 337–347.

Ogata, Y., Akaike, H. and Katsura, K. (1982). The application of linear intensity models to the investigations of causal relations between a point process and another stochastic process, *Ann. Inst. Statist. Math.*, **34**, 373–387.

Vere-Jones, D. (1978). Space-time correlations for microearthquakes—a pilot study, *Supplement to Adv. in Appl. Probab.*, **10**, 73–87.

Vere-Jones, D. (1988). On the variance properties of stress release models, *Austral. J. Statist.*, **30**, 123–135.

Vere-Jones, D. and Deng, Y. L. (1988). Point process analysis of historical earthquakes from North China, *Earthquake Research in China*, **4**, 8–19.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data, *Ann. Inst. Statist. Math.*, **34**, 189–207.

Vere-Jones, D. and Smith, E. (1981). Statistics in seismology, *Comm. Statist. Theory Methods*, **A 10**, 1559–1585.

Whittle, P. (1962). Topographic correlation, power law covariance functions and diffusion, *Biometrika*, **49**, 305–314.