Earthquake modelling at the country level using aggregated spatio-temporal point processes

M.N.M. van Lieshout†

CWI, Amsterdam and Eindhoven University of Technology, Eindhoven, The Netherlands.

A. Stein

Faculty of Geo-Information Science and Earth Observation of the University of Twente (ITC), Enschede, The Netherlands.

Summary. The goal of this paper is to derive a risk map for earthquake occurrences in Pakistan from a catalogue that contains spatial coordinates of shallow earthquakes of magnitude 4.5 or larger aggregated over calendar years. We test relative temporal stationarity and use the inhomogeneous J–function to test for inter-point interactions. We then formulate a cluster model, and deconvolve in order to calculate the risk map. The model is validated using the leave-one-out principle.

In memory of Julian E. Besag

1. Introduction

Disasters like earthquakes apparently occur at erratic seismic locations and at unexpected moments. Processes generating earthquakes are prominent in earthquake prone areas that are at least partly determined by geological faults and occur in particular close to subduction zones. An earthquake describes both a sudden slip on a fault, and the resulting ground shaking and radiated seismic energy caused by the slip, by volcanic or magmatic activity, or other sudden stress changes in the earth. The release of energy at an unanticipated moment may then appear at the earth surface and is registered as the main shock. Main shocks are usually followed by aftershocks that are smaller than the main shock and within 1-2 rupture lengths distance from the main shock. Aftershocks can continue over a period of weeks, months, or years. In general, the larger the main shock, the larger and more numerous the aftershocks, and the longer they will continue. Other types of clusters known as swarms also occur. Such clusters are more diffuse and can be distinguished from the aftershock sequences by their showing no clear correlation with a main shock, nor the typical decay in frequency and magnitude common to aftershock patterns.

Earthquakes are well-known to cause vast destruction and panic among the affected population. Having a better knowledge on where the earthquakes, e.g. as major events or as aftershocks occur in relation to geological features, may thus result in identification of hazard zones. Modelling earthquake data has since long been a focus of research by seismologists and statisticians. Stochastic geometry offers various tools and procedures to contribute to a better understanding by means of spatial testing, spatial modelling and

†*Address for correspondence:* CWI, PO. Box 94079, 1090 GB Amsterdam, The Netherlands
E-mail: Marie-Colette.van.Lieshout@cwi.nl

mapping. In that sense, data collected routinely in public databases may reveal patterns that are otherwise unknown.

The literature on earthquakes in space and time has a long history. Important contributions were made by Ogata and his co-authors in an outstanding series of papers (e.g. (18; 19; 25)). Other contributions in the point process literature include ((23; 24)). Many, more applied, scientific papers have been published in the disciplinary literature (e.g. (12; 1)).

These papers consider space-time data, often for the island states of Japan and New Zealand. In this context, a temporal point process marked by location of occurrence is appropriate and conditional intensity functions given past occurrences can be written down explicitly. These in turn suffice to write down a likelihood function on which inference can be based. Moreover, edge effects are no issue.

The aim of this study is to explore spatial statistical techniques for data aggregated over time for which an explicit likelihood function is not available. Attention focuses on Pakistan, for which country annual patterns of earthquakes have been recorded for more than thirty five years. During this period, two major earthquakes of magnitude larger than seven were recorded: one occurring in 1997 and the major Kashmir earthquake of 2005. Moreover, seismic activity in neighbouring countries may well influence occurrences inside Pakistan.

The plan of this paper is as follows. The first sections are exploratory in nature. In Section 2 we discuss the data, in Section 3 we give a kernel estimator for the intensity of earthquake occurrence. The next two sections test for relative temporal stationarity (Section 4) and inter-point interactions (Section 5). We look at the patterns of aftershocks in the major earthquake years 1997 and 2005 in Section 6 and formulate a cluster model in Section 7 from which a risk map is derived. The model is validated using the leave-one-out principle in Section 8. The paper closes with a critical discussion.

2. Background and data

Pakistan is a country that is regularly affected by earthquakes. The reason for the vulnerability of the country to earthquakes is the subduction of the Indo-Australian continental plate under the Eurasian plate with its two associated convergence zones. One such zone crosses the country from approximately its South-West border with the Arabic Sea to Kashmir in the North-East. The other convergence zone crosses the Northern part of the country in the East-West direction and is the direct cause of the Himalayan orogeny. Pakistan-administered Kashmir lies in the area where the two zones meet. The geological activity born out of the collisions is the cause of unstable seismicity in the region.

Earthquakes can be severe with a devastating effect on human life and property. Two major earthquakes were recorded in 1997 and 2005 with magnitudes of 7.3 and 7.6 respectively. The 1997 earthquake occurred along the convergence zone running from the South-West to the North-East and resulted in about seventy casualties. The 2005 Kashmir earthquake, however, was devastating with at least $86,000$ casualties. The Pakistan Meteorological Department estimated a 5.2 magnitude on the Richter scale, whereas the United States Geological Survey (USGS) measured its magnitude as at least 7.6 on the moment magnitude scale with its epicentre about 19 km north-east of the city of Muzaffarabad. The earthquake is classified as 'major' by the USGS. The hypocenter was located at a depth of 26 km below the surface. Such big earthquakes are accompanied by many aftershocks.

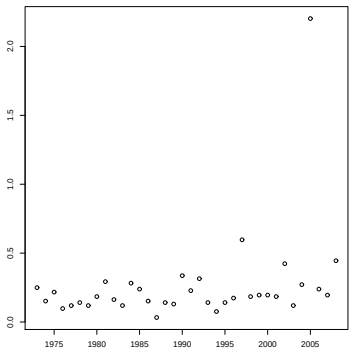

Fig. 1. The annual number of shallow earthquakes of magnitude 4.5 or higher per square degree latitude-longitude recorded in Pakistan during the years 1973–2008.

Aftershocks can even be stronger than the main earthquake itself. The city of Karachi (more than 1,000 km away) experienced a minor aftershock. There were many secondary earthquakes in the region, mainly to the northwest of the original epicentre. A total of 147 aftershocks were registered in the first day after the initial quake. On October 19, a series of strong aftershocks occurred about 65 km north-northwest of Muzaffarabad.

In addition to such major shocks, that are still relatively rare, many smaller shocks have been recorded (see `http://earthquake.usgs.gov/earthquakes` for a list of earthquakes since 1973). The majority of tectonic earthquakes originate at depths not exceeding tens of kilometres. Those occurring at a depth of less than 70 km are classified as 'shallow'. Earthquakes that originate below this upper crust are classified as 'intermediate' or 'deep'. See (16) for further details. Clearly, the impact of an earthquake depends on its epicentre, its depth as well as its magnitude. Minor earthquakes occur very frequently and may not even be noticed or recorded. Therefore we focus on those having a magnitude of at least 4.5 for which records are believed to be exhaustive (Van der Meijde, *pers. comm.*).

To summarize, our data (available from the authors on request) consist of the annual patterns of shallow earthquakes of magnitude 4.5 or higher in Pakistan during the period 1973–2008. The country level is appropriate, as most political and risk management actions are taken at this level. However, to avoid edge effects, we sometimes refer to data on earthquakes across the border. For each event, its location and magnitude is recorded. For the major earthquake years 1997 and 2005, also the times at which shocks occurred in the month following the main one are available.

The annual number of such earthquakes per square degree latitude-longitude is given in Figure 1 over the period 1973–2008. Note that the clearly visible outlier corresponds to the Kashmir earthquake in 2005 that generated a large number of aftershocks. The number of aftershocks in 1997 was considerably less and more diffuse. A histogram of the observed magnitudes is presented in Figure 2. In accordance with the Gutenberg–Richter power law,

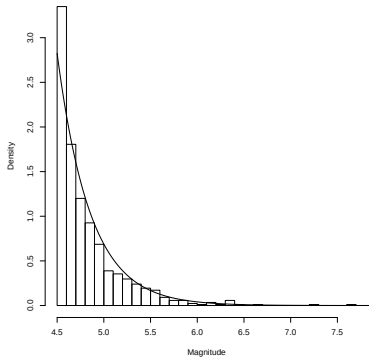

Fig. 2. Scaled histogram of the magnitudes restricted to $[4.5, \infty)$ of shallow Pakistan earthquakes occurring in the period 1973–2008 and the fitted shifted exponential density (black line).

we fit a shifted exponential probability density $\beta e^{-\beta(m-4.5)}$ for $m \geq 4.5$ and 0 elsewhere. The maximum likelihood estimator is $\hat{\beta} = 1/\overline{m}_{875} = 2.82$, where $\overline{m}_{875}$ is the sample mean over the set of 875 pooled magnitudes. Figure 2 indicates an adequate fit.

3. Spatial intensity

We first consider the spatial intensity function $\rho(x, y)$, where $(x, y) \in W$, the subset of $\mathbb{R}^2$ representing the Pakistan territory. To avoid edge effects, earthquake locations in Pakistan and in neighbouring countries within a distance of about one degree from the Pakistan border are aggregated into a single pattern as illustrated in Figure 3.

We then exclude the major earthquake years 1997 and 2005, pool the remaining thirty four years together and calculate the kernel estimator of intensity (8) using an isotropic Gaussian kernel with standard deviation 0.5. This corresponds to approximately 50 km, a value that is well in line with the spatial extent of zones affected by a major earthquake. The result is given in Figure 4.

High intensity parts occur in the north of the country, in the region bordering Afghanistan and Tajikistan, near the junction of plate boundaries, and in a smaller region in the east. A second zone of high earthquake activity lies in the mid-west of the country. In fact, the epicentre of the 1997 earthquake is located in this area.

4. Annual relative earthquake rates

The first analytical stage is an investigation into the spatial intensity function of earthquake events. Recall that visual and geological evidence suggest enhanced earthquake intensity in the northern and mid-western parts of the country. In this section, we test whether this

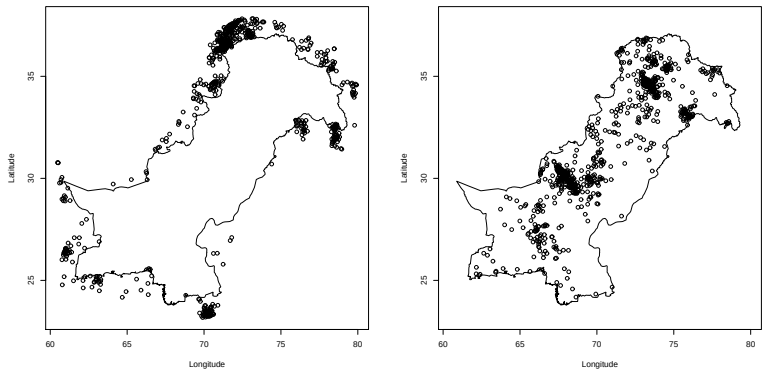

Fig. 3. Shallow earthquakes of magnitude 4.5 or higher in (left) a zone up to about one degree removed from the Pakistan border and (right) within the Pakistan territory during the years 1973–2008.

pattern persists over the years.

Write $W \subset \mathbb{R}^2$ for the compact set representing the Pakistan territory and let X_i be the point process of locations of shallow earthquakes of magnitude at least 4.5 that occur in Pakistan in year $i = 1973, \ldots, 2008$. Denote the intensity function of X_i by μ_i. In other words, for every Borel subset A of W,

$$\int_A \mu_i(x, y)\, dx\, dy$$

is the expected number of points of X_i falling in A.

In order to test whether μ_i is constant over the years except for a time-dependent scalar multiplication factor, we divide W in two disjoint areas A_N and A_S of equal pooled integrated intensity (cf. Figure 4). More precisely, A_N is the subset of W lying north of the 31.4° latitude line, and $A_S = W \setminus A_N$. Introduce the random variables

$$Y_i = \frac{N(X_i \cap A_N)}{N(X_i \cap W)},$$

where $N(X_i \cap A)$ is the cardinality of $X_i \cap A$ for every Borel set $A \subset W$ and $i \in \{1973, \ldots, 2008\}$. Thus, Y_i is the fraction of shocks above 31.4° latitude in year i.

To test the null hypothesis H_0 that $(Y_i)_i$ is stationary, we applied a test developed by Kwiatkowski *et al.* (14) known by the acronym KPSS referring to the first characters of the author's surnames. Given a time series of length n (in our study $n = 36$), define the partial sums process $S_n(i) = \sum_{j=1}^{i}(Y_j - \bar{Y}_n)$ for $i \in \{1, \ldots, n\}$, where $\bar{Y}_n$ is the sample mean $\sum_{i=1}^{n} Y_i/n$. Under the null hypothesis, the limit $\tau^2 = \lim_{n\to\infty} n \operatorname{Var} \bar{Y}_n = \sum_{j=1}^{\infty} c_j$ is well-defined provided the autocovariances $c_j = \operatorname{Cov}(Y_1, Y_{1+j})$ at lag j exist and their sum

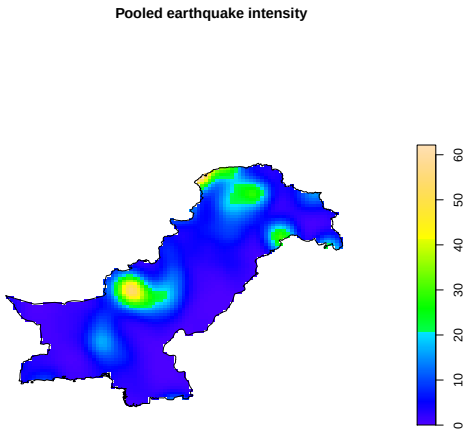

Fig. 4. Kernel estimator of intensity estimated from shallow earthquakes of magnitude at least 4.5 during the period from 1973 until 2008 but excluding the years 1997 and 2005.

is absolutely convergent. Under mild regularity conditions, $\tau^{-2}n^{-2}\sum_{i=1}^{n} S_n(i)^2$ converges in distribution to the integral of the squared Brownian bridge $\int_0^1 V(t)^2 dt$ (see (20) and (11, Cor. 1)). As τ^2 is unknown, we set

$$s_n^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{l}\left(1-\frac{j}{n}\right)\left(1-\frac{j}{l+1}\right)\hat{\gamma}_j,$$

where $\hat{\gamma}_j = (n-j)^{-1}\sum_{i=1}^{n-j}(Y_i-\bar{Y}_n)(Y_{i+j}-\bar{Y}_n)$ are the sample autocorrelations and l defines the maximal temporal lag taken into consideration. The weights $1-j/(l+1)$ were shown by Newey and West (17) to lead to a non-negative estimator. The authors also proved weak consistency, whereas strong consistency was proved by De Jong (13). In summary, the KPSS test statistic is given by

$$T_n = \frac{1}{n^2 s_n^2}\sum_{i=1}^{n} S_n(i)^2$$

which is asymptotically distributed as the integral of the squared Brownian bridge. Critical values of the test are reported in (14, Table 1).

For the data discussed in Section 2 with $l = 4$, the test statistic takes the value 0.1297 with a p-value exceeding 10%, so the null hypothesis cannot be rejected. We conclude that there is no statistical evidence of a temporal trend in intensity patterns of shallow earthquakes based on the north-south divide during the 36 years of study. Clearly, we could have divided W into more than two subsets, but in some years the number of events is so low (cf. Figure 1) that we did not choose to do so.

5. The $J-$function for inhomogeneous point processes

The next step is to test whether a series of inhomogeneous Poisson point processes could explain the annual patterns of earthquakes. For stationary point processes, a wide range of summary statistics exists that lend themselves to Monte Carlo testing ((4; 7)). The best known examples are the empty space function, the nearest neighbour distance distribution function, the J-function and the K-function. For an up to date account of these and other summary statistics, the reader is referred to (10).

For inhomogeneous point processes, Baddeley *et al.* (3) modified the K-function, which was extended to space-time point processes by Gabriel and Diggle (9). Recently, one of the authors proposed adaptations of the empty space function, the nearest neighbour distance distribution function and the J-function for intensity reweighted moment stationary point processes and discussed how to deal with marked and space-time point processes. As in the stationary case, the K-function can be seen as a second order approximation of the J-function. Note that weighting by the intensity compensates for the inhomogeneity.

Below, we shall use the inhomogeneous J-function. The basic idea is to compare the intensity reweighted point pattern around a typical point in the map to that around an arbitrarily chosen origin in space in order to gain insight in the interaction structure of the point process that generated the data. More formally, let X be a simple point process on $\mathbb{R}^2$ whose intensity function ρ exists and is bounded away from zero with $\inf_{x,y} \rho(x, y) = \rho_0 > 0$, write $B(0, t)$ for the closed ball of radius t centred at the origin, and let G and $G^{!0}$ denote the generating functionals of X and its reduced Palm distribution, respectively. Write, for $t \geq 0$, $u_t(x, y) = \rho_0 \, 1\{(x, y) \in B(0, t)\}/\rho(x, y)$ and define

$$J_{\mathrm{inhom}}(t) = \frac{G^{!0}\,(1 - u_t)}{G\,(1 - u_t)} \tag{1}$$

for all $t \geq 0$ for which the denominator is non-zero. See (15) for further details. The numerator can be interpreted as the inhomogeneous counterpart of 1 minus the nearest neighbour distance distribution function; the denominator is an inhomogeneous analogue of 1 minus the empty space function. Under the Poisson null hypothesis, numerator and denominator are identical, so that $J_{\mathrm{inhom}} \equiv 1$.

Below, we plot estimators of (1) together with their 5% Monte Carlo envelopes ((4; 7)) based on 19 independent samples from an inhomogeneous Poisson process with the same intensity function. Of course, the 'true' intensity function is unknown and has to be estimated. Based on the results of Section 4, we assume that the intensity function of X_i, the pattern in year i, is given by

$$\mu_i(x, y) = w_i \, \mu(x, y),$$

where μ is supposed to be normalized, that is, $\int_W \mu(x)\, dx = 1$. Under this assumption, the weights w_i can be estimated by $N(X_i \cap W)$, the number of shocks in year i. Furthermore,

for a set I of year indices, the intensity function μ_I of $X_I = \cup_{i\in I} X_i$ is $\sum_{i\in I} w_i\,\mu(x,y)$. Note that Figure 4 depicts an estimator $\hat{\mu}_I$ for $I = \{1973,\ldots,2008\} \setminus \{1997, 2005\}$. Plugging $\hat{\mu}_I$ and $\hat{\mu_0} = \min_{(x,y)\in W} \widehat{\mu_I(x,y)}$ into the estimators

$$\widehat{G(1-u_t^0)} = \frac{\sum_{l_k\in L\cap W_{\ominus t}} \prod_{(x,y)\in X_I\cap B(l_k,t)} \left[1-\frac{\rho_0}{\rho(x,y)}\right]}{\#L\cap W_{\ominus t}},$$

where $L \subseteq W$ is a finite grid, $B(a,t)$ the closed ball of radius t centred at a, and $W_{\ominus t}$ the eroded set $\{(x,y)\in W : d((x,y),\partial W)\geq t\}$, and

$$\widehat{G^{!0}(1-u_t)} = \frac{\sum_{(x_k,y_k)\in X_I\cap W_{\ominus t}} \prod_{(x,y)\in X_I\setminus\{(x_k,y_k)\}\cap B((x_k,y_k),t)} \left[1-\frac{\rho_0}{\rho(x,y)}\right]}{\#X_I\cap W_{\ominus t}},$$

we obtain a ratio estimator $\widehat{J_{\mathrm{inhom}}(t)}$ for the point pattern obtained by pooling all locations of earthquakes except those happening in 1997 and 2005. The result is shown in Figure 5. We conclude that X_I is more clustered than a Poisson process with the same intensity function up to $t \approx 1$.

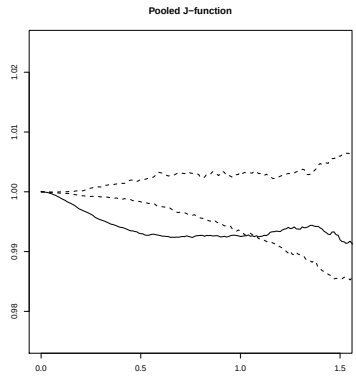

Fig. 5. Estimated inhomogeneous J-function for the pooled locations of shallow earthquakes of magnitude at least 4.5 during the period from 1973 until 2008 but excluding the years 1997 and 2005 with 5% upper and lower envelopes based on 19 independent realisations of an inhomogeneous Poisson process.

For the pooled data considered in Figure 5 we were forced to use the same data to estimate the intensity and J_{inhom} functions. For individual years, this is not needed. As an illustration, we considered the last three years. To estimate the intensity function, we excluded the year of interest as well as 1997 and 2005, then calculated the kernel estimator using an isotropic Gaussian kernel with standard deviation 0.5 taking into account also earthquakes happening within a distance of about 1 degree away from the Pakistan border

and normalized the result so as to place unit mass on W. The intensity function for year i is obtained by scaling by the number of earthquakes in year i. Therefore, the intensity function estimator is mass preserving. The resulting $\widehat{J_{\mathrm{inhom}}(t)}$ are shown in Figure 6. In 2006, the empirical J_{inhom}-function lies below the 5% envelopes, indicating clustering above that due to the inhomogeneity; the pattern in 2008 also exhibits strong attraction, especially for small t. In 2007, there is significant but milder clustering at intermediate range.

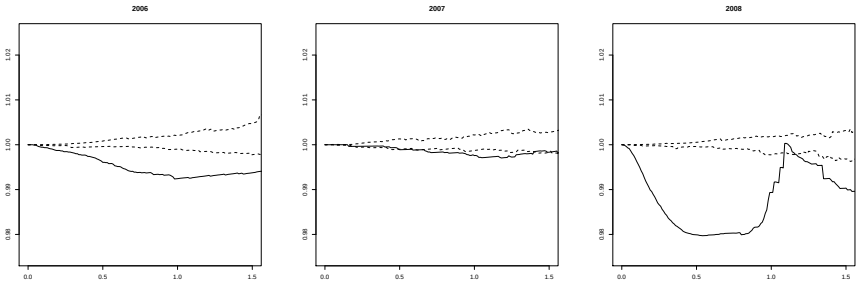

Fig. 6. Estimated inhomogeneous J-function for the locations of shallow earthquakes of magnitude at least 4.5 in 2006 (leftmost frame), 2007 (middle frame) and 2008 (rightmost frame) with 5% upper and lower envelopes based on 19 independent realisations of an inhomogeneous Poisson process.

For the years in which an earthquake of magnitude of at least 7 occurred, that is, for 1997 and 2005, the estimated inhomogeneous J-function strongly suggests clustering over and above that caused by the inhomogeneity, cf. Figure 7.

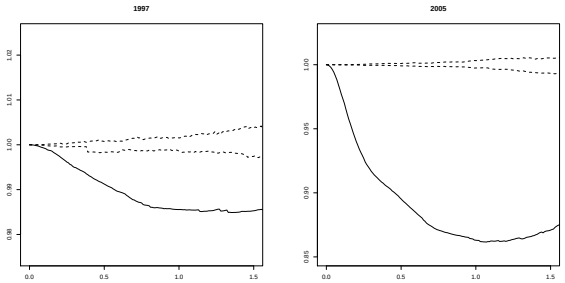

Fig. 7. Estimated inhomogeneous J-function for the locations of shallow earthquakes of magnitude at least 4.5 in 1997 (leftmost frame) and 2005 (rightmost frame) with 5% upper and lower envelopes based on 19 independent realisations of an inhomogeneous Poisson process.

6. Aftershocks in the major earthquake years

In this section, we focus on the two years (1997 and 2005) in which earthquakes of magnitude larger than 7 occurred and for which we have access to more detailed data. The scientific literature distinguishes a general earthquake intensity pattern from a pattern of aftershocks. These usually form a concentrated cluster during a relatively short time interval after a major earthquake event.

The 2005 Kashmir earthquake was studied by Anwar *et al.* (1). They found that the seismicity in Northern Pakistan decreased sharply following the shock on October 8. In particular, the number of earthquakes (of magnitude at least 4) that occurred more than a month after the first shock was found to be negligible compared to the number in the first month. We therefore concentrate on the 176 events (out of a total of 203 earthquakes during the whole of 2005) in our data that happen between October 8, 2005, and November 7, 2005. A histogram of the time (counted in days from October 8th) of these earthquakes is given in Figure 8. The figure confirms the picture painted above: a large number of aftershocks occurred during the first couple of days following the main shock, with a steep decline in numbers afterwards.

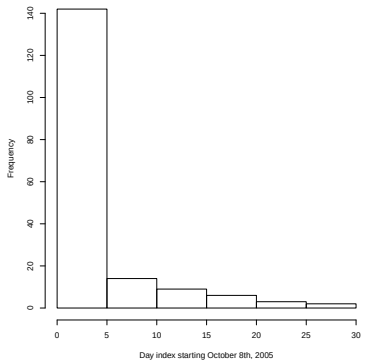

Fig. 8. Histogram of the times, counted in days starting October 8, 2005, at which shallow earthquakes of magnitude at least 4.5 occur.

It is interesting to note that all events are close to the epicentre of the main shock, cf. Figure 9. The pattern can be well described by two clusters: one corresponding to the main earthquake, the other to the next largest earthquake with a magnitude equal to 6.4 occurring approximately seven hours later. Pretending that the displacements in the two clusters follow a two-dimensional Gaussian distribution with mean zero and equal covariance matrices $\sigma^2 I_2$ for some $\sigma^2 > 0$, we may apply Fisher's linear discriminant function to assign the aftershocks to either of the clusters. Here I_2 denotes the two by two identity matrix. The result is given in Figure 9.

From the picture it is clear that the variances of the displacements are not identical, but

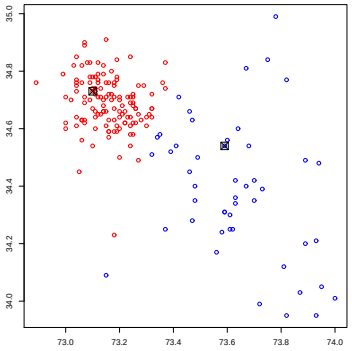

Fig. 9. Locations of shallow earthquakes of magnitude at least 4.5 occurring in the period October 8–November 7, 2005. The locations of the two shocks with the largest magnitude are identified by a crossed box; their clusters are colour coded.

nevertheless a visually good separation is achieved. The estimated within group variances are given in the table below.

Anwar *et al.* (1) did not consider the year 1997. In contrast to the pattern of aftershocks in 2005, in 1997 there is no clear decrease in aftershock occurrence following the main shock and the pattern is more diffuse, a type of behaviour typical of a swarm. We therefore extracted the cluster by hand and estimated the variance of the deviations from the epicentre in longitude and latitude (see the table below). Instead of looking at displacements with respect to the epicentre, we could have computed the sample variance, which would have led to slightly smaller values but would be difficult to integrate with the values obtained for 2005. The pooled sample variance is 0.038 (standard deviation 0.19). It may be conjectured that the spread of aftershocks is related to the magnitude of the earthquakes (cf. the difference in aftershock patterns between the two largest shocks in 2005) but we do not have enough data to support this conjecture.

	$\hat{\sigma}_x^2$	$\hat{\sigma}_y^2$
1997	0.0817	0.0969
2005	0.0394	0.0813
2005	0.0117	0.0105

7. Model

In the analysis of earthquakes, spatio-temporal Hawkes processes, in particular Ogata's Epidemic Type Aftershock-Sequences (*ETAS*) model, have become the standard first approximation for seismic catalogue data that come in the form of a list of earthquake locations,

their times and magnitudes. An excellent review is given by Ogata (19) who also gives historical references. In such a Hawkes process, 'immigrants' arrive according to a temporal Poisson process and are marked by their spatial location and other attributes as required. Each immigrant generates a finite marked Poisson process of 'offspring' with an intensity function that depends on the parent, independently of other immigrants. The offspring in their turn also generate offspring independently of all others, and so on. In other words, each immigrant produces a branching process of descendants. Therefore, a Hawkes process can also be described as a marked Poisson cluster process.

In the earthquake context, the offspring are swarms or aftershock clusters; their number typically depends on the magnitude of the parent, their temporal displacement follows the Omori (Pareto) power law. With regard to the marks, the magnitudes are widely assumed to follow a shifted exponential distribution; for the spatial displacements, various probability distributions have been tried. Examples, including the spherical Gaussian distribution we used in Section 6, are given in (19; 12; 6).

The advantage of focussing on the temporal dimension and treating other variables of interest (magnitude and spatial location) as marks is that a conditional intensity can be written down. Consequently, a likelihood function is available in closed form and can be used for inference. For details see (19).

The Pakistan data at our disposal, however, are only available as aggregated point patterns over calendar years marked by magnitude (with the exception of some information on aftershocks following the 1997 and 2005 catastrophes). We therefore model these data as a multivariate marked point process. Taking into account that aftershocks occur over relatively short periods of time only and the low point counts in most years, we set

$$Z = (Z_1, \ldots, Z_{36})$$

where the Z_i are independent but not identically distributed marked point processes with locations in W and marks in $[4.5, \infty)$ that represent the magnitudes. For the marks we assume random labelling according to a shifted exponential distribution $f_M(m) = \beta \exp[-\beta(m - 4.5)]$, $m \geq 4.5$, cf. Section 2.

We model the marked point processes Z_i, $i = 1, \ldots, 36$, as Poisson cluster processes: each 'parent' generates a Poisson number of offspring with a mean number $A(m)$ that depends on the magnitude m of the parent. The offspring are independent and normally distributed with probability density $f_N(\cdot - (x, y))$ centred at parent (x, y) and having covariance matrix $\sigma^2 I_2$ where I_2 is the 2×2 identity matrix (cf. Section 6). Note that a parent in the Pakistan territory W may generate offspring across the border, and that some earthquakes recorded in Pakistan may arise from a parent in a different country. We therefore assume that the parent locations form a point process on the set $W_b \supseteq W$ consisting of W and a buffer zone large enough to make the probability of a parent in $\mathbb{R}^2 \setminus W_b$ generating offspring in W negligible. We assume that this point process of parents is Poisson with locally finite intensity measure $\alpha_i \lambda(x, y)$, $i = 1, \ldots, 36$, $(x, y) \in W_b$. For identifiability reasons, we normalize the process so that λ is a probability density on W, i.e. $\int_W \lambda(x, y)\, dx\, dy = 1$.

We base inference on the first order moment measure. By (5, Prop. 6.3.III), the intensity function of X_i can be written as

$$\mu_i(x, y) = \alpha_i \int_{4.5}^{\infty} A(m)\, f_M(m)\, dm \int_{W_b} f_N((x, y) - (u, v))\, \lambda(u, v))\, du\, dv. \tag{2}$$

The joint intensity function of locations and marks in year i at $((x, y), k)$ is simply $\mu_i(x, y)$ $f_M(k)$. Equation (2) should be seen in the light of Section 5: for $I = \{1973, \ldots, 2008\} \setminus$

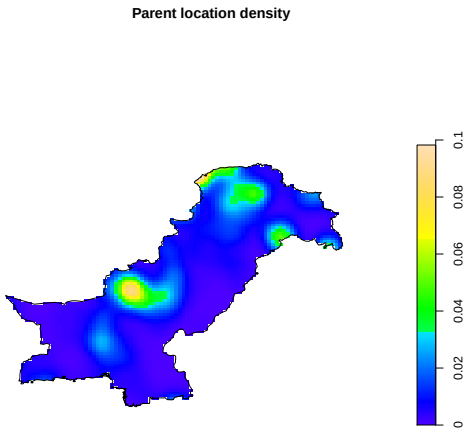

Fig. 10. Earthquake risk map estimated from shallow earthquakes of magnitude at least 4.5 during the period from 1973 until 2008 but excluding the years 1997 and 2005.

$\{1997, 2005\}$,

$$\mu_I(x,y) = \left(\sum_{i\in I} \alpha_i\right) \int_{4.5}^{\infty} A(m)\, f_M(m)\, dm \int_{W_b} f_N((x,y)-(u,v))\, \lambda(u,v))\, du\, dv.$$

Therefore μ_I is proportional to a convolution of the normal density f_N with λ. We refrain from formulating a model for $A(m)$ due to the paucity of data on cluster sizes resulting from a parent event of magnitude m.

It remains to estimate λ. To do so, we work in the Fourier domain: the Fourier transform of μ_I is proportional to the product of those of f_N and λ. Plugging in the estimators for μ_I and σ, the standard deviation of the normal distribution, transforming back and normalizing yields an estimator of λ. The resulting risk map $\hat{\lambda}$ is shown in Figure 10. The high risk areas reflect the geology, with high values of $\hat{\lambda}$ along the convergence zones (cf. Section 2). The region where the 1997 earthquake occurred is clearly visible, as are high risk area in the north.

8. Model validation

To validate the risk map (Figure 10), we take the leave-one-out approach. More specifically, for every year $i \in I = \{1973, \ldots, 2008\} \setminus \{1997, 2005\}$ we estimate the intensity function $\mu_{I\setminus\{i\}}$ and deconvolve to obtain another estimate $\widehat{\lambda_{(-i)}}$ of the risk map λ. A graph of

$$\int_W \left(\widehat{\lambda_{(-i)}(x,y)} - \widehat{\lambda(x,y)}\right)^2 dx\, dy,$$

where $\hat{\lambda}$ is the risk map estimated using all years except the two years in which a major earthquake occurred, plotted against i is given in Figure 11. The same figure also contains the integrated squared difference between the normalized intensity function μ (cf. Section 5) estimated using all years in I and using those in $I \setminus \{i\}$, $i \in I$, only.

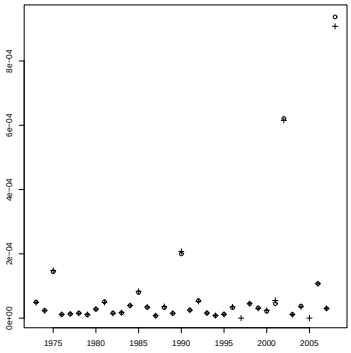

Fig. 11. Integrated squared difference between the earthquake risk map (circles) respectively intensity function (crosses) estimated from shallow earthquakes of magnitude at least 4.5 during the period from 1973 until 2008 but excluding the years 1997 and 2005 and that estimated by excluding also year i plotted against $i \in \{1973, \ldots, 2008\}$.

It can be seen from Figure 11 that the effect of leaving a year out is small. For most years, the integrated squared difference is smaller than 0.0002 or 2.2×10^{-6} per square degree latitude-longitude. For comparison, the mean value of $\hat{\lambda}$ is 0.01. The outliers are the years 2002 and 2008 in which clusters of shallow earthquakes occur.

9. Discussion

In this manuscript, we have addressed the occurrence of earthquakes at the national scale. We have focused on several questions that are relevant within this context. Earthquakes, like most natural disasters, are not forced to have their effects in a single country, and we have included as well disasters that occurred across the different borders, as much as these

might have an effect in the country. With the advent of a more inter-regional approach, however, a similar analysis might be done, where attention focuses, for example, on a group of countries around one particular fault. Adversely, a similar analysis might be applicable as well to a region within a country, that is in particular vulnerable.

An interesting issue that we discovered in passing when carrying out the analyses was that the major shock in 2005 could be modelled in a more convincing way when we applied a double model. The major shock was followed by another large shock, and both generated aftershocks in an almost perfect way. The second shock and all its aftershocks could, alternatively, have been included as a set of aftershocks of the first event. This raises the issue at which stage one should distinguish a 'shock;' from an 'aftershock'. Although these terms are intuitively clear, we were not successful in finding a sharp and unambiguous definition in the literature.

Relevant information may be included in this analysis, but that was not available to us when carrying out the analysis. In particular the transition of shock waves through the earth crust could serve to support the model that we applied. This would require additional data, in particular referring to the geological complexities. We felt that this was outside the scope of the current study, and may be of little relevance as well when the emphasis would be on the support after the occurrence of a disaster. The approach described in the paper could be modified by including such additional variation.

Acknowledgements

We are grateful to Ms. Salma Anwar, Mrs. Ellen-Wien Augustijn and Mr. Mark van der Meijde at the Faculty of Geo-Information Science and Earth Observation of Twente University who assisted us during the analysis. This research was done when the second author spent his sabbatical leave at the Centre for Mathematics and Computer Science (CWI) in Amsterdam. He is grateful for the hospitality received during that period.

Calculations were done in the software package `R`. The libraries `spatstat` (2) and `tseries` (22) were especially useful.

References

[1] Anwar, S., Stein, A. and Genderen, J. van (2011) Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks. In *Advances in Geo-Spatial Information Science* (ed W. Shi), ISPRS Book Series. London: Taylor & Francis.

[2] Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, 1–42.

[3] Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerlandica*, **54**, 329–350.

[4] Besag, J. and Diggle, P. J. (1977) Simple Monte Carlo tests for spatial pattern. *Appl. Statist.*, **26**, 327–333.

[5] Daley, D. J. and Vere–Jones, D.(2003) *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*, 2nd edn. New York: Springer–Verlag.

[6] Daley, D. J. and Vere–Jones, D.(2008) *An Introduction to the Theory of Point Processes. Volume I: General Theory and Structure*, 2nd edn. New York: Springer–Verlag.

[7] Diggle, P. J. (1979) On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics*, **35**, 87–101.

[8] Diggle, P. J. (1985) A kernel method for smoothing point process data. *Appl. Statist.*, **34**, 138–147.

[9] Gabriel, E. and Diggle, P. J. (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. *Statist. Neerlandica*, **63**, 43–51.

[10] Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010) *Handbook of Spatial Statistics*, Boca Raton: CRC.

[11] Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, **12**, 141–153.

[12] Holden, L., Sannan, S. and Bungum, H. (2003) A stochastic marked point process model for earthquakes. *Natural Hazards and Earth System Sciences*, **3**, 95–101.

[13] Jong, R. M. de (2000) A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory*, **16**, 262–268.

[14] Kwiatkowski, D., Phillips P. C. B., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**, 159–178.

[15] Lieshout, M. N. M. van (2011) A J-function for inhomogeneous point processes. *Statist. Neerlandica*, to appear.

[16] Molnar, P. and Chen, W. P. (1982) Seismicity and mountain building. In *Mountain Building Processes* (ed. K. J. Hsu), London: Academic Press.

[17] Newey, W. K. and West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

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

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

[20] Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in time series regression. *Biometrika*, **75**, 335–346.

[21] Stoyan, D., Kendall, W. S. and Mecke, J. (1995) *Stochastic Geometry and its Applications*, 2nd edn. Chichester: Wiley.

[22] Trapletti, A. and Hornik, K. (2009) *tseries: Time Series Analysis and Computational Finance*, R package version 0.10-22.

[23] Vere—Jones, D. (1970) Stochastic models for earthquake occurrence. *J. Roy. Statist. Soc. Ser. B*, **32**, 1–62.

[24] Vere–Jones, D. and Musmeci, F. (1992) A space-time clustering model for historical earthquakes. *Ann. Inst. Statist. Math.*, **44**, 1–11.

[25] Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002) Stochastic declustering of space-time earthquake occurrences. *J. Amer. Statist. Assoc.*, **97**, 369–380.