

Comparing the Hawkes and Trigger Process Models for Aftershock Sequences Following the 2005 Kashmir Earthquake

K. Türkyilmaz · M.N.M. van Lieshout · A. Stein

Received: 25 January 2012 / Accepted: 6 December 2012 / Published online: 15 January 2013
© International Association for Mathematical Geosciences 2013

Abstract In an earlier study (Van Lieshout and Stein in *Math Geosci* 44(3):309–326, 2012) we postulated the existence of two major earthquakes in the 2005 Kashmir disaster instead of a single one, based upon the pattern of aftershocks. In this study, we explore this hypothesis further by fitting several spatial point pattern models. In particular, we discuss the Hawkes and the trigger process models for earthquake aftershock sequences following the Kashmir catastrophe in 2005. The minimum contrast method is used for estimation of the parameters. The study shows that the trigger model fits better than the Hawkes model. The most likely number of main shocks is rounded to 2 generating the almost 200 aftershocks, whereas the Hawkes model would estimate a parent process of approximately 18 parents with on average about 10 descendants. We conclude that the spatial pattern of aftershocks can best be understood as a mixture of two bivariate normal distributions centered around two major shocks and estimate the parameters.

Keywords Cluster process · EM method · Minimum contrast method · Mixture model · Nearest neighbor distance distribution function · Pair correlation function

1 Introduction

Due to the orogenic belt caused by the collision of the Indo-Australian continental plate and the Eurasian plate, Pakistan is vulnerable to earthquakes. This vulnerability varies across the country (Van Lieshout and Stein 2012); a hot spot occurs

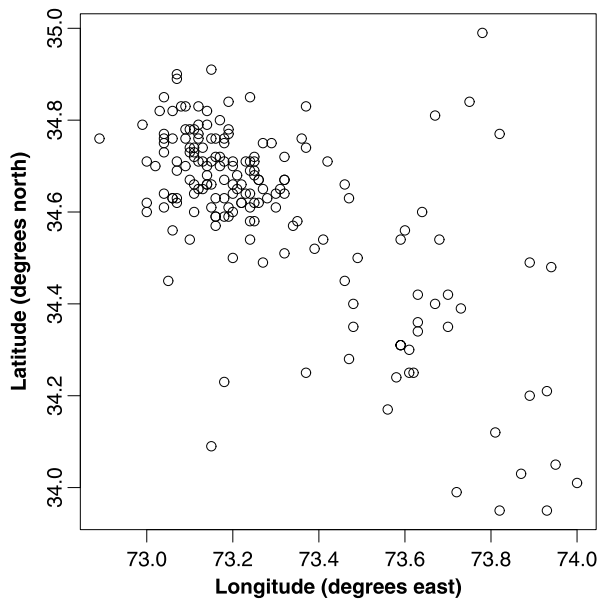
K. Türkyilmaz · M.N.M. van Lieshout
Probability, Networks and Algorithms, CWI, Amsterdam, The Netherlands

A. Stein (✉)
Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: a.stein@utwente.nl

in Pakistan-administered Kashmir, a region at which the two convergence zones associated with the subduction meet. Here, on the 8th of October 2005, a catastrophic earthquake occurred that resulted in at least 86,000 casualties. The Pakistan Meteorological Department estimated its magnitude at 5.2 on the Richter scale and the United States Geological Survey (USGS) measured it at least 7.6 on the moment magnitude scale. The epicenter was at approximately 19 km north-east of Muzaffarabad with a hypocenter at 26 km below the surface. The earthquake area is located along the great Himalayan main boundary thrust fault along the Himalayan arc (Avouac et al. 2006). Here, the Indian plate collides with the stable Eurasian plate. The fault, which generally runs parallel to the Himalayan mountains along the south eastern to north western direction, makes a rather peculiar bend in northern Kashmir, called the Hazara Syntaxis. Seen from the south-east, it forms a semicircle, then turns and close to the city of Islamabad continues in a western direction. The area of the semicircle itself shows a chaotic pattern of cracks and faults due to former energy releases and earthquakes. The main earthquake originally started close to the city of Muzaffarabad, some 30 km to the south-east of the northern most top of the semicircle. This major earthquake was a sudden release of seismic energy that progressed along the fault into south-eastern and north-western directions, releasing energy at many places that was registered as aftershocks: earthquakes that are generated shortly after a major earthquake (Lay and Wallace 1995). These aftershocks were apparently related to a fault plane that slipped during the event. The main shock introduces a major stress adjustment to a complex system by its sudden slip and regions within the rupture zone or adjacent to it may require adjustment to the new stress state, thus generating aftershocks. Aftershocks typically begin immediately after a main shock and are distributed throughout the source volume. In general, the largest aftershock is more than a magnitude unit smaller than the main shock (Båth's law). They can still be quite dangerous, though, due to the damage to structures caused by the main shock. In our study area, for example, a large build-up of energy took place in the north of the Hazara Syntaxis, leading to the complete destruction of the city of Balakot.

The cluster of aftershocks following the main event on October 8 was studied in Anwar et al. (2012) who concluded that the seismicity decreased so sharply that the number of earthquakes that occurred more than a month after the first shock was negligible. This picture was confirmed by the analysis carried out in Van Lieshout and Stein (2012) who found some rather surprising evidence that the spatial pattern formed by the earthquake locations in the period of October 8 to November 7, 2005, could well be described by two clusters rather than the expected single cluster: one corresponding to the main earthquake, the other to the next largest earthquake with a magnitude equal to 6.4 occurring approximately 7 hours later. The goal of the present paper is to investigate this hypothesis in further detail. To do so, we begin by describing the data and the two models most commonly used in seismology for the occurrence of aftershocks, namely the trigger and Hawkes processes, in Sect. 2. These two models are based on different interpretations of spatial patterns and the best fitting model would give insight into the process underlying the observed pattern. The trigger process is fit to the Kashmir data in Sect. 3, the Hawkes process in Sect. 4. For both models, a minimum contrast method based on the pair correlation function is used. Results prove that the trigger process with two clusters gives

Fig. 1 Locations of shallow earthquakes of magnitude 4.5 or higher recorded during the period October 8 to November 7, 2005



the better fit. This model is validated in Sect. 5. Based on these results, we include temporal information and the location of the main earthquakes to fit a mixture of two bivariate normal distributions by the EM algorithm in Sect. 6. The paper closes with discussion and conclusions.

2 Data and Models

The data consist of the locations and times of shallow earthquakes of magnitude 4.5 or higher at a depth of less than 70 km that happened between October 8 and November 7, 2005, in the Kashmir area. See Anwar et al. (2012) or Van Lieshout and Stein (2012) for further details of these data in a historical context. Figure 1 shows the pattern of the 176 spatial locations in the rectangular window $W = [72.65, 74.25] \times [33.70, 35.25]$, the position of which in Pakistan is indicated in Fig. 2. The bubble plot in Fig. 3 gives a visual impression of the spatial distribution of magnitudes. Figure 4 plots the frequency of earthquake occurrences against time (in hours counted from the 8th of October). The sharp decline in seismic activity is clearly visible; indeed 76 % of shocks happen within the first 48 hours from the main shock. The magnitude of earthquake occurrences against time (in hours counted from the 8th of October) is shown in Fig. 5, with Fig. 6 zooming in on those occurrences happening within the first 48 hours from the main shock. Both plots show two clear magnitude extremes of 7.6 (the main shock) and 6.4. From here on, therefore, it is assumed that Fig. 1 is an exhaustive map of the aftershocks with estimated spatial intensity $\hat{\lambda} = 176/|W| = 70.97$ per square degree. The natural model for data of the form described above is the Poisson cluster process X in the plane defined as the union

Fig. 2 Location of the observation window within Pakistan

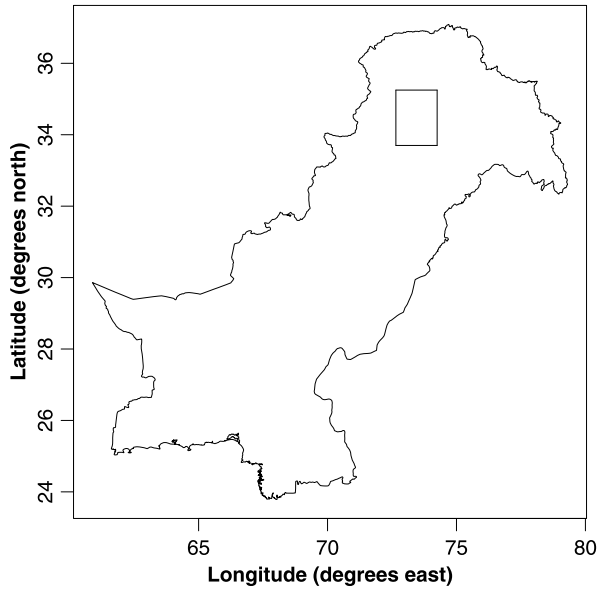
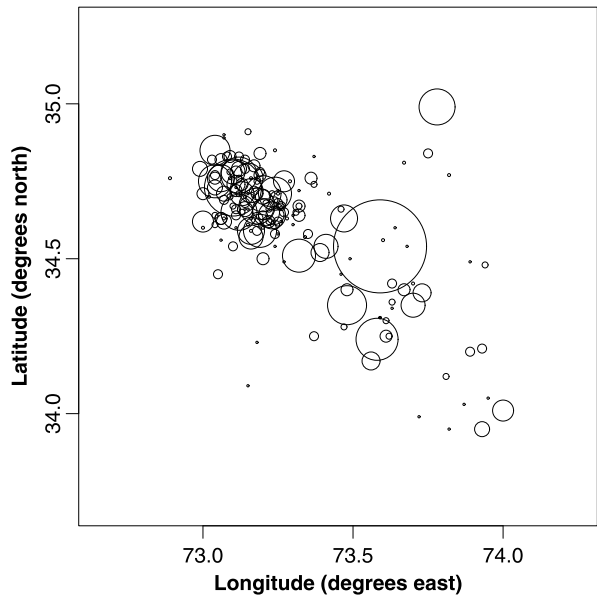


Fig. 3 Discs centered at the locations of shallow earthquakes of magnitude 4.5 or higher recorded during the period October 8 to November 7, 2005, with radius proportional to the excess magnitude



of independent, identically distributed clusters Z_x , centered around the points x of a stationary planar (marked) Poisson process Φ of parents, that is

$$X = \bigcup_{x \in \Phi} (x + Z_x). \tag{1}$$

Fig. 4 Frequency of occurrences of shallow earthquakes of magnitude 4.5 or higher recorded during the period October 8 to November 7, 2005, plotted against time (in hours)

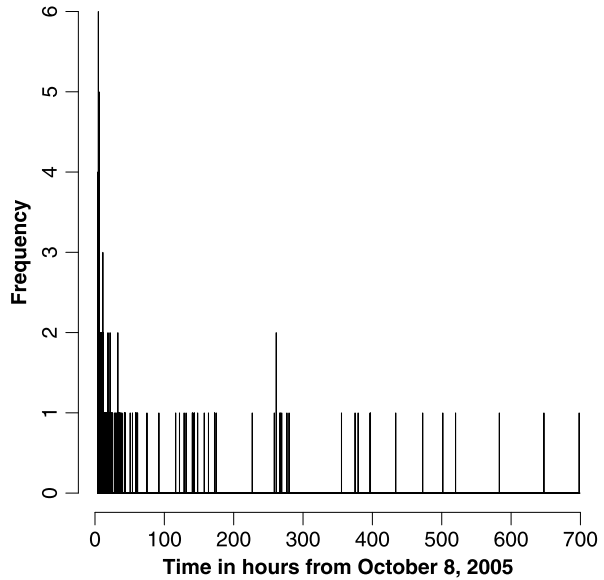
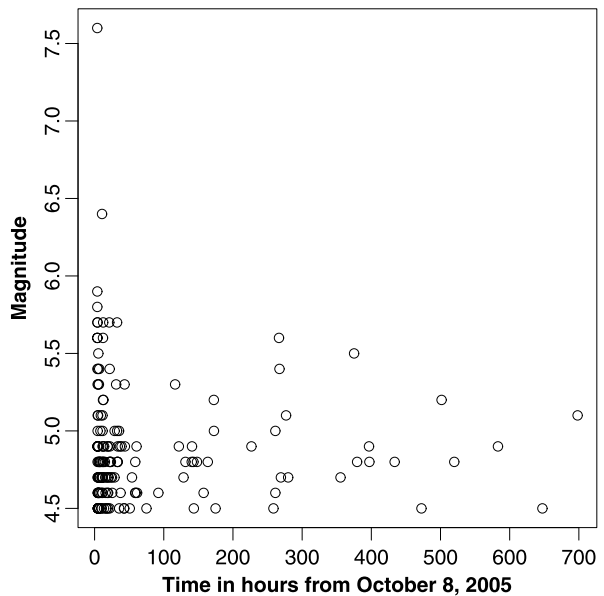
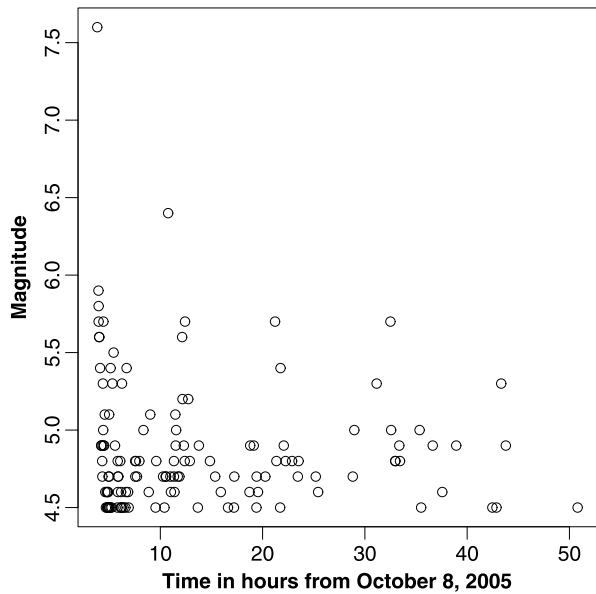


Fig. 5 Magnitude of occurrences of shallow earthquakes of magnitude 4.5 or higher recorded during the period October 8 to November 7, 2005, plotted against time (in hours)



If the Z_x consist of a random number of (marked) points that are scattered independently with identical distribution around the unobserved parent x , X is a Neyman–Scott process and, under the further assumption that the cluster size distribution is Poisson, X is known as a trigger process in the seismological literature (Adamopoulos 1976; Lomnitz and Hax 1966; Vere-Jones 1970; Vere-Jones and Davies 1966). Trigger processes are convenient to work with. Indeed, closed form expressions exist for many fundamental point process characteristics such as the intensity, the gener-

Fig. 6 Magnitude of occurrences of shallow earthquakes of magnitude 4.5 or higher recorded within 48 hours from the main Kashmir earthquake on October 8, 2005, plotted against time (in hours)



ating functional, Palm distribution, J -function, and pair correlation function (Daley and Vere-Jones 2003, 2008; Gelfand et al. 2010; Stoyan et al. 1995). This is not the case for the other subclass of Poisson cluster processes that dominates the seismology literature, namely that of Hawkes processes. In Hawkes processes, each parent independently generates a (marked) Poisson process of offspring with an intensity function that depends on the parent. The offspring in turn also generate offspring independently of all other offspring, and so on. Thus, Z_x is a branching process. An example is Ogata's Epidemic Type Aftershock-Sequences (ETAS) model, which is now the standard first approximation for seismic catalogue data that come in the form of a list of earthquake locations marked by time. An excellent review including historical references is Ogata (1998). The temporal component is important in that a conditional intensity can be written down. Consequently, a likelihood function is available in closed form (Ogata 1998). In the spatial case, on the other hand, only a series representation of the pair correlation function is available (Møller and Torrisi 2007) and it seems to be hard to generalize this expression to marked point processes except for predictable marks (Brémaud et al. 2005).

3 Fitting a Trigger Process

The first model we consider is a trigger process X in the plane. The parent process Φ is assumed to be a stationary Poisson process of intensity $\kappa > 0$, which in view of the small region under consideration, is a viable assumption to make. The number of offspring generated by each parent is taken to be Poisson with parameter $\nu > 0$. Conditional on the number of offspring, they are independently and identically distributed according to a bivariate normal distribution centered at the parent location

with covariance matrix $\sigma^2 I$, where I is the 2×2 identity matrix. This assumption is a simplification as the number and spread of aftershocks may well depend upon the magnitude of the parent shock. We shall get back to this issue in Sect. 6. As an aside, note that the trigger model is known in stochastic geometry as the modified Thomas process. The trigger process X is stationary with intensity $\lambda = \kappa \nu$ and has second-order product density (Stoyan et al. 1995)

$$\rho(x, y) = \lambda^2 + \frac{\kappa \nu^2}{4\pi\sigma^2} \exp[-\|x - y\|^2/(4\sigma^2)]. \tag{2}$$

Heuristically, $\rho(x, y) dx dy$ can be interpreted as the probability of points falling in each of two infinitesimal areas dx and dy . Note that $\rho(x, y)$ is a function of the distance $r = \|x - y\|$ alone. Upon standardization, one obtains the pair correlation function

$$g(r) = \frac{\rho(x, y)}{\lambda^2} = 1 + \frac{1}{4\pi\kappa\sigma^2} \exp[-r^2/(4\sigma^2)]. \tag{3}$$

Since $g(r) = g(r; \kappa, \sigma^2)$ does not depend on ν , we fix the theoretical intensity $\lambda = \kappa \nu$ at its empirical level $\hat{\lambda} = 176/|W|$. The minimum contrast method can then be used to estimate κ and σ^2 . More precisely, this method minimizes the integrated L_p distance between the q th power of the estimated pair correlation function \hat{g}^q and its model counterpart

$$\int_{r_1}^{r_2} |\hat{g}(r)^q - g(r; \kappa, \sigma^2)^q|^p dr, \tag{4}$$

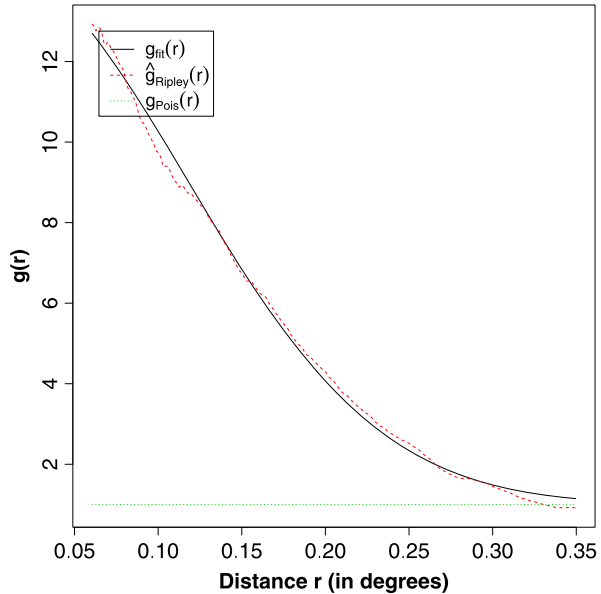
with respect to the parameters κ and σ^2 where the parameters p and q and the integration range from r_1 to r_2 are free to choose. In our study, we follow the rule of thumb in Illian et al. 2008 and set $q = 1/4, p = 2$. We set $[r_1, r_2]$ to $[0.06, 0.35]$. To estimate the pair correlation function, we consider the pattern of earthquakes, that is, the observation $\{x_1, \dots, x_n\}$ of X , that occur within the window W . With $\partial b(x_i, \|x_i - x_j\|)$ denoting the circle centered at x_i of radius $\|x_i - x_j\|$, the estimator is

$$\hat{g}_{\text{Ripley}}(r) = \frac{1}{(\hat{\lambda})^2} \frac{1}{2\pi r} \sum_{i=1}^n \sum_{j \neq i \in \{1, \dots, n\}} \frac{k(r - \|x_i - x_j\|)}{\alpha_{ij} |\{z : \partial b(z, \|x_i - x_j\|) \cap W \neq \emptyset\}|}. \tag{5}$$

Here, k is a kernel function, for example, the Epanechnikov kernel, α_{ij} is Ripley’s (1976) edge correction, that is, the proportion of the circle centered at x_i of radius $\|x_i - x_j\|$ that lies in W , and $|\cdot|$ denotes area (see also Ohser 1983). The estimator is appropriate for isotropic point processes X . Alternatives for anisotropic point processes can be found in Gelfand et al. (2010). Upon plugging in the estimated values $\hat{\kappa}$ and $\hat{\sigma}^2$ into $g(r; \kappa, \sigma^2)$, one obtains the fitted pair correlation function $g_{\text{fit}}(r) = g(r; \hat{\kappa}, \hat{\sigma}^2)$.

For the data in Fig. 1, the minimum contrast estimators are $\hat{\kappa} = 0.88, \hat{\sigma} = 0.08$ and $\hat{\nu} = 81$. Essentially, 2.17 parents corresponding to slightly over two major earthquakes are expected in the region with on average 81 offspring as aftershocks each scattered around their parent with standard deviation 0.08 for the displacements in latitude and longitude. The results should be compared to those of Van Lieshout and Stein (2012). These authors, working with a long time series of earthquake patterns,

Fig. 7 Fitted (solid line) and estimated (dashed line) pair correlation functions against distance for the trigger process

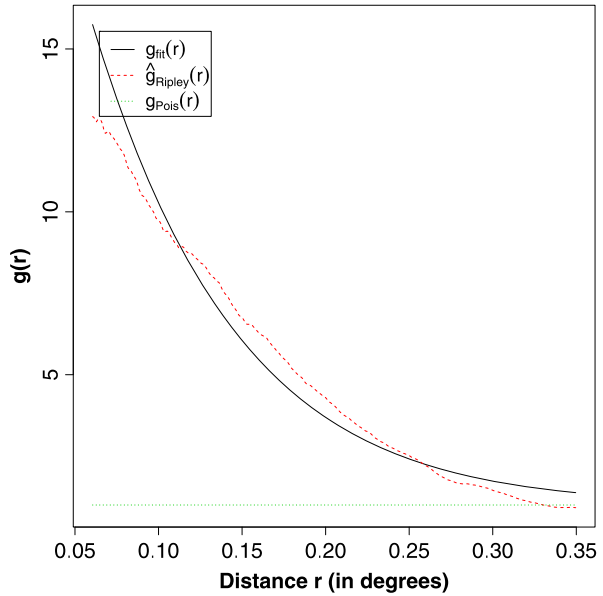


applied Fisher's linear discriminant function to assign earthquake locations in Kashmir in 2005 to two clusters and pooled the variance estimates with those of a diffuse swarm of aftershocks in another earthquake year. The longitude and latitude standard deviations ranged from 0.1 to 0.3, leading to a larger pooled estimate (0.19) for the displacements' standard deviation. The fitted and estimated pair correlation functions are shown in Fig. 7. They are fairly close and differ markedly from $g_{\text{Pois}}(r) \equiv 1$, the pair correlation function of a Poisson process without interpoint interaction.

4 Fitting a Hawkes Process

The second model that we consider is a planar Hawkes process. Again, we use the generic notation X for the point process that consists of all offspring and assume that the parent process Φ is a stationary Poisson process of intensity $\kappa > 0$. In contrast to the trigger processes of Sect. 3, the points of Φ form a subset of X , called generation zero. As for the Thomas process, each point of Φ generates a Poisson number of offspring with parameter $\nu > 0$. Conditional on the number of offspring, they are independently and identically distributed according to a bivariate normal distribution centered at the parent location with covariance matrix $\sigma^2 I$. The combined offspring of all parents in generation zero forms generation one. This construction is iterated: conditional on generation j , each of its points acts as parent and produces a Poisson number of offspring with parameter $\nu > 0$; conditional on the number of offspring, they are independently and identically distributed according to a bivariate normal distribution centered at the parent location with covariance matrix $\sigma^2 I$ and the combined offspring of all parents in generation j forms generation $j + 1$. The iteration terminates at the first empty generation. In order to ensure this is the case,

Fig. 8 Fitted (solid line) and estimated (dashed line) pair correlation functions against distance for the Hawkes process



assume $\nu > 1$. The point process X is stationary with intensity $\lambda = \kappa/(1 - \nu)$ and second-order product density (Møller and Torrisi 2007)

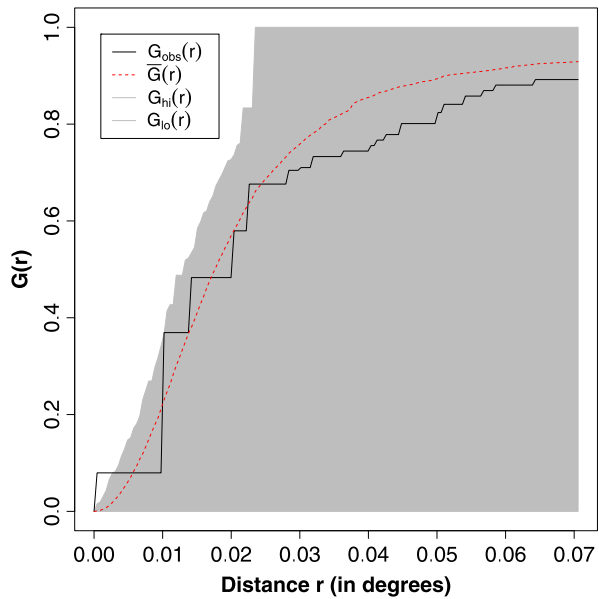
$$\rho(x, y) = \frac{\kappa^2}{(1 - \nu)^2} + \frac{\kappa}{1 - \nu} \sum_{n=1}^{\infty} (n + 1)\nu^n \frac{1}{2\pi n\sigma^2} \exp[-\|x - y\|^2/(2n\sigma^2)]. \tag{6}$$

Note that $\rho(x, y)$ is a function of the distance $r = \|x - y\|$ alone. Upon standardization, the pair correlation function is given by

$$g(r) = 1 + \frac{1 - \nu}{\kappa} \sum_{n=1}^{\infty} (n + 1)\nu^n \frac{1}{2\pi n\sigma^2} \exp[-r^2/(2n\sigma^2)]. \tag{7}$$

To estimate the parameters, we equate the intensity $\kappa/(1 - \nu)$ with its empirical counterpart $\hat{\lambda} = 176/|W|$ and use the minimum contrast method with $q = 1/4$, $p = 2$ and $[r_1, r_2] = [0.06, 0.35]$ as before. The parameter estimates are $\hat{\kappa} = 7.44$, $\hat{\sigma} = 0.03$, and $\hat{\nu} = 0.9$. Essentially, 18.45 parents are expected as major shocks in the region with on average $1/(1 - \nu) = 9.54$ descendants as aftershocks each. Note that due to the branching process nature of a Hawkes process, the estimate of the scatter standard deviation is smaller than that of the trigger process of Sect. 3, and that the number of parents is larger. The fitted and estimated pair correlation functions are shown in Fig. 8. The estimated line is apparently too high in the middle range and too low beyond, indicating that the branching structure of the Hawkes clusters may not be appropriate.

Fig. 9 Empirical (solid line) nearest neighbor distance distribution function with upper (G_{hi}) and lower (G_{lo}) envelopes over 100 simulations of the fitted trigger process against distance



5 Model Validation

From a visual comparison of Fig. 7 to Fig. 8, it is clear that a trigger process yields the better fit. To validate this model, we consider the nearest neighbor distance distribution function (Stoyan et al. 1995)

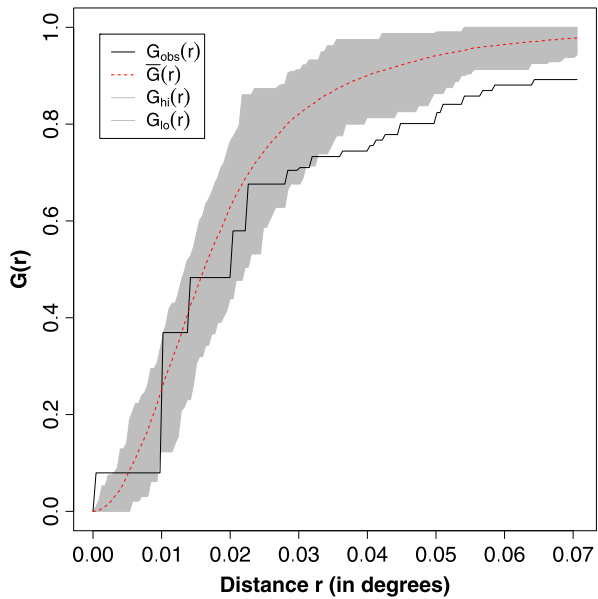
$$G(r) = \mathbb{P}(d(x, X \setminus \{x\}) \leq r | x \in X). \tag{8}$$

Since X is stationary, the definition does not depend on the choice of x . Figure 9 plots the empirical nearest neighbor distance distribution function $G_{obs}(r)$ (together with the upper and lower envelopes $G_{hi}(r)$ and $G_{lo}(r)$, respectively) based on 100 independent samples from the fitted model. The point-wise average of the sample estimates $\hat{G}_i(r), i = 1, \dots, 100$, is the dashed line $\bar{G}(r)$. It can be seen that $G_{obs}(r)$ lies almost entirely within the grey region bounded by the upper and lower envelopes, indicating a good fit (Besag and Diggle 1977; Diggle 1979). The lower envelope, however, takes the constant value zero due to the fact that the void probability (Stoyan et al. 1995)

$$v(W) = \exp \left[-\kappa \int [1 - e^{-vP(W_{-x}; \sigma^2)}] dx \right] \tag{9}$$

of finding no points in the rectangular region W is nonnegligible. Here, $P(\cdot; \sigma^2)$ denotes the bivariate normal distribution measure centered at the origin with covariance matrix $\sigma^2 I$, and W_{-x} is a translation over the vector $-x$ of the observation window W . Tighter envelopes can be obtained by using the information that an earthquake did happen on October 8th with the epicenter at $e = (73.59, 34.54)$ and using sampling conditional on having a parent at the epicenter while adapting κ in such a way that the expected number of parents in W remains constant, that is, $\tilde{\kappa} = 0.47$.

Fig. 10 Empirical (solid line) nearest neighbor distance distribution function with upper (G_{hi}) and lower (G_{lo}) envelopes over 100 simulations of the superposition of a cluster with normally distributed displacements centered at the epicenter of the main Kashmir shock and the fitted trigger process plotted against distance

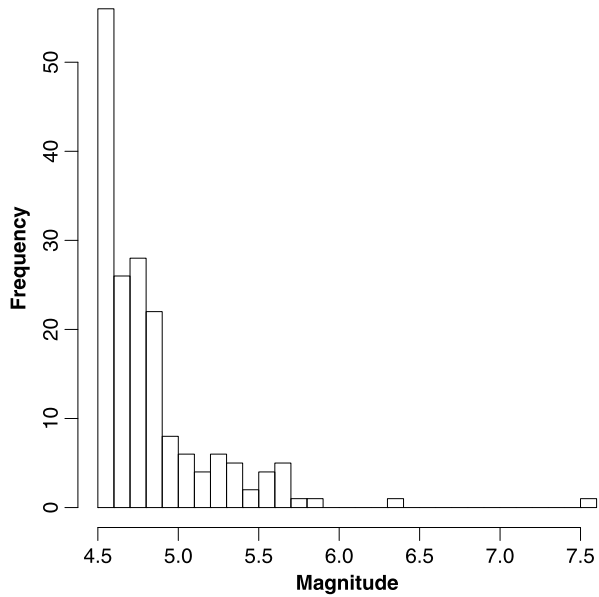


The result is shown in Fig. 10 though it should be stressed that the plot is not a model validation in the strict sense as conditioning affects the pair correlation function. Indeed, the conditional model is neither stationary nor isotropic. The estimates of κ and σ^2 depend on our initial choices of the parameters $r_1, r_2, p,$ and q . To check the sensitivity of our choices, we reestimated them for a range of settings. These proved to have hardly any effect on the outcome.

6 Gaussian Mixture Model

The conclusion drawn from the previous sections is that a trigger process with independent normally distributed offspring is a plausible model. We also found evidence for two clusters. This observation is supported by the histogram of magnitudes restricted to the range $[4.5, \infty)$ of shallow earthquakes in the month following the Kashmir disaster shown in Fig. 11. Indeed, two outliers are clearly visible and are separated by a wide gap from the bulk of the observations. The largest magnitude is associated with the main shock, the second largest with a severe aftershock some 7 hours after the main shock. In this section, we refine the trigger model by relaxing the assumption that all clusters follow the same normal distribution. Specifically, we fit a mixture of two bivariate normal distributions without any restrictions on the covariance matrices. The first cluster is centered around the main shock located at $\mu_1 = e = (73.59, 34.54)$. For the second cluster, we take as center point the location $\mu_2 = (73.10, 34.73)$ of the second largest shock. Naturally, all earthquakes happening before the second shock are allocated to the first cluster. We shall denote their locations by $\mathbf{y} = (y_1, \dots, y_m)$ where $m = 69$. For the other earthquake locations to be allocated, denoted by $\mathbf{x} = (x_1, \dots, x_n)$ with $n = 105$, we introduce latent variables (Z_1, \dots, Z_n) that indicate the cluster. Thus, if the random variable Z_i takes the

Fig. 11 Histogram of magnitudes



value 1, then x_i is allocated to the cluster at μ_1 , whereas if $Z_i = 2$ we allocate x_i to the cluster with center μ_2 . Let $f(\cdot | \mu, \Sigma)$ denote the bivariate normal probability density function with mean vector μ and covariance matrix Σ . Then the likelihood function $L(\theta; \mathbf{x}, \mathbf{y}, \mathbf{z})$ is equal to

$$\prod_{i=1}^m f(y_i | \mu_1, \Sigma_1) \prod_{i=1}^n [p \cdot 1_{\{z_i=1\}} \cdot f(x_i | \mu_1, \Sigma_1) + (1 - p) \cdot 1_{\{z_i=2\}} \cdot f(x_i | \mu_2, \Sigma_2)]. \tag{10}$$

The parameter vector $\theta = (p, \Sigma_1, \Sigma_2)$ consists of the unknown allocation probability $p \in (0, 1)$ and the cluster covariance matrices Σ_1 and Σ_2 . Essentially, the Z_i are assumed to be independent and identically Bernoulli distributed with $\mathbb{P}(Z_i = 1) = p$ and conditionally on $Z_i = j, j \in \{1, 2\}$, X_i follows a bivariate normal distribution with mean vector μ_j and covariance matrix Σ_j . To estimate the parameters, we use the EM algorithm (Dempster et al. 1977; Sundberg 1971; Wu 1983), an iterative procedure to approximate the maximum likelihood estimator $\hat{\theta}$. The essence of this method is to compute the expected log likelihood under the conditional expectation of the latent variables given the observed ones and the current value of the parameters, which is then optimized over the parameter vector. For our model, the new value of the allocation probability is the average over the x_i of the probability of x_i belonging to the first cluster under the current parameter values; the new value of the covariance matrices is the classic estimator weighted by the probabilities of the latent variables taking the appropriate value, again under the current parameter values. More precisely, set $t = 0$, initialize $p_t, \Sigma_{1,t}$, and $\Sigma_{2,t}$, and compute

$$\begin{aligned}
 p_{t+1} &= \frac{1}{n} \sum_{i=1}^n \frac{p_t f(x_i | \mu_1, \Sigma_{1,t})}{p_t f(x_i | \mu_1, \Sigma_{1,t}) + (1 - p_t) f(x_i | \mu_2, \Sigma_{2,t})}; \\
 \Sigma_{1,t+1} &= \frac{\sum_{i=1}^m (y_i - \mu_1)(y_i - \mu_1)^T + \sum_{i=1}^n \mathbb{P}(Z_i = 1 | X_i = x_i; \theta_t)(x_i - \mu_1)(x_i - \mu_1)^T}{m + \sum_{i=1}^n \mathbb{P}(Z_i = 1 | X_i = x_i; \theta_t)}; \\
 \Sigma_{2,t+1} &= \frac{\sum_{i=1}^n \mathbb{P}(Z_i = 2 | X_i = x_i; \theta_t)(x_i - \mu_2)(x_i - \mu_2)^T}{\sum_{i=1}^n \mathbb{P}(Z_i = 2 | X_i = x_i; \theta_t)},
 \end{aligned} \tag{11}$$

where

$$\mathbb{P}(Z_i = 1 | X_i = x_i; \theta_t) = \frac{p_t f(x_i | \mu_1, \Sigma_{1,t})}{p_t f(x_i | \mu_1, \Sigma_{1,t}) + (1 - p_t) f(x_i | \mu_2, \Sigma_{2,t})} \tag{12}$$

with $\mathbb{P}(Z_i = 2 | X_i = x_i; \theta_t) = 1 - \mathbb{P}(Z_i = 1 | X_i = x_i; \theta_t)$. The above steps are repeated for $t = 1, 2, \dots$ until the parameter estimates stabilize. Note that the first m earthquakes only affect the $\Sigma_{1,t}$ s.

For the Kashmir data of Sect. 2, we obtained $\hat{p} = 0.28$ and

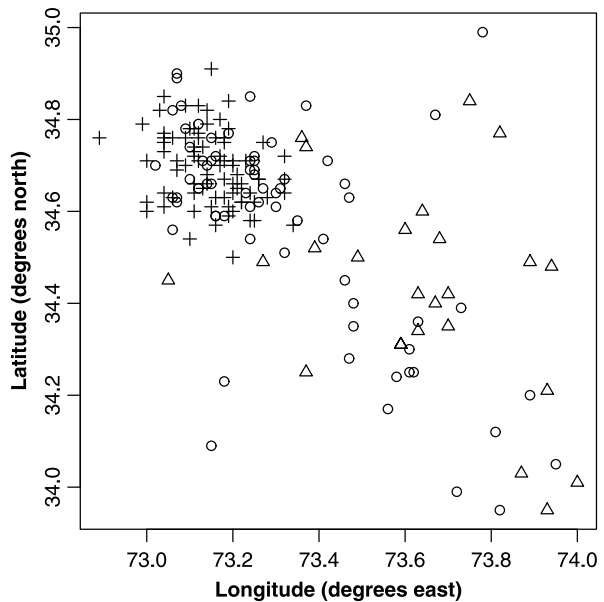
$$\hat{\Sigma}_1 = \begin{bmatrix} 0.110 & -0.039 \\ -0.039 & 0.056 \end{bmatrix}; \quad \hat{\Sigma}_2 = \begin{bmatrix} 0.0093 & -0.0037 \\ -0.0037 & 0.0077 \end{bmatrix} \tag{13}$$

after 1,000 steps. We took as initial allocation probability $p_0 = 1/2$ and covariance matrices $\Sigma_{1,0} = \Sigma_{2,0} = \hat{\sigma}^2 I_2$ where I_2 is the identity matrix and the multiplication constant $\hat{\sigma}^2 = 0.0068$ is the estimated variance in the trigger process of Sect. 3. Note that in both clusters, longitude and latitude are negatively correlated, reflecting the tilt of the convergence zone in northern Pakistan. These are caused by strike lines that are parallel to the main Himalayan thrust strikes. As expected, the cluster of aftershocks of the first shock is more widely scattered than the tighter pattern surrounding the second most severe earthquake. The variation in longitude is larger than that in latitude in both cases, though more pronounced in the main shock’s cluster. The clusters can be found by allocating an earthquake location x_i to the first cluster if $f(x_i | \mu_1, \hat{\Sigma}_1)/f(x_i | \mu_2, \hat{\Sigma}_2) > (1 - \hat{p})/\hat{p}$, that is, if the likelihood of the first cluster exceeds that of the second. The result is shown in Fig. 12. It differs from that obtained from Fisher’s linear discriminant function: the inclusion of temporal information and the relaxation of the equal variance assumption leads to more points in the north west being allocated to the main cluster. This phenomenon is the main explanation for the larger variance in longitude compared to latitude. The results also improve upon those of Kayabol (2011) that did not include the temporal dimension.

7 Discussion and Conclusions

In this paper, the trigger and Hawkes process models were fit to data on aftershocks following the Kashmir earthquake on 8 October, 2005. Although both Poisson cluster processes, these models differ in their offspring generating mechanism. In a trigger process, each unobserved parent gives rise to one generation of offspring, whereas in the Hawkes process multiple generations of offspring are formed according to a branching process.

Fig. 12 Locations of earthquakes recorded before the second largest shock (*circles*), those of later earthquakes allocated to the main cluster (*triangles*), and the cluster of the second largest shock (*crosses*)



The study led to several conclusions:

- (i) The three dimensional distribution of seismic waveforms in the underground exhibits a complicated 3-dimensional pattern. Registration of the depth of an earthquake is relatively imprecise. In the dataset at our disposal, often the same depth was given. To analyze the data in three dimensions, seismological models might be of help. Such models are being developed to properly understand the mechanism with which the energy waves proceed toward the earth's surface and generate earthquakes. Even if understood, such a pattern is difficult to validate and in the current study area most likely extremely complicated. In order to avoid such complications, we focused on a 2-dimensional point pattern analysis and restricted ourselves to shallow earthquakes.
- (ii) The trigger model fits the pattern of aftershocks better than the Hawkes process. Thus, the observed pattern can best be understood as being generated by approximately 2 parents, followed by 176 aftershocks in total. This picture is supported visually by the spatial pattern of the aftershocks. The Hawkes process model identifies a pattern of 18.45 main shocks with on average 9.54 aftershocks. The trigger model can be further improved upon by including temporal information and by relaxing the constraints on the covariance matrix.
- (iii) An unambiguous definition of aftershock does not exist. The existence of the second main shock, in particular, could as well be (and in fact most likely is) an aftershock of the first earthquake. However, this second major shock almost independently generates a clearly recognizable pattern of aftershocks. Increasing the number of main shocks to higher values such as 18.45 is possible, and the definition of aftershocks does not prohibit this. This combination of the numbers of main shocks and aftershocks though is less likely and intuitively less

appealing. Note that in order to describe the pattern in terms of a Hawkes process with an expected number of two main shocks, one would need 0.99 for the mean number of offspring which is dangerously close to the critical value of 1 at which the process explodes.

- (iv) The trigger process is based upon an entirely different assumption than the Hawkes process and, therefore, leads to a different interpretation of the mechanism generating earthquakes. This study shows that for the Kashmir data aftershocks are most likely generated by a Poisson cluster process with clusters of aftershocks scattered independently around the points of a stationary planar (marked) Poisson process of main shocks. The Hawkes model, that is, a branching process in which each main shock independently generates a (marked) Poisson process of aftershocks with an intensity function that depends on the parent, each of which again generates aftershocks independently of all others and so on, is less likely. This is remarkable, as the ETAS model (Ogata 1998) that is commonly used for aftershocks is a Hawkes process.
- (v) The study as presented can be characterized as a spatial statistical analysis of earthquake and aftershock data. The general impression that there was only one main shock in 2005 was already disputed (Van Lieshout and Stein 2012), and is further criticized by the current study. It all depends upon the definition of a main shock in relation to an aftershock. In principle, such a definition is somewhat arbitrary and it can be questioned whether an analysis with 2 main shocks is scientifically more sound than an analysis with one main shock, or with say 18 main shocks. The present analysis, however, shows that the spatial statistical modeling with point patterns and subsequent spatial distributional modeling of shocks and aftershocks conforms to the process of aftershock generation, and in that sense may help to separate the term of shock from the term aftershock.

Acknowledgements This research was supported by The Netherlands Organisation for Scientific Research NWO (613.000.809). The authors would like to acknowledge the contributions of Dr. Van der Meijde to this manuscript in strengthening its geological component, discussions on the seismology and pointing us to the paper of Avouac et al. (2006). Calculations were done using the R library *spatstat* (Baddeley and Turner 2005).

References

- Adamopoulos L (1976) Cluster models for earthquakes: regional comparisons. *Math Geol* 8(4):463–475
- Anwar S, Stein A, van Genderen, JL (2012) Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks. In: Shi W, Goodchild M, Lees B, Leung Y (eds) *Advances in geo-spatial information science*. CRC Press, Boca Raton, pp 125–140
- Avouac J-P, Ayoub F, Leprince S, Konca O, Helmberger DV (2006) The 2005, M_W 7.6 Kashmir earthquake: sub-pixel correlation of ASTER images and seismic waveforms analysis. *Earth Planet Sci Lett* 249(3–4):514–528
- Baddeley A, Turner R (2005) *Spatstat: an R package for analyzing spatial point patterns*. *J Stat Softw* 12(6):1–42
- Besag J, Diggle PJ (1977) Simple Monte Carlo tests for spatial pattern. *Appl Stat* 26(3):327–333
- Brémaud P, Massoulié L, Ridolfi A (2005) Power spectra of random spike field and related processes. *Adv Appl Probab* 37(4):1116–1146
- Daley DJ, Vere-Jones D (2003) *An introduction to the theory of point processes. Elementary theory and methods*, vol I, 2nd edn. Springer, New York

- Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes. General theory and structure, vol II, 2nd edn. Springer, New York
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39(1):1–38
- Diggle PJ (1979) On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* 35(1):87–101
- Gelfand AE, Diggle PJ, Fuentes M, Guttorp P (2010) Handbook of spatial statistics, Part IV. CRC Press, Boca Raton
- Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. Wiley, Chichester
- Kayabol K (2011) A latent variable Bayesian approach to spatial clustering with background noise. Technical report, CWI, Amsterdam
- Lay T, Wallace TC (1995) Modern global seismology. Academic Press, San Diego
- Lomnitz C, Hax A (1966) Clustering in aftershock sequences. In: Steinhart JS, Jefferson Smith T (eds) The earth beneath the continents. Amer Geophys Union, Washington, pp 502–508
- Møller J, Torrisi GL (2007) The pair correlation function of spatial Hawkes processes. *Stat Probab Lett* 77(10):995–1003
- Ogata Y (1998) Space-time point-process models for earthquake occurrences. *Ann Inst Stat Math* 50(2):379–402
- Ohser J (1983) On estimators for the reduced second moment measure of point processes. *Math Operforsch Stat, Ser Optim* 14(1):63–71
- Ripley BD (1976) The second-order analysis of stationary point processes. *J Appl Probab* 13(2):255–266
- Stoyan D, Kendall WS, Mecke J (1995) Stochastic geometry and its applications, 2nd edn. Wiley, Chichester
- Sundberg R (1971) Maximum likelihood theory and applications for distributions generated when observing a function of an exponential family variable. Dissertation, Stockholm University, Sweden
- van Lieshout MNM, Stein A (2012) Earthquake modelling at the country level using aggregated spatio-temporal point processes. *Math Geosci* 44(3):309–326
- Vere-Jones D (1970) Stochastic models for earthquake occurrence. *J R Stat Soc B* 32(1):1–62
- Vere-Jones D, Davies RD (1966) A statistical survey of earthquakes in the main seismic region of New Zealand. *NZ J Geol Geophys* 9:251–284
- Wu CFJ (1983) On the convergence properties of the EM algorithm. *Ann Stat* 11(1):95–103