


APPLICATION NOTE



kppmenet: combining the kppm and elastic net regularization for inhomogeneous Cox point process with correlated covariates

Achmad Choiruddin , Tabita Yuni Susanto, Ahmad Husain and Yuniar Mega Kartikasari

Departement of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

ABSTRACT

The `kppm` is a standard procedure to estimate the parameters of the inhomogeneous Cox point process. However, the procedure cannot handle the problem when the models involve correlated covariates. In this study, we develop the `kppmenet`, the modified version of the `kppm`, for the inhomogeneous Cox point process involving correlated covariates by considering elastic net regularization. We compare the methodology in a simulation study and apply it to model major-shallow earthquake distribution in Sumatra, Indonesia. We conclude that the `kppmenet` outperforms `kppm` when correlated covariates are involved.

ARTICLE HISTORY

Received 22 May 2022
Accepted 19 April 2023



KEYWORDS

Earthquake modelling; lasso; natural disaster; spatial point process; `spatstatR` package

1. Introduction

1.1. Background

Spatial point processes are stochastic processes which model point patterns distributed in space, such as random locations of earthquake occurrences, species of trees, disease cases, and crime events [2,4,12]. Cox point process is one of the main important point process models to analyse point patterns which tend to be clustered or aggregated due to its natural mechanism [8,24,25,29]. For parameter estimation, the Cox point process however requires heavy computation using Markov chain Monte Carlo (MCMC) due to its intractable likelihood function involving random intensity [21]. To tackle this issue, the popular `spatstat` R package procedure `kppm` [2,19,30] developed a two-step procedure where the parameters related to the first and second moments are estimated consecutively in two steps. In particular, the composite likelihood [22,29] is developed in the first step to estimate the first-order intensity. Given the estimated first-order intensity, minimum contrast estimation, second-order composite likelihood, or palm likelihood is derived to estimate the second-order intensity in the second step of the procedure, see [19,27,30]. Such a procedure performs well and satisfies asymptotic properties [30].

CONTACT Achmad Choiruddin  choiruddin@its.ac.id  Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Jl. Teknik Mesin No. 175 Surabaya, 60115, Indonesia

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

In recent years, more data and information such as spatial covariates are often included in the point pattern data analysis to enhance more insight and interpretation. To include the effect of spatial covariates on a Cox point process model, the random intensity involves parametric terms depending on spatial covariates, see Equations (2) and (6). Such an adaptation does not affect the estimation procedure, i.e. the two-step procedure (`kppm`) can still be performed for such a Cox point process, where the parameters quantifying the effect of spatial covariates are estimated in the first step through regression analysis [9,19,22]. However, when highly correlated covariates or a large number of covariates are involved, the procedure `kppm` is not able to tackle this issue (see Section 1.2 for more details).

In regression analysis, the regularization method has been developed to handle multicollinearity and perform variable selection. For example, elastic net regularization [16] combines ridge and lasso methods to prevent estimates suffering from large variances and to conduct variable selection. Recently, the regularization method has been developed for the spatial point process to model its intensity depending on a number of spatial covariates. In particular, the technique centres around the regularized first-order composite likelihood, see, e.g. [5,6,32]. This technique gains much attention and development in different contexts within the point process framework (e.g. [8,10,17,23]). However, since Cox point processes depend on the two groups of parameters related to the first-order and second-order characteristics of the point process, it is not clear how to employ the regularization procedure.

In this study, we extend the two-step procedure [19,30] for the inhomogeneous Cox point process involving correlated covariates. In particular, we modify the first step of the procedure by introducing elastic net regularization [16] and then mimic the second step by considering minimum contrast, composite likelihood, or palm likelihood. We term this procedure `kppmenet`. The `kppmenet` combines the `kppm` function of the `spatstat` package [2] and `glmnet` function of the `glmnet` package [13], which makes the procedure very practical. Apart from elastic net regularization, the `kppmenet` also covers ridge and lasso (and also its adaptive versions, see [5]), depending on the set-up of the tuning parameters (see Remark 2.1). The procedure `kppmenet` is integrated in the `spatstat` package procedure `kppm` by setting the option `ppm.improve.type = 'enet'`.

The remainder of the paper is organized as follows. We present in Section 1.2 the dataset motivating the study. Section 2 details the methodology, applied to the simulation experiment in Section 3 and real data analysis in Section 4. Conclusion is given in Section 5.

1.2. Dataset

The dataset includes the pattern of earthquake coordinates in the Sumatra region, Indonesia, and three geological variables. In particular, Figure 1 presents the locations of 6900 major-shallow earthquakes ($\text{magnitude} \geq 4M$, $\text{depths} \leq 60$ km) during 2004-2018 in the Sumatra ($D = [104.517, 119.295] \times [-6.518, 7.378]$ 100 km^2), along with the locations of volcano, fault, and subduction in the same region. These three variables are transformed into pixel images by measuring the distance of each coordinate within the region to each of the nearest geological variables and then treated as spatial covariates (see Figure 2). Using the point process framework, several studies [1,4,26] prove the effect of the geological

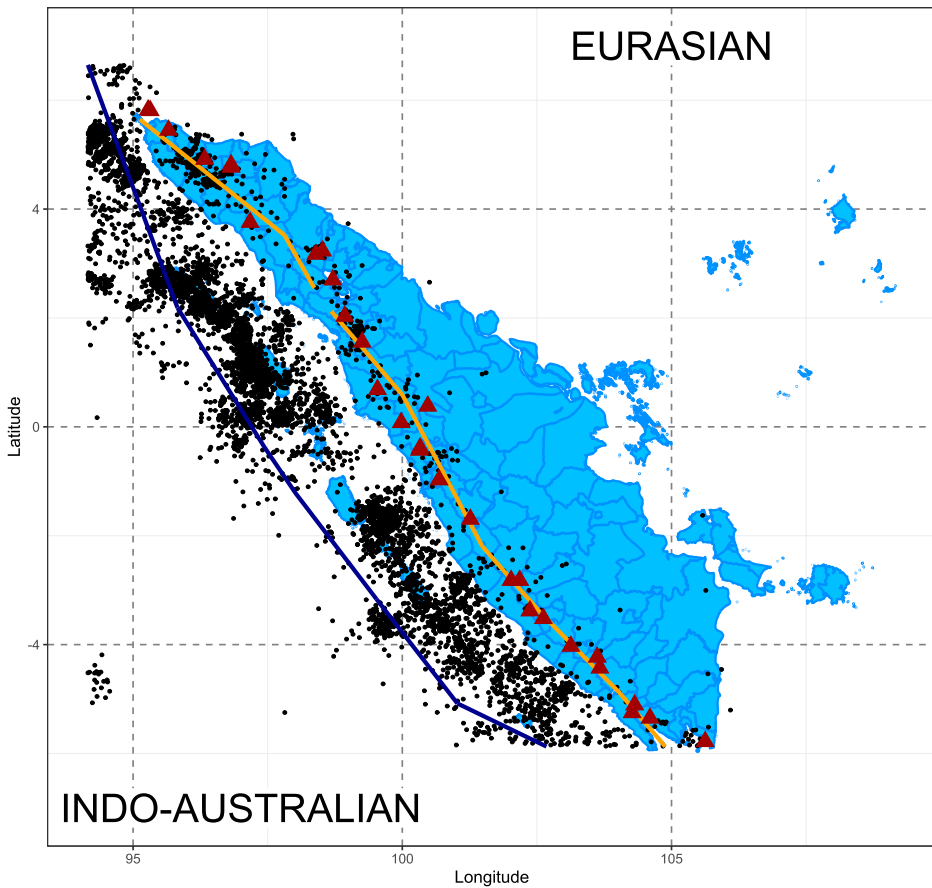


Figure 1. Locations of earthquake occurrences with $M \geq 4$ and depths ≥ 60 km in Sumatra (circle) during 2004–2018 and locations of volcano (triangle), fault (left-side line), and subduction zone (right-side line) in the same region. Note that volcano and fault concur closely each other.

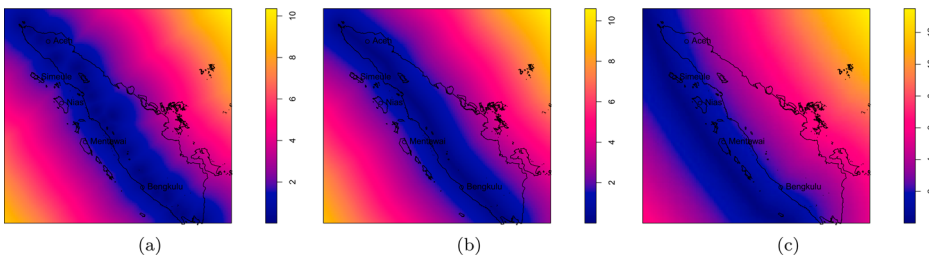


Figure 2. Spatial covariates representing the nearest distance per 100 km of each coordinate within the Sumatra region to the geological variables: (a) volcanoes, (b) faults, and (c) subduction zones.

factors on the distribution of earthquake occurrences respectively in Kashmir (Pakistan), Greece, and Sulawesi-Maluku region (Indonesia).

Previously, Choiruddin *et al.* [9] analysed the distribution of earthquake locations in Sumatra depicted in Figure 1 and further investigated the inhomogeneity due to three geological covariates and clustering due to main shock and aftershock activity using variants of Neyman–Scott Cox point processes. Although the models perform well in terms of the envelope test and intensity prediction, the model might have issue on interpretation and estimation due to multicollinearity because all geological covariates have strong correlation (see Figures 1 and 2 and Section 4). In particular, when all three geological variables are included in the model, the volcano, which is supposed to perform a negative impact, gives a positive effect (the closer the location to a volcano, the less risk of having a major-shallow earthquake). To avoid such an issue, volcano was eliminated in the model and was considered as an insignificant covariate. In this study, we propose a remedy to improve the analysis previously conducted by Choiruddin *et al.* [9] by performing the `kppmnet`. The results are discussed in Section 4.

2. Methodology

Let $\mathbf{x} = \{x_1, \dots, x_m\}$ denote a set of earthquake locations observed in a bounded region D , where for each $i = 1, \dots, m$, x_i represents the coordinate of the i th earthquake and m is the number of earthquake events in D . Let \mathbf{X} be a spatial point process on \mathbb{R}^2 which models the point pattern \mathbf{x} and has first- and second-order intensity functions λ and λ_2 . We refer the readers to [2,21] for more materials of the spatial point process.

In this study, \mathbf{X} is a class of Cox point processes involving covariates detailed in Section 2.1. Cox point processes have been considered for the analysis of earthquake distribution, whether or not covariates are incorporated in the model (e.g. [4,9,25,28]). We present the estimation technique used in the procedure `kppmnet` in Section 2.2.

2.1. Cox point processes

A Cox point process is defined in terms of a non-negative random process Λ , where given $\Lambda = \{\Lambda(u), u \in D\}$, the conditional distribution of \mathbf{X} is a Poisson point process with intensity Λ [21]. The first and second order intensity functions of a Cox process are

$$\lambda(u) = \mathbb{E}\Lambda(u), \quad \lambda_2(u, v) = \mathbb{E}[\Lambda(u)\Lambda(v)], \quad u, v \in D, \quad (1)$$

where the intensity $\lambda(u)$ quantifies the probability of observing an event or object in a coordinate u , and where the second order intensity $\lambda_2(u, v)$ measures the probability of observing a pair of points at u and v .

The two popular classes of Cox point processes are the log-Gaussian Cox process and Neyman–Scott Cox point process.

2.1.1. Log-Gaussian Cox process

Let \mathbf{G} be a Gaussian random field and $\Lambda(u) = \exp(\mathbf{G}(u))$. If conditionally on Λ , the point process \mathbf{X} is a Poisson process, then \mathbf{X} is a log-Gaussian Cox process (LGCP) driven by Λ [21,22].

Choiruddin *et al.* [4] argue that the random intensity Λ obtained from the LGCP could represent a random environment generating the aftershock events and could be

decomposed into different sources of variation. In particular, the log-linear model

$$\log \Lambda(u) = \boldsymbol{\beta}^\top \mathbf{z}(u) + \phi(u), \tag{2}$$

could represent the sources of variation due to observed geological covariates and unobserved variables. The first term includes $\boldsymbol{\beta}^\top \mathbf{z}(u)$, where $\mathbf{z}(u) = \{z_1(u), \dots, z_p(u)\}^\top$ is a vector of geological variables and $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_p\}^\top$ is the corresponding regression parameters. When $\mathbf{z}(u)$ involves many and/or correlated variables, there would be some issues in estimating $\boldsymbol{\beta}$ using the standard procedure. Next, the second term ϕ is a zero-mean stationary Gaussian random field with covariance function $c(u, v; \boldsymbol{\psi}) = \sigma^2 \exp(-\|u - v\|/\alpha)$, $u, v \in D$. Therefore, $\boldsymbol{\psi} = (\sigma^2, \alpha)^\top$ plays a role of an interaction parameter vector, where $\sigma^2 > 0$ is the variance and $\alpha > 0$ is the correlation scale parameter. By (1), the intensity function of an LGCP with random intensity (2) is

$$\lambda(u; \boldsymbol{\beta}) = \mathbb{E}\Lambda(u) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u) + \sigma^2/2) = \exp(\beta_0 + \boldsymbol{\beta}^\top \mathbf{z}(u)), \tag{3}$$

where $\beta_0 = \sigma^2/2$ [22,30].

2.1.2. Neyman–Scott Cox point processes

A Neyman–Scott Cox process (NSCP) is defined by $\mathbf{X} = \cup_{c \in \mathbf{C}} \mathbf{X}_c$, where \mathbf{C} is a ‘main-shock’ Poisson point process with intensity $\kappa > 0$. Given \mathbf{C} , $\mathbf{X}_c, c \in \mathbf{C}$ are ‘aftershock’ Poisson point processes with intensity $\lambda_c(u, \boldsymbol{\beta}), u \in D$ [4,21,30], where a log-linear form is employed to incorporate geological variables

$$\lambda_c(u; \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u))h(u - c; \omega). \tag{4}$$

More precisely, $\boldsymbol{\beta}^\top \mathbf{z}(u)$ plays the same role as the one in (2). The inclusion of a large number of and/or correlated covariates in the NSCP model would also cause a major issue. In addition, h is a probability density function specifying the distribution of aftershock points around the main shocks parameterized by a scaling parameter $\omega > 0$. For example, a bivariate Gaussian density

$$h(u - c; \omega) = (2\pi\omega^2)^{-1} \exp(-\|u - c\|^2/(2\omega^2)) \tag{5}$$

would result in a Thomas point process [21]. See, e.g. [4,19,27] for other variants of NSCP models. By (4), the NSCP \mathbf{X} is a Cox point process driven by a random intensity

$$\Lambda(u) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u)) \sum_{c \in \mathbf{C}} h(u - c, \omega). \tag{6}$$

The interaction parameter is constituted by $\boldsymbol{\psi} = (\kappa, \omega)^\top$, where smaller values of κ and ω lead to a stronger correlation (points tend to be closer and more clustered). Taking the $\mathbb{E}\Lambda(u)$, the intensity of an NSCP with random intensity (6) is

$$\lambda(u; \boldsymbol{\beta}) = \kappa \exp(\boldsymbol{\beta}^\top \mathbf{z}(u)) = \exp(\beta_0 + \boldsymbol{\beta}^\top \mathbf{z}(u)), \tag{7}$$

where $\beta_0 = \log \kappa$.

2.2. Parameter estimation

Suppose $\theta = (\beta^\top, \psi^\top)^\top$ is the parameter vector to estimate, decomposed into two vectors consisting of β and ψ . Such parameter vectors will be estimated sequentially in a two-step procedure. Here, we adjust the two-step estimation used in the procedure `kppm` [19,30] to develop the `kppmenet`.

2.2.1. Step 1: estimation of first-order parameter and selection of tuning parameter

The first step is to obtain β estimate, denoted by $\hat{\beta}$. To handle multicollinearity and perform covariate selection, we estimate β by maximizing the regularized composite likelihood

$$Q(\beta) = CL_1(\beta) - |D| \sum_{k=1}^p p_\rho(|\beta_k|), \tag{8}$$

where

$$CL_1(\beta) = \sum_{u \in X} \log \lambda(u; \beta) - \int_D \lambda(u; \beta) du \tag{9}$$

is the first-order composite likelihood function for β [22], $\lambda(u, \beta)$ is the intensity given by either (3) or (7), $|D|$ is the area of observation window D , and $p_\rho(\beta) = \rho(\gamma|\beta| + \frac{1}{2}(1 - \gamma)\beta^2)$, $\beta \in \mathbb{R}$, is the elastic net penalty function [16] with tuning parameters $\rho \geq 0$ and $\gamma \in [0, 1]$. Note that the elastic net reduces to ridge or lasso by setting $\gamma = 0$ or $\gamma = 1$. In addition, the penalty function can be extended into more general convex or non-convex function (see Remark 2.1). For a point process with a log-linear intensity, including the LGCP and NSCP with intensity (3) and (7), Choiruddin *et al.* [5,6] derive the asymptotic properties of $\hat{\beta}$. In addition, Choiruddin *et al.* [5,6] also demonstrate the numerical aspects of the procedure by combining the Berman–Turner approximation and coordinate descent algorithm.

The first step of the `kppmenet` relies on the selection of a proper tuning parameter ρ to avoid large bias or variance due to respectively the choice of too small or too large ρ . To specify $\rho \geq 0$, we define $[\rho_{max}, \rho_{min}]$ and select the one which minimizes the Bayesian information criteria [7] defined by

$$BIC(\rho) = -2CL_1(\hat{\beta}(\rho)) + q(\rho) \log(|D|), \tag{10}$$

where $CL_1(\hat{\beta}(\rho))$ is the maximum of (9) given ρ , $q(\rho)$ represents the number of non-zero elements of $\hat{\beta}(\rho)$, and $|D|$ is the area of observation window. We refer the reader to [5] for the details of defining $[\rho_{max}, \rho_{min}]$. Although two-dimensional BIC can be considered for the optimum search of both ρ and γ , ρ takes a more important role to control bias and variance of estimates [16,18]. Therefore, we treat γ as a fixed parameter to avoid more complex computation.

2.2.2. Step 2: estimation of second-order parameter

Given $\hat{\beta}$ obtained in the first step, we maximize the second-order composite likelihood with respect to ψ [2],

$$\begin{aligned}
 CL_2(\hat{\beta}, \psi) &= \sum_{\substack{\neq \\ u, v \in \mathbf{x}}} w(u, v) \log \lambda_2(u, v; \hat{\beta}, \psi) \\
 &\quad - \left(\sum_{\substack{\neq \\ u, v \in \mathbf{x}}} w(u, v) \right) \log \int_D \int_D w(u, v) \lambda_2(u, v; \hat{\beta}, \psi) du dv, \quad (11)
 \end{aligned}$$

where \neq over the summation sign indicates that the sum goes over all pairwise distinct points $u, v \in \mathbf{x}$, $w(u, v) = \mathbf{1}\{\|u - v\| < R\}$, $R > 0$ is an upper bound on the correlation distance model and $\lambda_2(u, v; \hat{\beta}, \psi)$ is the second-order intensity function of the specified Cox point processes described in Section 2.1 (see also [4,19,27]). The resulting estimator is denoted by $\hat{\psi}$. The alternatives are to replace (11) by minimum contrast or palm likelihood (see Remark 2.1). Using the two-step estimation (kppm), Waagepetersen and Guan [30] derive the joint-asymptotic properties of $(\hat{\beta}^\top, \hat{\psi}^\top)^\top$, where $\hat{\beta}$ is obtained from unregularized composite likelihood (9) and $\hat{\psi}$ is obtained from minimum contrast estimation. Combined with the asymptotic results of $\hat{\beta}$ obtained by Choiruddin *et al.* [5], the results of [30] might be adapted in our problem, see Remark 2.2. We end this section by the following remarks.

Remark 2.1: We highlight that our procedure can be extended into more general cases with no additional difficulty. First, the penalty term in (8) can be defined as a general convex or non-convex penalty function $p_\rho(\beta)$, see [5]. Second, the second step of the procedure (11) can be replaced by a palm likelihood or minimum contrast, see [27,30]. Our procedure kppmenet covers this generality. The kppmenet also covers the kppm by letting $\rho = 0$.

Remark 2.2: One may concern about the asymptotic properties of $\hat{\theta} = (\hat{\beta}^\top, \hat{\psi}^\top)^\top$ obtained from (8) and (11). A detailed study of asymptotic properties is outside the scope of this article and here we only sketch out the key arguments. To verify the joint properties of $\hat{\theta}$, one may extend the technique derived by Waagepetersen and Guan [30] and combined with the one studied by Choiruddin *et al.* [5,6].

We may view $\hat{\theta}$ as a solution of $\mathbf{u}(\theta) = \mathbf{0}$ where $\mathbf{u}(\theta) = (\mathbf{u}_1^\top(\beta), \mathbf{u}_2^\top(\psi))^\top$ is constructed by concatenating the gradients of respectively (8) and (11). Within increasing domain framework, one can prove under some conditions that $|D|^{1/2} \Sigma^{-1/2}(\hat{\theta} - \theta)$ is asymptotically distributed as $\mathbb{V}ar^{-1/2}[\mathbf{u}(\theta)]\mathbf{u}(\theta)$ and $\mathbb{V}ar^{-1/2}[\mathbf{u}(\theta)]\mathbf{u}(\theta)$ is asymptotically standard normal. Therefore, $|D|^{1/2}(\hat{\theta} - \theta)$ is asymptotically zero-mean normal with the asymptotic covariance matrix $\Sigma = |D|A^{-1}\mathbb{V}ar[\mathbf{u}(\theta)]A^{-1}$, where $A = -\mathbb{E}d\mathbf{u}(\theta)/d\theta$. Note that for a square matrix M , M^{-1} means the inverse of M and $M^{-1/2}$ is defined such that $M^{-1/2}(M^{-1/2})^\top = M^{-1}$.

3. Simulation study

To study the performance of the `kppmenet` estimates and compare it with the `kppm` estimates in a finite sample setting, we conduct a simulation study whose set-up is described in Section 3.1. The results are discussed in Section 3.2.

3.1. Simulation set-up

We set the Sumatra region (Section 1.2) to be the spatial domain D . In domain D , we generate spatial point patterns from a Neyman–Scott Cox point process with density (5), resulting in a Thomas point process model, see Section 2.1.2. We generate 500 point patterns with an average number of points equal to 750 using the `rThomas` function of the `spatstat` package. The cluster parameters $\boldsymbol{\psi} = (\kappa, \omega)^\top$ are $\kappa = 25$ or 5 ($\times 0.072$) and $\omega = 0.122$. We consider different values of $\boldsymbol{\psi}$ to observe the behaviour of the estimates for different cluster tightness. We do not present the results for different values of ω since we obtain similar results.

To consider a more realistic setting, we use three geological covariates from the real dataset depicted in Figure 2. The correlations among covariates are presented in Table 2. Following (7), the intensity function is

$$\lambda(u, \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 z_1(u) + \beta_2 z_2(u) + \beta_3 z_3(u)),$$

where $\beta_1 = -0.25$ represents the effect of distance to volcano (z_1), $\beta_2 = \beta_3 = -1$ represents the effect of distance to fault (z_2) or subduction (z_3), and β_0 is designed such that we obtain 750 points in average. The β_1 is set to have different magnitude to test how the procedure could handle the issue since the covariates are highly correlated.

All the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top)^\top = (\beta_0, \beta_1, \beta_2, \beta_3, \kappa, \omega)^\top$ are estimated using the two step procedure through the `kppm` and `kppmenet`. For the `kppmenet`, ridge, lasso, and elastic-net regularizations are considered by setting respectively $\gamma = 0$, $\gamma = 1$ or $\gamma = 0.5$. To compare the `kppm` and `kppmenet`, we consider the modified version of bias, standard deviation (SD), and root of mean squared errors (RMSE) of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\psi}}$ defined by

$$\text{Bias} = \left[\sum_{k=1}^p \left\{ \hat{E}(\hat{\theta}_k) - \theta_k \right\}^2 \right]^{\frac{1}{2}}, \quad \text{SD} = \left[\sum_{k=1}^p \hat{\sigma}_{\hat{\theta},k}^2 \right]^{\frac{1}{2}}, \quad \text{RMSE} = \left[\sum_{k=1}^p \hat{E}(\hat{\theta}_k - \theta_k)^2 \right]^{\frac{1}{2}},$$

where θ_k and $\hat{\theta}_k$ represent the true value and estimate of the k th element of either $\boldsymbol{\beta}$ or $\boldsymbol{\psi}$ of dimension p , and where $\hat{E}(\hat{\theta})$ and $\hat{\sigma}_{\hat{\theta}}^2$ are the empirical mean and variance of $\hat{\theta}$. In addition, we also compare the 10% and 90% quantiles of $\hat{\boldsymbol{\beta}}$ and the maximum value of the second-order composite likelihood (11). We seek the method yielding higher value of maximum second-order composite likelihood and smaller values of modified bias, SD, and RMSE. We do not compare the maximum first-order composite likelihood since the `kppm` and `kppmenet` use different optimization, see (8)–(9).

Table 1. The 10% and 90% quantiles of $\hat{\beta}$, along with the modified bias, standard deviation (SD), and RMSE of $\hat{\beta}$ and $\hat{\psi}$, maximum composite likelihood $CL_2(\hat{\beta}, \hat{\psi})(\times 10^5)$, and the empirical mean and variance of selected tuning parameter ρ obtained by the `kppm` and `kppmenet` (ridge, lasso, and elastic-net). The κ is scaled by 0.072.

			$[q_{10}, q_{90}]$		$\hat{\beta}$	$\hat{\psi}$	CL_2	$\hat{E}\rho$ ($\hat{V}ar\rho$)
$\kappa = 25$	<code>kppm</code>	$\hat{\beta}_1$	$[-1.74, -0.32]$	BIAS	0.04	0.08		
		$\hat{\beta}_2$	$[-0.93, 0.57]$	SD	0.79	0.78	3.60	0 (0)
		$\hat{\beta}_3$	$[-1.17, -0.79]$	RMSE	0.79	0.79		
	ridge	$\hat{\beta}_1$	$[-0.43, -0.35]$	BIAS	0.83	0.75		
		$\hat{\beta}_2$	$[-0.37, -0.28]$	SD	0.06	0.51	4.59	8.21 (4.54)
		$\hat{\beta}_3$	$[-0.49, -0.38]$	RMSE	0.84	0.91		
	enet	$\hat{\beta}_1$	$[-1.52, -0.38]$	BIAS	0.07	0.05		
		$\hat{\beta}_2$	$[-0.84, 0.47]$	SD	0.68	0.77	3.63	0.10 (0.002)
		$\hat{\beta}_3$	$[-1.16, -0.79]$	RMSE	0.68	0.77		
lasso	$\hat{\beta}_1$	$[-1.54, -0.34]$	BIAS	0.04	0.04			
	$\hat{\beta}_2$	$[-0.88, 0.46]$	SD	0.69	0.76	3.64	0.24 (0.04)	
	$\hat{\beta}_3$	$[-1.15, -0.78]$	RMSE	0.70	0.76			
$\kappa = 5$	<code>kppm</code>	$\hat{\beta}_1$	$[-2.32, 0.70]$	BIAS	0.22	0.17		
		$\hat{\beta}_2$	$[-2.01, 1.10]$	SD	1.72	0.21	5.37	0 (0)
		$\hat{\beta}_3$	$[-1.41, -0.64]$	RMSE	1.73	0.16		
	ridge	$\hat{\beta}_1$	$[-0.46, -0.28]$	BIAS	0.85	0.17		
		$\hat{\beta}_2$	$[-0.40, -0.24]$	SD	0.12	0.13	7.14	8.91 (25.53)
		$\hat{\beta}_3$	$[-0.52, -0.33]$	RMSE	0.85	0.12		
	enet	$\hat{\beta}_1$	$[-2.20, 0.59]$	BIAS	0.24	0.17		
		$\hat{\beta}_2$	$[-1.85, 0.98]$	SD	1.57	0.20	5.41	0.09(0.005)
		$\hat{\beta}_3$	$[-1.37, -0.62]$	RMSE	1.58	0.15		
lasso	$\hat{\beta}_1$	$[-2.20, 0.60]$	BIAS	0.23	0.17			
	$\hat{\beta}_2$	$[-1.86, 0.98]$	SD	1.58	0.20	5.41	0.20 (0.04)	
	$\hat{\beta}_3$	$[-1.38, -0.61]$	RMSE	1.59	0.15			

Table 2. Pixel correlation among geological variables.

	Volcano	Fault	Subduction
Volcano	1.00	–	–
Fault	0.99	1.00	–
Subduction	0.70	0.72	1.00

3.2. Simulation results

The modified bias, SD, and RMSE of β and ψ estimates are resumed in Table 1, along with the 10% and 90% quantiles of $\hat{\beta}$, the average of maximum values of the second-order composite likelihood, and the selection of tuning parameter ρ . The mean and variance of the selected ρ are slightly higher when the point pattern shows stronger clustering structure (smaller κ). The `kppmenet` with ridge tends to apply greater penalization than the one with elastic net and lasso by selecting higher values of ρ . This results in smaller SD. For the maximum composite likelihood CL_2 , the `kppmenet` with any regularization shows improvement over the `kppm`. In particular, the `kppmenet` with ridge regularization reaches the highest value of CL_2 . The value of CL_2 increases when κ gets smaller.

Regarding the parameter estimated results, we observe that the `kppmenet` outperforms the `kppm` by producing smaller RMSE. For the `kppmenet` with lasso or the elastic net, the procedure retains similar bias to the one produced by the `kppm`, but gains smaller

Table 3. The clustering estimates $\hat{\psi} = (\hat{\kappa}, \hat{\omega})$, the maximum composite likelihood CL_2 and AIC for each of fitted model to the Sumatra earthquake data using the `kppm` and `kppmenet` with ridge ($\gamma = 0$), elastic net ($\gamma = 0.5$), and lasso ($\gamma = 1$) regularizations.

	<code>kppm</code>	Ridge	Elastic Net	Lasso
$\hat{\kappa}$	0.3696	0.2281	0.3661	0.3663
$\hat{\omega}$	0.1554	0.1766	0.1557	0.1558
$CL_2 (\times 10^7)$	3.9388	4.7675	3.9516	3.9495
$AIC (\times 10^7)$	-7.8176	-9.4749	-7.8432	-7.8389

SD, resulting in producing smaller RMSE. For the `kppmenet` with ridge, the bias increases but the SD gets much smaller, leading to smaller RMSE.

When the point pattern gets more clustered (smaller κ), the overall metrics (bias, SD, RMSE) increase for $\hat{\beta}$ but decrease for $\hat{\psi}$. For a point pattern with a weaker clustering structure, the `kppmenet` with lasso or the elastic net is more preferable by producing the smallest RMSE, while the `kppmenet` with ridge performs best for a point pattern with strong clustering. Note that by employing `kppm` or `kppmenet` with lasso or the elastic net, the resulting β estimates may contradict the reality since these methods sometimes produce positive value of $\hat{\beta}_1$ and $\hat{\beta}_2$ (see 90% quantile of the estimator) while the true value is negative. This could be dangerous and affects model interpretation. Therefore, together with the comparison based on the maximum CL_2 , we would recommend employing the `kppmenet` with ridge regularization to model a point pattern which exhibits clustering and involves correlated covariates.

4. Application to earthquake data

To improve the analysis previously conducted by Choiruddin *et al.* [9], we first perform the pixel correlation among geological covariates depicted in Table 2. There exists a strong correlation between all pairs of geological variables (correlation is above 0.7). The fault and volcano are even almost perfectly correlated (0.99). Together with the issue presented in Section 1.2, it clearly indicates the multicollinearity problem.

We follow Choiruddin *et al.* [4,9] to model the earthquake occurrences in Sumatra by using the Cauchy cluster process since this model outperforms the other Cox models. The parameters are estimated using the `kppm` and `kppmenet` with ridge ($\gamma = 0$), elastic net ($\gamma = 0.5$), and lasso ($\gamma = 1$) regularizations, see Section 2.2.1. Table 3 presents the cluster parameter estimates, maximum composite likelihood values CL_2 , and AIC values. The AIC is defined by

$$AIC(\hat{\beta}, \hat{\psi}) = -2CL_2(\hat{\beta}, \hat{\psi}) + 2t,$$

where t is the dimension of $\hat{\theta} = (\hat{\beta}^\top, \hat{\psi}^\top)^\top$. The `kppmenet` with any regularization provides an advantage over the `kppm` by generating higher maximum CL_2 and lower AIC values. In particular, the `kppmenet` with ridge outperforms the other methods. The ridge regularization obtains different cluster parameters from the other methods by estimating the lowest main shock intensity (smallest $\hat{\kappa}$) but the highest scaling parameters (highest $\hat{\omega}$). By ridge regularization, there are 45 estimated main shocks (or 45 clusters). In

Table 4. The geological variables and their corresponding regression β estimates by the `kppmenet` with ridge regularization ($\alpha = 0$).

Parameter	Estimate ($\hat{\beta}$)	$\exp(\hat{\beta})$	$1 / \exp(\hat{\beta})$
Intercept ($\hat{\beta}_0$)	5.4348	229.2563	
Volcano ($\hat{\beta}_1$)	-0.0894	0.9145	1.0935
Fault ($\hat{\beta}_2$)	-0.1733	0.8409	1.1893
Subduction ($\hat{\beta}_3$)	-0.5569	0.5730	1.7452

addition, the aftershocks are scattered around each main shock with a scale of approximately 18 km (Table 3). The β estimates using the `kppmenet` with ridge are reported in Table 4. Except for $\hat{\beta}_0$, all other β estimates are negative, meaning that the closer a location to any of these three geological variables, the more likely great-shallow earthquakes will tend to occur. This is an indication that the `kppmenet` with ridge does the job to handle

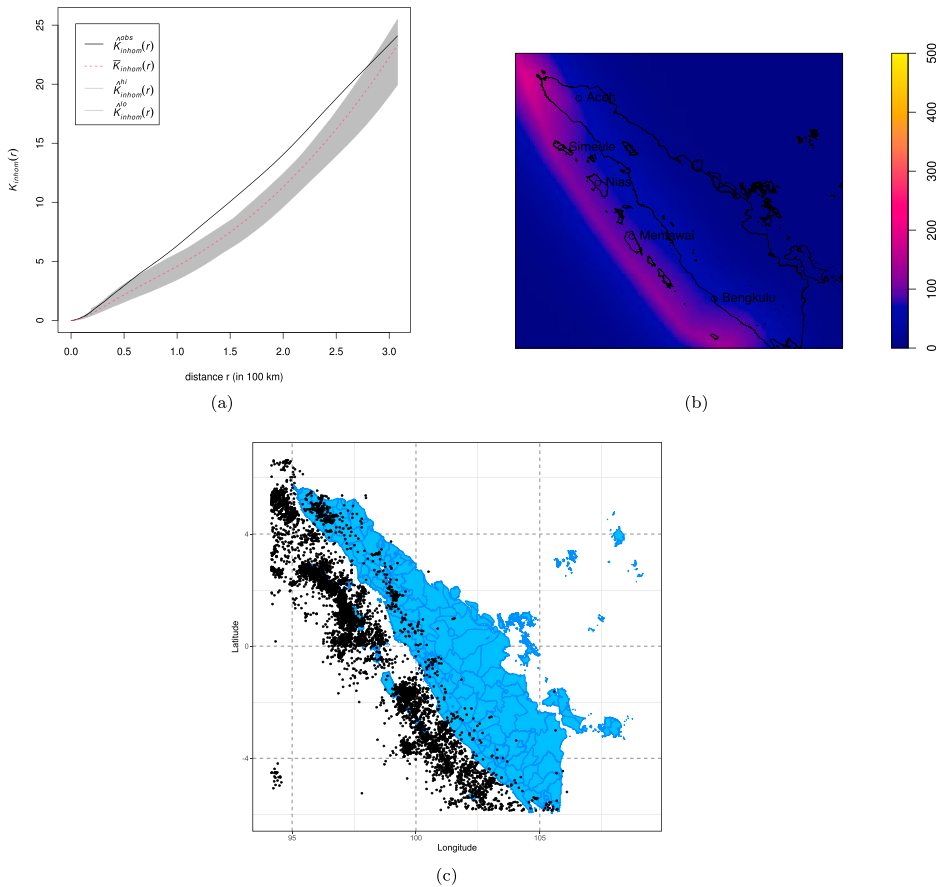


Figure 3. (a) The K -function envelope for the earthquake data in Sumatra based on `kppm` with ridge regularization, (b) predicted intensity map based on the Cauchy model fitted to the earthquake risk in Sumatra, and (c) distribution of the locations of earthquakes in Sumatra.

multicollinearity issues (see the previous problem in Section 1.2 where, e.g. the volcano produces a positive regression estimate). In particular, in an area with a distance of 100 km closer to a volcano, fault, or subduction, the risk of a major-shallow earthquake occurrence increases respectively by 1.09, 1.19, and 1.75 times. The subduction contributes the highest effect to the major-shallow earthquake occurrence in Sumatra by 75% since most of the great-shallow earthquakes in Sumatra are triggered by subduction movement, see [11,14]. Earthquake occurrence due to volcano activity is relatively rare in Sumatra, indicating the risk increases only by 9% when the distance of a location to the volcano gets closer by 100 km, see [20,31].

The plot of inhomogeneous K -function envelope and predicted intensity map are depicted in Figure 3. In general, the model fits the earthquake distribution well although the K -function envelope plot seems to be slightly underestimated within 150–200 km. The predicted intensity map is similar to the distribution of the major-shallow earthquakes in Sumatra, where three main areas have the highest risk of earthquake occurrences: (1) Aceh, (2) Simuelue, Nias, and Mentawai, and (3) Bengkulu. The results are inline with the previous finding [9]. However, in this study, we enrich the previous results by allowing the contribution of a volcano to the earthquake distribution using the `kppm` considering ridge regularization.

5. Conclusion

In this study, we introduce the `kppmenet`, the modified version of the `kppm`, for inhomogeneous Cox point process involving correlated covariates by considering elastic net regularization. The proposed method performs better than the standard `kppm` in simulated data and in application to the earthquake dataset when correlated covariates are involved. The `kppmenet` is integrated in the `spatstat` R package procedure `kppm` by setting the option `ppm.improve.type = 'enet'`.

In general, the `kppmenet` improves the `kppm` by producing more efficient estimates and higher values of maximum composite likelihood. However, the `kppmenet` tends to obtain more biased estimates, especially when ridge regularization is considered. To improve the estimates, one may consider approximating (9) by the logistic regression based technique [2,3]. Baddeley *et al.* [3] prove to obtain less biased estimates using such an approach in the unregularized estimation and Choiruddin *et al.* [5] further extend to the regularized setting. However, the logistic-based method is not yet covered in `kppm`. The development of `kppm` and `kppmenet` with such an approach could be a direction for future study.

The tuning parameter selection within the `kppmenet` is an important task. We have an empirical finding that the choice of ρ for the `kppmenet` with ridge is more varied (Section 3.2) might be due to improper way to define the degree of freedom. One may investigate the theoretical justification in future and consider, e.g. the composite version of the BIC (10) [7]. The variable selection procedure can also be derived from the Bayesian technique. The study by Giudici *et al.* [15] could serve as a basis for an extension to the spatial point process.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The study is supported by Institut Teknologi Sepuluh Nopember grant number 1292/PKS/ITS/2021. We thank the associate editor and three reviewers for the constructive comments.

ORCID

Achmad Choiruddin  <http://orcid.org/0000-0003-2568-2274>

References

- [1] S. Anwar, A. Stein, and J. van Genderen, *Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks*, In W. Shi, M. F. Goodchild, B. Lees, & Y. Leung (Eds.), *Adv. Geo-spat. Inf. Sci. (ISPRS book series)*, CRC Press (Taylor & Francis), (2012), pp. 125–140.
- [2] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R*, CRC Press, Boca Raton, FL, 2015.
- [3] A. Baddeley, J.F. Coeurjolly, E. Rubak, and R.P. Waagepetersen, *Logistic regression for spatial Gibbs point processes*, *Biometrika* 101 (2014), pp. 377–392.
- [4] A. Choiruddin, A.F. Trisnisa, and N. Iriawan, *Quantifying the effect of geological factors on distribution of earthquake occurrences by inhomogeneous Cox processes*, *Pure Appl. Geophys.* 178 (2021), pp. 1579–1592.
- [5] A. Choiruddin, J.F. Coeurjolly, and F. Letué, *Convex and non-convex regularization methods for spatial point processes intensity estimation*, *Electron. J. Stat.* 12 (2018), pp. 1210–1255.
- [6] A. Choiruddin, J.F. Coeurjolly, and F. Letué, *Adaptive Lasso and Dantzig selector for spatial point processes intensity estimation*, *Bernoulli* 29 (2023), pp. 1849–1876.
- [7] A. Choiruddin, J.F. Coeurjolly, and R. Waagepetersen, *Information criteria for inhomogeneous spatial point processes*, *Aust. N. Z. J. Stat.* 63 (2021), pp. 119–143.
- [8] A. Choiruddin, F. Cuevas-Pacheco, J.F. Coeurjolly, and R. Waagepetersen, *Regularized estimation for highly multivariate log Gaussian Cox processes*, *Stat. Comput.* 30 (2020), pp. 649–662.
- [9] A. Choiruddin, T.Y. Susanto, and R. Metrikasari, *Two-step estimation for modeling the earthquake occurrences in Sumatra by Neyman–Scott Cox point processes*, *Commun. Comput. Inf. Sci.* 1489 (2021), pp. 146–159.
- [10] J.F. Coeurjolly, F. Cuevas-Pacheco, and M.H. Descary, *A convolution type model for the intensity of spatial point processes applied to eye-movement data*, *Spatial Stat.* 51, October 2022, 100651. https://www.sciencedirect.com/science/article/abs/pii/S2211675322000331?casa_token=6uerZamJzHgAAAAA:6lEkK0COOy460MGFTG6lnHy_JS_G0MuOekyAWr0W3_MTdy3ugZNtPuTaZS2uBnZSJ2hT-jZ2hSR
- [11] R. Collings, D. Lange, A. Rietbrock, F. Tilmann, D. Natawidjaja, B. Suwargadi, M. Miller, and J. Saul, *Structure and seismogenic properties of the Mentawai segment of the sumatra subduction zone revealed by local earthquake travel time tomography*, *J. Geophys. Res. Solid Earth* 117 (2012), pp. 1–23.
- [12] P. Diggle, *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*, CRC Press, Boca Raton, FL, 2013.
- [13] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, *J. Stat. Softw.* 33 (2010), pp. 1–22.
- [14] V. Gahalaut, C. Subrahmanyam, B. Kundu, J. Catherine, and A. Ambikapathy, *Slow rupture in Andaman during 2004 Sumatra–Andaman earthquake: A probable consequence of subduction of 90° e ridge*, *Geophys. J. Int.* 180 (2010), pp. 1181–1186.

- [15] P. Giudici, M. Mezzetti, and P. Muliere, *Mixtures of products of Dirichlet processes for variable selection in survival analysis*, J. Stat. Plan. Inference. 111 (2003), pp. 101–115.
- [16] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press, Boca Raton, FL, 2015.
- [17] K.B. Hessellund, G. Xu, Y. Guan, and R. Waagepetersen, *Second-order semi-parametric inference for multivariate log Gaussian Cox processes*, J. R. Stat. Soc. Ser. C Appl. Stat. 71 (2022), pp. 244–268. DOI:10.1111/rssc.12530.
- [18] A. Husain and A. Choiruddin, *Poisson and logistic regressions for inhomogeneous multivariate point processes: A case study in the Barro Colorado Island plot*, Communications in Computer and Information Science, Springer Singapore, Singapore, 2021, pp. 301–311. DOI:10.1007/978-981-16-7334-4_22.
- [19] A. Jalilian, Y. Guan, and R. Waagepetersen, *Decomposition of variance for spatial Cox processes*, Scand. J. Stat. 40 (2013), pp. 119–137.
- [20] B. Kundu, D. Legrand, K. Gahalaut, V.K. Gahalaut, P. Mahesh, K. Kamesh Raju, J. Catherine, A. Ambikapathy, and R. Chadha, *The 2005 volcano-tectonic earthquake swarm in the andaman sea: Triggered by the 2004 great Sumatra-Andaman earthquake*, Tectonics 31 (2012), pp. 1–11.
- [21] J. Møller and R. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, CRC Press, Boca Raton, FL, 2003.
- [22] J. Møller and R. Waagepetersen, *Modern statistics for spatial point processes*, Scand. J. Stat. 34 (2007), pp. 643–684.
- [23] S. Rakshit, G. McSwiggan, G. Nair, and A. Baddeley, *Variable selection using penalised likelihoods for point patterns on a linear network*, Aust. N. Z. J. Stat. 63 (2021), pp. 417–454.
- [24] G. Shen, F. He, R. Waagepetersen, I.F. Sun, Z. Hao, Z.S. Chen, and M. Yu, *Quantifying effects of habitat heterogeneity and other clustering processes on spatial distributions of tree species*, Ecology 94 (2013), pp. 2436–2443.
- [25] M. Siino, G. Adelfio, and J. Mateu, *Joint second-order parameter estimation for spatio-temporal log-Gaussian Cox processes*, Stoch. Environ. Res. Risk. Assess. 32 (2018), pp. 3525–3539.
- [26] M. Siino, G. Adelfio, J. Mateu, M. Chiodi, and A. D’alessandro, *Spatial pattern analysis using hybrid models: An application to the hellenic seismicity*, Stoch. Environ. Res. Risk. Assess. 31 (2017), pp. 1633–1648.
- [27] U. Tanaka, Y. Ogata, and D. Stoyan, *Parameter estimation and model selection for Neyman–Scott point processes*, Biom. J. 50 (2008), pp. 43–57.
- [28] K. Türkyilmaz, M. van Lieshout, and A. Stein, *Comparing the Hawkes and trigger process models for aftershock sequences following the 2005 Kashmir earthquake*, Math. Geosci. 45 (2013), pp. 149–164.
- [29] R. Waagepetersen, *An estimating function approach to inference for inhomogeneous Neyman–Scott processes*, Biometrics 63 (2007), pp. 252–258.
- [30] R. Waagepetersen and Y. Guan, *Two-step estimation for inhomogeneous spatial point processes*, J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2009), pp. 685–702.
- [31] T.R. Walter and F. Amelung, *Volcanic eruptions following $m \geq 9$ megathrust earthquakes: Implications for the Sumatra-Andaman volcanoes*, Geology 35 (2007), pp. 539–542.
- [32] Y. Yue and J.M. Loh, *Variable selection for inhomogeneous spatial point process models*, Can. J. Stat. 43 (2015), pp. 288–305.