

# Spatial Analysis of Forest Fire Hotspots in Kalimantan Using Inhomogeneous Neyman-Scott Cluster Area

Berlian A. Kindhi<sup>1\*</sup>, Rifqi A. Mahmada<sup>2</sup>, Muhammad F. Zaqi<sup>2</sup>, Dimas A. Saputra<sup>2</sup>, and Narsischa S. Ningrum<sup>2</sup>

<sup>1</sup>Department of Electrical Automation Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

**Abstract.** Forest and land fires in Kalimantan pose a serious environmental challenge with significant impacts on ecosystems, public health, and the achievement of the Sustainable Development Goals (SDGs). In recent years, fire intensity has increased due to the combined effects of human activities, peatland characteristics, and extreme climate conditions associated with El Niño. Spatial analysis shows that fire hotspots are unevenly distributed, with a strong concentration in peatland areas. This study applies the Inhomogeneous Neyman–Scott Cluster Process (NSCP) to model the spatial clustering of hotspots under heterogeneous conditions by incorporating covariate information. The objective is to identify hotspot distribution patterns in Kalimantan during 2022–2024 and to generate fire risk maps to support targeted mitigation and environmental management policies. Model comparison using the Akaike Information Criterion (AIC) indicates that the Thomas Cluster Process with distance to BPBD facilities as a covariate provides the best fit, with the lowest AIC value (-456,944.7). The results reveal a negative relationship between distance to BPBD facilities and hotspot intensity, indicating that fire occurrences decrease with increasing distance. These findings confirm that hotspot distribution in Kalimantan is not random and highlight the importance of spatially informed fire management strategies.

## 1 Introduction

Forest and land fires are a serious environmental problem in Indonesia, particularly in Kalimantan, which is known as one of the world's tropical forest centres. This issue is closely related to the Sustainable Development Goals (SDGs), particularly SDG 13 on climate action and SDG 3 on good health and well-being. In recent years, forest fires in Kalimantan have become increasingly widespread. As reported in [1], during the period from January to April 2024, East Kalimantan was the most severely affected region, with more than 18,000 hectares of land burned, the largest amount compared to other provinces in Indonesia.

The main factors causing forest fires include peatland conversion, land clearing practices involving burning, and extreme climatic conditions such as El Niño, which prolongs the dry season [2]. Spatio-temporal analysis using hotspot data shows that dry peatlands are highly flammable and accelerate the spread of fire, thereby increasing the risk of fires in surrounding areas. As a result, forest fires have widespread impacts, ranging from ecosystem damage and loss of biodiversity to threats to public health due to haze pollution.

In addition to these widespread impacts, forest fires in Kalimantan tend to be concentrated in areas with high peat density. Several previous studies have examined forest fire patterns using various approaches. For example, satellite image-based analysis and hotspot data have been used to map fire-prone areas and identify environmental factors that influence the spread of fire [2, 3]. Other studies have conducted spatio-temporal analyses to understand the dynamics of fire events over time and the interactions between hotspots in different locations [4]. These results show that hotspots tend to form spatial clusters, where fires in one location can increase the risk of fires in surrounding areas.

---

\*Corresponding author: [berlian@its.ac.id](mailto:berlian@its.ac.id)

The pattern of fires forming clusters is relevant for analysis through a point process framework such as the Cox process, in which large fires can increase the likelihood of subsequent fires in the surrounding area due to the spread of heat or embers. Therefore, the Neyman-Scott Cluster Process (NSCP) is used, as this model is able to capture spatial intensity variations and provide a more accurate representation of fire patterns. Study [4] shows that NSCP is effective in mapping the frequency, intensity, and level of damage caused by fires in Kalimantan's peatlands, thus providing a basis for the development of more targeted forest fire risk maps. Previous studies have demonstrated the relevance of this approach. Study [5] developed a non-homogeneous Cox process to model the distribution of large earthquakes in Sulawesi and Maluku using geological variables. In the context of forest fires, existing studies are still limited. Study [6] examined fire hotspot patterns in Kalimantan using descriptive analysis based on SNPP-VIIRS satellite imagery. However, this study did not use a point process approach or produce forest fire risk maps. Furthermore, study [7] conducted a combined spatio-temporal analysis of hotspots, burned area, and carbon emissions with drought indicators; however, this study emphasised the relationship between climate and forest fires rather than point process-based spatial intensity modelling.

## 2 Literature Review

### 2.1 Fire Forest

In Indonesia, forest fires are a common environmental calamity, especially in Kalimantan, which is home to extensive tropical forests and peatland ecosystems. Study [8] states that peatland fires are more severe because they spread underground and are challenging to put out, while these fires harm vegetation and biodiversity in addition to degrading the environment through increased carbon emissions, soil damage, and health issues from smoke. Haze lowers air quality, causes respiratory illnesses, interferes with community activities, and even has cross-border effects, according to socio economic perspectives. Climate variability, particularly the El Niño phenomenon that lowers water levels and prolongs dryness, exacerbates human activities like land clearance for plantations, peatland drainage, and land burning [3, 4]. Thus, forest fires in Kalimantan are regarded as both ecological and social calamities, necessitating community empowerment and sustainable land management in addition to technological suppression. As of October 2023, the affected area in Indonesia was 994.313 hectares, with Kalimantan being the most affected, according to KLHK data. This led to a number of prevention and monitoring initiatives, including the SiPongi+ portal and policies under Regulation No. 32 of 2016 on Forest and Land Fire Control, which place an emphasis on law enforcement, integrated patrols, and monitoring.

### 2.2 Spatial Point Process

Spatial Point Process (SPP) is a stochastic framework for representing the positions of events as a collection of points in  $R^2$ . This analysis aims to describe the intensity (first order), the pattern of dependence between points (second order), and the relationship between point patterns and spatial variables. Where the realization of this process is represented as a set of points  $\{x_i\}$ , so this model is suitable for various ecological and environmental applications because it intuitively distinguishes between regular patterns (inhibition), clustering, and random patterns (Poisson) [9]. First-order statistics focus on intensity  $\lambda(x)$ , which is defined such that for a small area  $B$  around location  $x$ .

$$E[N(B)] = \lambda(x) \cdot |B| \quad (1)$$

In a homogeneous process,  $\lambda(x)$  is constant with respect to  $(\lambda)$ , while the intensity of an inhomogeneous process varies depending on location and is often modeled using regression against spatial covariates,  $\lambda(x) = \exp(\beta^T z(x))$ . Prior to second-order analysis, it is generally necessary to estimate  $\lambda(x)$  (e.g., smoothing kernel or GLM/GAM model) to prevent misinterpretation of clustering patterns that are caused by intensity inhomogeneity [10].

Second-order statistics focus on interactions between points. The probability of finding points in a limited area around  $x$  and  $y$  at the same time is estimated by  $\int x \int y \rho^2(x, y) dx dy$ , which is the definition of the second order product density  $\rho^2(x, y)$ . The pair correlation function (PCF) is obtained from  $\rho^2$  as below.

$$g(x, y) = \frac{\rho^2(x, y)}{(\lambda(x)\lambda(y))} \tag{2}$$

This PCF becomes a function of distance  $r = \|x - y\|$  for isotropic/stationary processes denoted by  $(r)$ . If  $(r) > 1$ , there are more pairs of points at distance  $r$  (clustering), whereas is  $(r) < 1$ , there are fewer pairs (inhibition).

The Ripley's K-Function is a popular second-order summary besides PCF. For stationary processes, this function is defined as follow.

$$K(r) = \frac{1}{\lambda} \mathbb{E}[N(B)] \tag{3}$$

Where the expected complete spatial randomness CSR is  $(r) = \pi r^2$ . The relationship between  $(r)$  and  $(r)$  in two dimensions is expressed below.

$$g(r) = \frac{K'(r)}{2\pi r} \tag{4}$$

To prevent the actual clustering signal from being obscured by uneven intensity patterns, it is important to use the appropriate intensity-reweighted version of the K-Function and an estimate of  $g(r)$  that takes into account the variation  $\lambda(x)$  for inhomogeneous data [11].

The inhomogeneous Kfunction was developed to address the limitation of Ripley's K-function, which assumes that point intensity is constant across space. In real-world cases like forest fire hotspots, this assumption is unrealistic since fire risks vary significantly due to factors such as peatland distribution, precipitation, and human activity. Introduced by study, the inhomogeneous K function adjusts for this non-stationarity by weighting point pairs with an estimated intensity,  $\lambda(x)$ . This correction makes it possible to separate first-order effects (environmental influences on fire density) from second-order effects (spatial clustering that reflects fire spread or local interactions). Inhomogeneous K-function  $K_{inhom}(r)$  for the Neyman-Scott family in the case where the parent (cluster centers) process is homogeneous (constant intensity  $\kappa_0$ ), then the simplification that occurs when parents are homogeneous, and give the Thomas (Gaussian) closed form as an example plus practical note. For a point process if  $N \sim Poiss(\mu)$  then  $v = \mu^2$  and the prefactor simplifies to  $\frac{1}{\kappa_0}$ . So with intensity function  $(x)$  we can express.

$$K_{inhom}(r) = \pi r^2 + \frac{1}{\kappa_0} \int_{\|h\| \leq r} (f * f)(h) dh \tag{4}$$

Where  $(f * f)(h)$  denotes the convolution of the offspring density with itself, capturing the contribution of point pairs that originate from the same cluster. The first term  $\pi r^2$  represents the theoretical K-function of a completely spatially random (CSR) process, while the second term introduces an additional component due to clustering. In empirical applications,  $(x)$  and its estimate  $\hat{\lambda}(x)$  represent the spatial intensity function and its nonparametric estimate, respectively, and  $W$  denotes the observation window within which the process is observed. Estimating  $K_{inhom}(r)$  thus involves correcting for spatial variation in point intensity by weighting each observed pair inversely by the product  $\lambda(x)\lambda(y)$ , ensuring that the result reflects second-order clustering rather than large-scale intensity inhomogeneity.

### 2.3 Inhomogeneous Neyman-Scott

An extensive overview of the theory of point processes and Neyman–Scott processes is presented, among others, by study [11]. In essence, a Neyman–Scott process  $X$  can be described as a union of clusters  $\cup_{c \in C} X_c$ , where  $C$  is a Poisson process of centers of clusters (parent process). Given  $C$ , the clusters  $X_c$ ,  $c \in C$  are independent Poisson processes, with an intensity function  $\lambda(u - c, \omega)$ , where  $k(u, \omega)$  is a probability density function parameterized by  $\omega$ , which determines the spread of daughter points around its mother and  $a$  is the expected number of daughters per cluster. If  $\phi(u)$  is the density of a symmetric normal distribution, then the process is called a modified Thomas Process [12].

An inhomogeneous Neyman–Scott process is often formulated by thinning the daughter points, producing what is known as a second order intensity reweighted stationary model. Our formulation differs by introducing inhomogeneity at the cluster center level  $C$ , treating  $C$  as an inhomogeneous Poisson process governed by an intensity function  $\lambda(u) : \kappa \exp(\beta^T z(u))$ . Where  $z = (z_1, \dots, z_k)$  is a vector of covariates and  $\beta = (\beta_1, \dots, \beta_k)$  is a regression parameter.

The intensity of the Neyman-Scott point process with inhomogeneous cluster centers is then.

$$\begin{aligned} \lambda(u) &= \int_{\mathbb{R}^2} \lambda(c) \mathbb{E}[N_c] f(u - c) dc \\ &= \mu \int \lambda(c) \phi(u - c) dc, u \in \mathbb{R}^2 \\ &= (\kappa * f)(u) \end{aligned} \tag{5}$$

where  $(\kappa * f)(u)$  denotes the spatial convolution of the parent intensity function  $\kappa$  with the offspring kernel  $f$ . This expression shows that the intensity of the observed process is a smoothed version of the underlying parent intensity, with the degree of smoothing determined by the spread of the offspring kernel  $f$ . Consequently, spatial variation in  $\lambda(c)$  is transferred to the offspring process through convolution, producing inhomogeneous clustering that reflects both the spatial heterogeneity of parent centers and the dispersal properties of the offspring distribution.

Thomas Cluster Process, Fire spread patterns are distributed around the ignition sources following a bivariate normal with variance parameter  $\sigma^2$  (isotropic) [13]. The density function is given by

$$f(u) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right) \tag{6}$$

Where  $\sigma^2$  denotes the variance of the Gaussian distribution, and  $\|u\|$  is the Euclidean distance between an offspring and its parent. This kernel ensures that offspring are symmetrically distributed around the parent center with expected radial dispersion proportional to  $\sigma$ . Smaller values of  $\sigma$  indicate that offspring points are more tightly concentrated near their parent, forming denser clusters, whereas larger  $\sigma$  values correspond to broader, more diffuse clustering patterns. Since  $f(u)$  integrates to one over the plane, it serves as a valid spatial probability density governing the offspring displacement mechanism in the homogeneous parent Neyman–Scott process. For the Thomas cluster process, the pair correlation function is expressed as

$$g(r, \sigma) = 1 + \frac{1}{4\pi\kappa\sigma^2} \exp\left(-\frac{r^2}{4\sigma^2}\right), r = \|u - v\| \tag{7}$$

while the K-function is given by

$$K(r) = \pi r^2 + \frac{1}{\kappa} \left\{1 - \exp\left(-\frac{r^2}{4\sigma^2}\right)\right\} \tag{8}$$

To estimate the estimation of inhomogeneous clustered processes is the first-order composite log-likelihood (treating observed points as an inhomogeneous Poisson process with intensity  $\lambda(x; \beta)$  and ignoring interactions). When parameterize  $(c; \beta)$  (parent inhomogeneity parameters  $\beta$ ) and cluster parameters  $\theta$  appear in  $(u; \theta)$  and  $\mu(\theta)$ , the first-order composite log-likelihood is.

$$\ell_{c1}(\beta) = \sum_{xi \in X \cap W} \log(\lambda(xi; \beta)) - \int_W \lambda(u; \beta) du \tag{9}$$

To maximize (7) we can use Hessian and the entries (8).

$$\frac{\partial^2 \ell_{c1}}{\partial \theta_j \partial \theta_k} = \sum_{i=1}^n \frac{\partial \theta_j \lambda(xi)}{\lambda(xi)^2} \frac{\partial \theta_k \lambda(xi)}{\lambda(xi)} - \sum_{i=1}^n \frac{\partial^2 \lambda(xi)}{\partial \theta_j \partial \theta_k} \int_W \frac{\partial^2 \lambda(u)}{\partial \theta_j \partial \theta_k} du \tag{10}$$

captures both the local curvature at observed data points and the compensating contribution over the observation window  $W$ .

Matérn Cluster Process (MCP) is a special case of the Neyman–Scott family of cluster processes, where offspring points are uniformly distributed within a fixed radius around their parent points [14-15]. In this model, parent points are generated from a homogeneous Poisson process with intensity  $\kappa$ , and each parent independently produces a Poisson-distributed number of offspring with mean  $\mu$ . The offspring points are then placed uniformly within a circular region of radius  $R$  centered at their parent location. This uniform displacement mechanism implies that all locations within the cluster radius are equally likely to contain offspring, resulting in clusters with homogeneous intensity within their boundaries. The probability density function governing the offspring distribution is expressed as below.

$$f(u) = \begin{cases} \frac{1}{\pi R^2} & \text{if } \|u\| \leq R, \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

*otherwise*

Here,  $R$  defines the spatial extent of the cluster, and  $\|u\|$  represents the Euclidean distance between an offspring and its parent. The uniformity of  $(u)$  ensures that the offspring are evenly scattered within the disk, creating clusters with sharp boundaries compared to the smoother Gaussian spread in the Thomas process [19].

The pair correlation function of the Matérn cluster process declines from a finite value at short distances to one as  $r$  increases, reflecting spatial aggregation limited to the neighborhood defined by  $R$ . The corresponding K-function takes the form (9).

$$K(r) = \pi r^2 + \frac{1}{\kappa} k(r, R) \tag{12}$$

Where  $(r,)$  represents the clustering contribution within radius  $R$ . Model estimation for inhomogeneous Matérn cluster processes typically relies on the first-order composite likelihood, treating observed points as arising from an inhomogeneous Poisson process with intensity  $\lambda(x; \beta, \theta)$  while ignoring higher-order interactions. Parameter estimation is then achieved by maximizing this likelihood with respect to both the parent process parameters  $\beta$  and cluster parameters  $\theta$ . The Matérn process is particularly suitable for modeling spatial phenomena where clusters exhibit uniform density within a finite spatial extent, such as vegetation patches, animal nests, or fire hotspots confined within localized regions.

## 2.4 Model Evaluation

Model selection plays a crucial role in spatial analysis, especially in point process modelling, to ensure that the selected model can accurately represent the spatial distribution of points. The Akaike Information Criterion (AIC) is commonly applied to achieve a balance between model fit and model complexity to minimize the risk of overfitting [19,20].

The AIC is defined by (10).  $AIC = -2L_{max} + 2p$  (13) Where  $p$  is the number of estimated parameters and  $L_{max}$  denotes the maximum value of the likelihood function for the fitted model. In the case of a Poisson process model,  $L_{max}$  is obtained from the first-order composite likelihood, while for a Cox process model, it is derived from the second-order composite likelihood. The model with the lowest AIC value is considered the most suitable, as it indicates minimal information loss and provides a more accurate and efficient spatial representation.

## 3 Methodology

### 3.1 Study Area and Data Descriptions

This study focuses on the island of Kalimantan, Indonesia, which consists of five provinces: West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. Kalimantan is selected due to its extensive peatland coverage and its recurrent forest and land fire events, particularly during prolonged dry seasons influenced by El Niño phenomena. The primary dataset consists of forest fire hotspot locations detected during the dry season of 2024. Only hotspots with high confidence levels were retained to reduce false detections and improve spatial reliability. Each hotspot is represented as a spatial point in a two-dimensional geographical coordinate system (longitude, latitude). In addition to the point pattern data, spatial covariates were incorporated to model intensity inhomogeneity. These covariates include:

1. Distance to the nearest **Regional Disaster Management Agency (BPBD)** office,
2. Distance to the nearest **water source (river)**.

Distances were computed using Euclidean distance metrics based on spatial layers obtained from Google Maps. The response variable is the spatial location of fire hotspots, while covariates are treated as continuous spatial predictors.

**Table 1.** Research Variables

| Variable   | Description                      | Label/Scale                             | Source of Data       | Year |
|------------|----------------------------------|---|----------------------|------|
| X          | Forest fire locations (hotspots) | Degree coordinate (latitude, longitude) | Ministry of Forestry | 2024 |
| Covariates | Distance to water sources        | Ratio                                   | Google Maps          | 2024 |
|            | Distance to BPBD                 | Ratio                                   | Google Maps          | 2024 |

### 3.2 Data Structure

In our purpose method,  $X$  represents forest fire point data, which is a coordinate point consisting of latitude and longitude. Then,  $z$  is the covariate, namely the distance from the point to the nearest river and the distance from the point to the nearest BPBD. Let

$$X = \{x_i = (x_{i1}, x_{i2}): i = 1, \dots, n\} \tag{13}$$

denote the observed spatial point pattern of forest fire hotspots within an observation window  $W \subset R^2$ , where  $x_{i1}$  and  $x_{i2}$  represent longitude and latitude, respectively. Each point  $x_i$  is associated with a vector of spatial covariates:

$$z(x_i) = z_1(x_i), z_2(x_i)). \tag{14}$$

where:

1.  $z_1(x_i)$  denotes the distance to the nearest river,
2.  $z_2(x_i)$  denotes the distance to the nearest BPBD office.

The following is the data structure used in this study.

**Table 2.** Data Structure

| <b>i</b> | <b>Longitude (<math>X_{long,i}</math>)</b> | <b>Latitude (<math>X_{lat,i}</math>)</b> | <b><math>Z_1(x_i)</math></b> | <b><math>Z_2(x_i)</math></b> |
|----------|--|--|------------------------------|------------------------------|
| 1        | $X_{long,1}$                               | $lat,1$                                  | $Z_{1,1}$                    | $Z_{2,1}$                    |
| 2        | $X_{long,2}$                               | $lat,2$                                  | $Z_{1,2}$                    | $Z_{2,2}$                    |
| ..       | ....                                       | ....                                     | ...                          | ....                         |
| n        | $X_{long,n}$                               | $lat,n$                                  | $Z_{1,n}$                    | $Z_{2,n}$                    |

### 3.3 . Exploratory Spatial Data Analysis

Prior to model fitting, exploratory spatial data analysis (ESDA) was conducted to assess the global structure and potential spatial dependence of hotspot locations. Quadrat counting analysis was employed to test the hypothesis of Complete Spatial Randomness (CSR). The observation window was partitioned into several subregions, and a chi-square test was applied to compare observed and expected counts under a homogeneous Poisson process.

To further examine spatial dependence, Ripley’s K-function and the inhomogeneous K-function were computed. The homogeneous K-function was used as an initial diagnostic to identify clustering tendencies, while the inhomogeneous K-function was employed to account for spatial intensity variation induced by covariates. This step is crucial to distinguish true second-order interaction (clustering) from first-order intensity inhomogeneity. The estimated intensity function  $\lambda$  was evaluated on a regular spatial grid covering the study area to produce a continuous hotspot risk map. Regions with higher estimated intensity indicate areas with greater susceptibility to forest and land fires, providing spatially explicit information for mitigation planning and early-warning strategies.

### 3.4 Analysis Steps

The analysis begins with the acquisition and preprocessing of forest fire hotspot data and spatial covariates. Only hotspots with high confidence levels were retained to ensure the reliability of detected fire events. Spatial covariates, including the distance to the nearest BPBD office and distance to water sources, were computed using geographic information system (GIS) tools based on Euclidean distance. All spatial layers were projected into a common coordinate reference system, and the study region was defined as a bounded observation window  $W \subset R^2$ . Data cleaning procedures were applied to remove duplicated points and ensure spatial consistency.

Exploratory spatial analysis was conducted to assess the global structure and heterogeneity of the hotspot distribution. Initial visualization through spatial point maps revealed substantial variation in point density across Kalimantan. To formally evaluate the assumption of Complete Spatial Randomness (CSR), a quadrat counting test was applied by partitioning the study area into several subregions and comparing observed hotspot counts with those expected under a homogeneous Poisson process. The rejection of CSR indicated strong spatial inhomogeneity, suggesting that fire occurrence is influenced by underlying spatial factors.

To investigate spatial dependence beyond first-order effects, second-order statistics were employed. Ripley’s K-function was first used to detect clustering patterns in the hotspot distribution. Because the intensity of fire occurrences varies spatially, an inhomogeneous K-function was subsequently applied to account for non-stationary intensity. The

persistence of clustering after intensity correction indicated the presence of genuine spatial interaction among hotspot events, thereby justifying the use of cluster process models.

Based on the exploratory findings, the hotspot pattern was modeled using an Inhomogeneous Neyman–Scott Cluster Process (NSCP), where cluster centers follow an inhomogeneous Poisson process with intensity expressed as a log-linear function of spatial covariates. Two alternative offspring dispersion mechanisms were considered: the Thomas cluster process, which assumes Gaussian dispersion of events around parent locations, and the Matérn cluster process, which assumes uniform dispersion within a fixed radius. These two models represent different spatial fire spread behaviors and clustering scales.

Model parameters were estimated using the first-order composite likelihood approach, treating the observed process as an inhomogeneous Poisson process while incorporating clustering effects through spatial convolution. The likelihood integral over the observation window was approximated using the Berman–Turner device to ensure computational efficiency. Backward variable elimination was applied to identify the most parsimonious covariate structure for each cluster model.

Model selection was performed using the Akaike Information Criterion (AIC) to compare competing specifications. The model with the lowest AIC value was selected as the optimal representation of the spatial distribution of forest fire hotspots. The estimated regression coefficients and cluster parameters were then interpreted to assess the influence of spatial covariates on hotspot intensity and to characterize the strength and spatial extent of clustering.

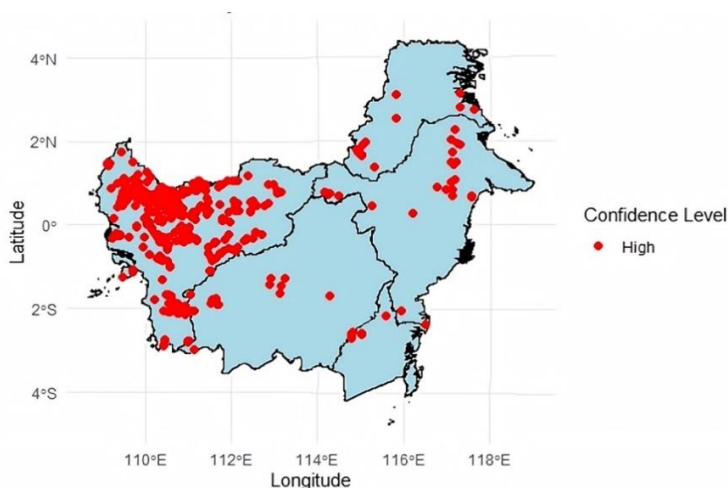
Finally, the selected NSCP model was used to generate spatial predictions by evaluating the estimated intensity function over a regular grid covering Kalimantan. The resulting intensity surface was visualized as a forest fire risk map, where higher intensity values indicate areas with greater susceptibility to fire occurrence. This risk map provides spatially explicit information to support targeted fire prevention, monitoring, and mitigation strategies.

## **4 Results and Discussion**

### **4.1 Exploratory Spatial Pattern of Forest Fire Hotspot**

The initial analysis focuses on the spatial distribution of forest fire hotspots across Kalimantan in 2024. Visual inspection of the hotspot map reveals a highly uneven spatial pattern, with clear concentrations in specific regions, particularly in the western and central parts of Kalimantan. These areas exhibit dense aggregations of hotspot points, while other regions show sparse or scattered occurrences. Such strong spatial heterogeneity indicates that forest fire events are unlikely to follow a random spatial process.

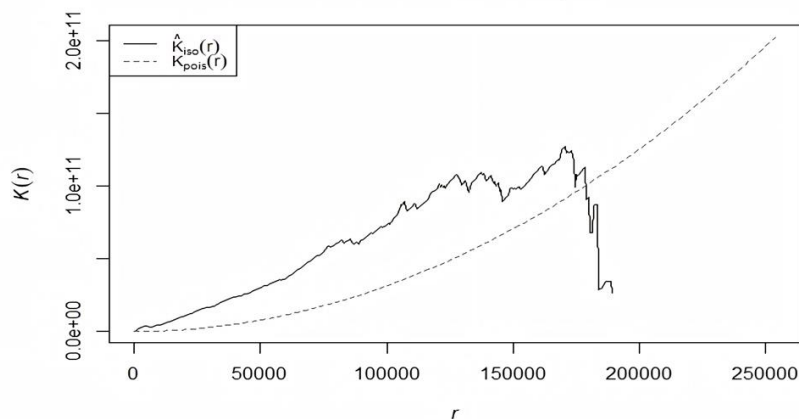
The quadrat counting analysis further supports this observation. The extremely large chi-square statistic and the corresponding p-value far below conventional significance levels strongly reject the hypothesis of Complete Spatial Randomness (CSR). This result confirms that hotspot occurrences are influenced by spatially varying factors, such as environmental conditions and human activities, rather than arising from a homogeneous Poisson process. Consequently, modelling approaches assuming spatial homogeneity would be inappropriate and potentially misleading for this dataset.



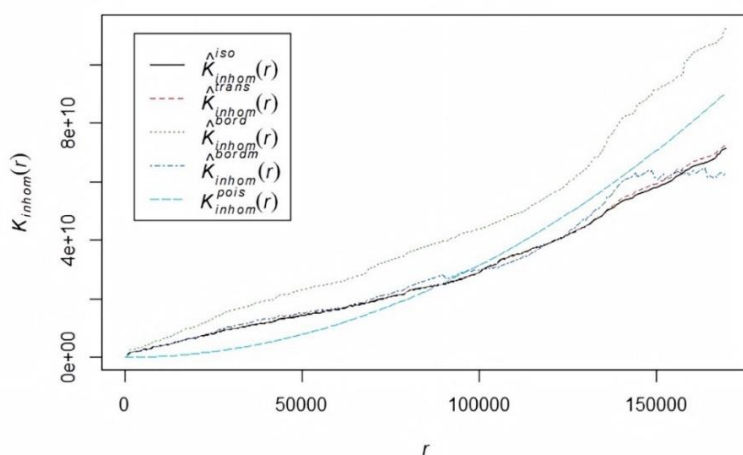
**Fig. 1.** Distribution of forest fire hotspots with confidence in Kalimantan in 2024.

The distribution of hotspots on the map (Fig. 1) is generally somewhat irregular or uneven across the island. This is indicated by extreme variations in point density. The western part of Kalimantan is the area with the most striking and massively concentrated hotspots density, with a main focus in the 110° E until 112° E longitude and 0° until 2° S latitude zones. This area is the most significant hotspots zone, as indicated by its high concentration. In contrast, other regions on the island of Kalimantan show much lower hotspot densities or are scattered sporadically. This distribution pattern clearly shows a tendency toward clustering, indicating that the formation of hotspots is not random but is influenced by certain local factors. Therefore, this visual evidence strongly indicates that the hypothesis of Complete Spatial Randomness (CSR) should most likely be rejected for this hotspot data. However, for further verification, a quadrat counting analysis was conducted to observe spatial trends, and then a K-function plot was used to detect interactions between points. To perform quadrat testing under the null hypothesis of Complete Spatial Randomness (CSR), the observation window was divided into nine irregular quadrats to assess whether the number of hotspot points within each sub-region differs significantly from what would be expected under a homogeneous Poisson process. The test result yielded a chi-square statistic of 1174,7 with 8 degrees of freedom and a p value  $< 2,2 \times 10^{-16}$ , indicating a strong rejection of the CSR hypothesis. This suggests that the hotspot pattern is not randomly distributed but exhibits clear spatial variation across the study area. Such a pattern implies the presence of spatial inhomogeneity, possibly due to geological and spatial factors.

Then, we will use the K-function to see the spatial correlation. Fig. 2 presents the result of the K-function analysis for hotspot locations. The empirical K-function lies above the theoretical Poisson reference line, indicating a clear clustering pattern among the hotspots. The sudden drop in the empirical K-function near the right end of is characteristic of edge effects or a lack of data coverage at those large distances,  $r$ . This suggests that the hotspots are not randomly distributed but tend to occur in groups, possibly influenced by environmental or geological factors. Next, we consider the inhomogeneous K-function (Fig. 3) to assess whether the clustering persists after accounting for spatial variation in intensity. The dashed blue line represents the K-function of an inhomogeneous Poisson process. In general, the edge corrected inhomogeneous K-function still indicates the presence of clustering, although the degree of clustering is weaker than in the homogeneous case. This implies that some clustering can be explained by spatial inhomogeneity, but a significant level of spatial dependence remains. Hence, assuming a stationary process may overlook important spatial trends affecting the hotspot distribution.



**Fig. 2.** The K-function plot for distribution of forest fire hotspots in Kalimantan in 2024.



**Fig. 3.** Inhomogeneous Kfunction plot for distribution of forest fire hotspots in Kalimantan in 2024.

#### 4.2 Model Selection

The AIC is a model selection metric used to compare competing statistical models by balancing goodness of fit and model complexity. Lower AIC values indicate a model that provides a better trade-off between fit and parsimony. In the context of spatial point process modeling, AIC helps identify which cluster process (Thomas or Matérn) and which spatial covariate best describe the observed pattern of events. Table 3 presents AIC values and spatial parameters for the Thomas and Matérn cluster process models, each fitted with a different spatial covariate below.

**Table 3.** Thomas and Matérn cluster process models parameter architecture

|          | Thomas                          |                         |                         | Matérn                          |                         |                         |
|----------|---------------------------------|-------------------------|-------------------------|---------------------------------|-------------------------|-------------------------|
|          | Distance between BPBD and river | Distance to BPBD        | Distance to river       | Distance between BPBD and river | Distance to BPBD        | Distance to river       |
| $K$      | $2,766 \cdot 10^{-10}$          | $2,628 \times 10^{-10}$ | $2,652 \times 10^{-10}$ | $2,652 \times 10^{-10}$         | $2,523 \times 10^{-10}$ | $2,604 \times 10^{-10}$ |
| $\omega$ | $9,621 \times 10^3$             | $9,88 \times 10^3$      | $1,993 \times 10^4$     | $2,041 \times 10^4$             | $2,041 \times 10^4$     | $2,011 \times 10^4$     |
| AIC      | -451714,3                       | -456944,7               | -452044,5               | -451665,9                       | -456901,1               | -451997,9               |

Based on AIC comparison, the Thomas process incorporating the Distance to BPBD covariate yields the most parsimonious fit to the observed spatial pattern (AIC = -456944,7). This indicates that the spatial distribution of events is best explained by a Gaussian-type cluster process in which intensity declines with increasing distance from BPBD sites. The Thomas model's consistently lower AIC values relative to the Matérn process further suggest that event clustering follows a smoother, more continuous decay pattern rather than a sharply bounded cluster structure.

In the Thomas model, the estimated  $\kappa$  parameter ranges from  $2,6 \times 10^{-10}$  to  $2,8 \times 10^{-10}$ , while the scale (weight) parameter varies between  $9,6 \times 10^3$  and  $9,9 \times 10^3$ . The relatively small  $\kappa$  value indicates a high degree of spatial clustering, suggesting that the parent process operates with low intensity. Meanwhile, the smaller scale value implies that the aftershock points tend to be concentrated in closer proximity to the main event centers. In comparison, the Matérn model yields slightly smaller  $\kappa$  estimates (approximately  $2,5 \times 10^{-10}$  to  $2,7 \times 10^{-10}$ ) but considerably larger scale values (around  $1,99 \times 10^4$  to  $2,04 \times 10^4$ ). This suggests that the Matérn process represents a more dispersed clustering structure, where events are distributed over a wider spatial range surrounding the mainshock centers.

Overall, the Thomas model exhibits a stronger clustering pattern with a lower parent process intensity compared to the Matérn model. Based on the spatial parameter estimates and the corresponding AIC values, it can be concluded that the Thomas model incorporating the Distance to BPBD covariate provides the most appropriate representation of the spatial structure observed in the data.

### 4.3 Model Interpretation

An in-depth analysis of the calculated parameters was conducted based on the model selection results, which showed that the Thomas Cluster Process Model with the BPBD Distance variable provided the most economical fit to the spatial pattern of hotspots. The Thomas model was able to capture the clustering effect and the intensity trend of the cluster center ( $\lambda p$ ) based on spatial covariates. The parameter that controls this intensity is  $\beta$ , which is modeled in the loglinear function  $\ln(\lambda p(\mathbf{u})) = \beta_0 + \beta_1(u)$ , where  $\beta_1$  is for BPBD Distance.

**Table 4.** Estimated Parameters

|                      | $\beta$                 | $\exp(\beta)$         |
|----------------------|-------------------------|-----------------------|
| <b>Intercept</b>     | -20,572                 | $1,16 \times 10^{-9}$ |
| <b>Z<sub>1</sub></b> | $-2,199 \times 10^{-6}$ | 0,999998              |

The spatial model estimation results show that the estimated value of the intercept parameter ( $\beta_0$ ) is -20,572 and  $\beta_1$  is  $-2,119 \times 10^{-6}$ . These estimates were obtained using the maximum composite likelihood approach, which aims to describe the effect of distance on the spatial distribution of fire hotspot occurrences in the study area.

The negative value of  $\beta_1$  indicates that the closer a point is to the BPBD office, the higher the tendency for a hotspot to occur. Conversely, as the distance from the BPBD increases, the likelihood of hotspot occurrence decreases. In other words, hotspots are more likely to be found in areas relatively close to BPBD locations. Based on the  $\beta_1$  value, it can be calculated that for every one kilometer decrease in distance toward the BPBD, the intensity of hotspot occurrence increases by approximately  $\exp(\beta_1) \approx 0,999998$ , meaning that the change in risk is relatively small per unit of distance. However, this effect becomes more evident over larger distances. For example, in an area located 100 km closer to the BPBD, the probability of hotspot occurrence increases by approximately 1,00021 times compared to more distant regions. This finding suggests that BPBD offices, which are typically located in areas with high human activity and accessibility, may be associated with a greater frequency of hotspots due to anthropogenic factors. The proximity to human settlements and infrastructure likely contributes to a higher fire risk or hotspot concentration.

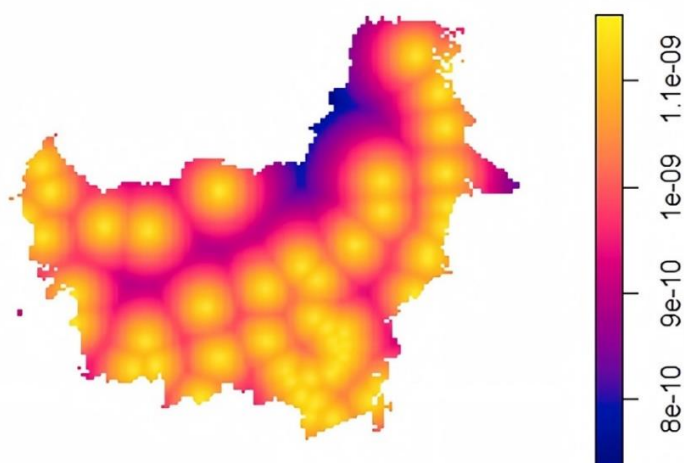
Meanwhile, the intercept ( $\beta_0$ ) represents the baseline log intensity under the condition where the distance to the BPBD is zero (i.e., exactly at the BPBD location). Based on this value, the baseline intensity of hotspots ( $\lambda$ ) can be calculated as.

$$\lambda = e(\beta_0) \times \kappa = (1,16 \times 10^{-19}) \times (2,628 \times 10^{-10}) = 3,06 \times 10^{-19} \quad (14)$$

The value  $\lambda = 3,06 \times 10^{-19}$  indicates that when the distance to the BPBD is zero, the intensity of hotspot occurrence is extremely low. This implies that even though the location is very close to BPBD, the probability of hotspot occurrence remains minimal. This condition may be attributed to stronger monitoring and fire-control measures implemented around BPBD offices. These results confirm that the distance to BPBD has a negative influence on the spatial intensity of hotspots, but the effect is relatively mild and tends to become more apparent over larger spatial scales. Overall, these findings support the idea that the distribution of hotspots in Kalimantan is determined by a clustering process that is strongly influenced by the relative distance to BPBD facilities.

#### 4.4 Model Prediction

The modeling results using the Non-Stationary Cluster Process (NSCP) with the Thomas cluster model approach and covariates of river proximity and BPBD location produced a predicted map of hotspot intensity distribution across Kalimantan, as shown in Fig. 4. This map presents an estimation of hotspot density (intensity) based on the developed spatial model, where lighter colors indicate areas with higher hotspot occurrence intensity.



**Fig. 4.** Predicted hotspot intensity using the Thomas NSCP model in Kalimantan.

The regions with the highest intensity are identified in the central and eastern parts of Kalimantan, along with several smaller concentrations in the southwestern and southern areas. This distribution pattern indicates that forest and land fire activities tend to cluster in specific areas, most likely influenced by proximity to triggering variables such as river distance and BPBD location. The predicted intensity map remains indicative and potentially biased. This condition is attributed to the limited number and uneven distribution of hotspot data used, as well as possible spatial imbalances in the covariate variables. The interpretation of this map should therefore be conducted with caution, and further validation is recommended using additional observational data or alternative modeling approaches.

## 5 Conclusion and Future Works

Based on the model comparison using Akaike Information Criterion (AIC), the Thomas Cluster Process model incorporating the “Distance to BPBD” covariate provides the best and most parsimonious fit to the observed spatial pattern of hotspot occurrences. With the lowest AIC value (AIC = -456944,7), this model effectively captures the

spatial clustering behavior of events, indicating that hotspot intensity decreases smoothly with increasing distance from BPBD facilities. The Thomas process demonstrates stronger spatial clustering ( $\kappa = 2,6 \times 10^{-10}$ ) and smaller cluster spread ( $\omega = 9,9 \times 10^3$ ) compared to the Matérn process, which exhibits a more dispersed clustering structure with larger scale values ( $\approx 2.0 \times 10^4$ ). This implies that hotspot events tend to concentrate more tightly around central points under the Thomas model, while the Matérn process represents a wider, less cohesive distribution.

From the log-linear intensity model  $\ln(\lambda \square(u)) = \beta_0 + \beta_1(u)$ , the intercept ( $\beta_0 = -20,572$ ) and slope ( $\beta_1 = -2,119 \times 10^{-6}$ ) reveal that hotspot intensity is inversely related to the distance from BPBD locations. In other words, areas closer to BPBD facilities are more likely to become cluster centers for fire events, while regions farther away show a significant reduction in clustering intensity.

In summary, the analysis concludes that the Thomas Cluster Process with the BPBD Distance covariate is the most suitable model to describe the spatial structure of hotspot occurrences in Kalimantan. The model reflects a Gaussian type clustering mechanism influenced strongly by proximity to BPBD facilities, suggesting that the spatial distribution of fire hotspots is not random but rather governed by localized clustering dynamics and accessibility to disaster management infrastructure.

Some suggestions for further research are to add several spatial covariates directly related to the location of the fire, such as land cover and soil type, or add some temporal covariates such as rainfall and wind direction throughout the year. In addition, because the data used is very large, analysis methods and tools that are robust and tested on high-volume data can be used.

## References

- [1] Madani Berkelanjutan, "40 Ribu Hektare Karhutla Diindikasikan Terjadi di Indonesia Selama Januari-April 2024, Terbesar di Kalimantan Timur," [madaniberkelanjutan.id](https://madaniberkelanjutan.id). Available: <https://madaniberkelanjutan.id/40-ribu-hektare-karhutla-diindikasikan-terjadi-di-indonesia-selama-januari-april-2024-terbesar-dikalimantan-timur/>. [Accessed: Feb. 10, 2025].
- [2] L. Syaufina and I. S. Sitanggang, "Peatland fire detection using spatio-temporal data mining analysis in Kalimantan, Indonesia," *J. Trop. For. Sci.*, vol. 30, no. 2, pp. 154–162, 2018. doi: 10.26525/jtfs2018.30.2.154162.
- [3] A. J. Horton et al., "Identifying key drivers of peatland fires across Kalimantan's Ex-Mega Rice Project using machine learning," *Earth and Space Sci.*, vol. 8, no. 9, Art. no. e2021EA001873, 2021. doi: 10.1029/2021EA001873.
- [4] A. Schmidt et al., "Fire frequency, intensity, and burn severity in Kalimantan's threatened peatland areas over two decades," *Front. Forests Global Change*, vol. 7, Art. no. 1221797, 2024. doi: 10.3389/ffgc.2024.1221797.
- [5] A. Choiruddin, Aisah, F. Trisnisa, and N. Iriawan, "Quantifying the effect of geological factors on distribution of earthquake occurrences by inhomogeneous Cox processes," *Pure Appl. Geophys.*, vol. 178, pp. 1579–1592, 2021.
- [6] R. Kumalawati et al., "Hotspot spatial patterns using SNNP-VIIRS for fire potential monitoring," *Int. J. Forestry Res.*, vol. 2023, Art. no. 3121862, 2023.
- [7] S. Nurdianti, A. Sopaheluwakan, P. Septiawan, and M. R. Ardhana, "Joint spatio-temporal analysis of various wildfire and drought indicators in Indonesia," *Atmosphere*, vol. 13, no. 10, Art. no. 1591, 2022.
- [8] F. Bioresita, D. E. Pongdatu, N. Hayati, U. W. Deviantari, and C. B. Pribadi, "Estimation of forest fire areas in Palangka Raya, Central Kalimantan, Indonesia using NBR2 and its impact on environment," *Indonesian J. Forestry Res.*, vol. 12, no. 1, pp. 1–12, 2025. doi: 10.59465/ijfr.2025.12.1.1-12.

- [9] Y. Wang, A. Degleris, A. Williams, and S. W. Linderman, “Spatiotemporal clustering with Neyman-Scott processes via connections to Bayesian nonparametric mixture models,” *J. Amer. Statist. Assoc.*, vol. 119, no. 547, pp. 2382–2395, 2024. doi: 10.1080/01621459.2023.2257896.
- [10] HA Sidharta, B Al Kindhi, E Mulyanto, MH Purnomo, “Semantic Segmentation of Pedestrian Groups Based on Directional-oriented Density Features with Shallow U-net Architecture”, *International Journal of Intelligent Engineering & Systems*, 18, 1, 2025
- [11] A. J. Baddeley, J. Møller, and R. Waagepetersen, “Non- and semiparametric estimation of interaction in inhomogeneous point patterns,” *Stat. Neerl.*, vol. 54, no. 3, pp. 329–350, 2000.
- [12] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester, U.K.: Wiley, 2008.
- [13] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R*. New York, NY, USA: Chapman and Hall/CRC, 2015.
- [14] S. Portet, “A primer on model selection using the Akaike Information Criterion,” *Infect. Disease Modelling*, vol. 5, pp. 111–128, 2020. doi: 10.1016/j.idm.2019.12.010.
- [15] J. Zhang, Y. Yang, and J. Ding, “Information criteria for model selection,” *WIREs Comput. Stat.*, vol. 15, no. 5, Art. no. e1607, 2023. doi: 10.1002/wics.1607.