



A SPATIAL NONHOMOGENEOUS POISSON PROCESS MODEL USING BAYESIAN APPROACH ON A SPACE-TIME GEOSTATISTICAL DATA

Anggun Yuliarum Qur'ani¹ and Subanar²

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. Email: anggun.yuliarum.q@mail.ugm.ac.id

²Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada.

Cite this article:

Anggun Y.Q., Subanar (2021), A Spatial Nonhomogeneous Poisson Process Model Using Bayesian Approach on a Space-Time Geostatistical Data. African Journal of Mathematics and Statistics Studies 4(3), 186-198. DOI: 10.52589/AJMSS-C4L7KHUC.

Manuscript History

Received: 19 Nov 2021

Accepted: 10 Dec 2021

Published: 27 Dec 2021

Copyright © 2020 The Author(s).

This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *In this research, we propose the nonhomogeneous Poisson process on geostatistical data by adding a time component to be applied in the study case of air pollution in the Special Region of Yogyakarta. We use the Bayesian approach to inference the model using the MCMC method. And to generate samples of the posterior distribution, we wield the Metropolis-Hastings algorithm, and we obtained it has good convergence for this case. And to show the goodness of fit of this model, we had the value of DIC.*

KEYWORDS: Spatial Nonhomogeneous Poisson Process (SNHPP), Space-time Geostatistical Data, MCMC, Gibbs Sampling Algorithm with Metropolis-Hasting Steps, Air Pollution.



INTRODUCTION

Poisson process both homogeneous and nonhomogeneous have an important and fundamental role both in theory and applications to applied in many study areas, e.g. growth of emission factors, especially in the air pollution case [13], [14]. They were applied nonhomogeneous Poisson process into this case. To observe air pollution, we use fixed locations in its region. By using spatial data that classified by [5] into 3 types, i.e. point, lattice, and geostatistical data, we find suitable types of spatial data in the air pollution observation locations is geostatistical data. We called its model here by spatial nonhomogeneous poisson model (SNHPP). The applications of nonhomogeneous Poisson process in geostatistical data is performed by [12]. We called its model here by spatial nonhomogeneous poisson model (SNHPP). By adding time component into its data, we obtain data that is called a space-time geostatistical data.

The SNHPP model inference, we use parametric approach more precisely using Bayesian approach. In Bayesian approach, we use Markov Chain Monte Carlo (MCMC) method to estimate of its model parameters. We focus to generating samples of posterior distribution using Metropolis-Hastings algorithm. Then generating samples, we also should to check its convergence, here we use Gelman-Rubin diagnostic, one of the convergence checking technique of MCMC samples [6]. And to get the best SNHPP model, we used DIC value as proposed by [10] in Bayesian case.

The purpose of this research is to modeling SNHPP in the study case of air pollution in 12 fixed observed locations in Special Region of Yogyakarta that indicated by NO_2 and SO_2 concentrations in the air using active method in 2012 to 2015 that measured twice in a year.

Spatial Autocorrelation Assumption Testing

The spatial autocorrelation means to measure existence of value dependence level of variable that close to each other in the geographic space. The existence of spatial autocorrelation can be seen by Moran's I value. Mathematically, statistics test of Global Moran's I can be written in the following equation.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

where n is the number of observation unit, x_i is the value of each predictor variables for i -th unit, and w_{ij} is the element of spatial weighting matrix power distance weight between i -th and j -th locations. Testing the parameter I can be done as follows.

$H_0 : I = 0$ (there is no spatial autocorrelation)

And there are two alternative hypothesis

$H_1 : I > 0$ (it has positive autocorrelation)

$H_1 : I < 0$ (it has negative autocorrelation)



Testing the hypothesis using statistic test as follows.

$$Z_I = \frac{I - E(I)}{\sqrt{V(I)}} \quad (2)$$

Hypothetical decision making based on the condition of H_0 is rejected if the value of $Z_I > Z_{\alpha/2}$ that means there is correlation between locations [8], [11].

Spatial Nonhomogeneous Poisson Process (SNHPP)

SNHPP is nonhomogeneous Poisson process that used to get geostatistical model using space-time approach with random intensity process of data that have two additional components, those are Gaussian spatial and temporal effect component. [12]. Let geographical space that be observation space, $D \subset R^p$ (with $p = 2$), there are n fixed points, $s_1, s_2, \dots, s_n \subset D$, with observation time at the interval $(0, T]$. So, we obtain t_{ij} , the i -th time observation at the location s_j , that is $0 < t_{1j} < t_{2j} < \dots < t_{ij} < \dots < t_{n,j} < T$, with $i = 1, 2, \dots, t; j = 1, 2, \dots, n$. SNHPP model [Morales, 2016] can be formulated as follows.

$$\log \Lambda(s, t) = W(s | \theta_W) + H(t | \theta_H) \quad (3)$$

From (3) can be obtained the following equation.

$$\lambda(s, t) = \exp(W(s | \theta_W) + H(t | \theta_H)) \quad (4)$$

where $\lambda(s, t)$ is intensity function of Poisson process, and $W(\cdot | \cdot)$ is non-stationary and isotropic Gaussian process as follows.

$$W(s | \theta_W) \sim GP(\mu(s), \sigma^2, \rho_\phi) \quad (5)$$

where $\mu(s) = x(s)\Psi$ is the process mean, $x(s)$ is the covariate vector at the location s based on the coefficient vector Ψ . σ^2 is the process variance, and ρ_ϕ is the function of the valid correlation i.e. the correlation between $W(s_i, \theta_W)$ and $W(s_j, \theta_W)$ for each $s_i, s_j \in D$ related to the isotropic process of $|s_i, s_j|$, the distance of Euclidean between s_i and s_j , where the correlation $\rho_\phi(s_i, s_j) = \exp(-\phi |s_i, s_j|)$. The cumulative hazard function $H(t)$ is the logarithm of hazard function $h(t)$ as follows.



$$H(t) = \log(\beta \alpha t^{\alpha-1} e^{-\beta t^\alpha}) \quad (6)$$

So, for each s_j , we have [7], [12] the general SNHPP with the intensity function as follows.

$$\lambda(s_j, t) = \theta_j \beta \alpha t^{\alpha-1} e^{-\beta t^\alpha} \quad (7)$$

and the function of mean value at the location s time t can be formulated in the following equation.

$$m(s_j, t) = \theta_j (1 - e^{-\beta t^\alpha}) \quad (8)$$

with $\theta_j = e^{W_j}$, $j = 1, 2, \dots, n$, the parameter θ_j is interpreted as the number of events that observed in the time interval $(0, T]$, β is the detected level of events occurred.

Prior and Posterior Distribution of Snhpp Model

Prior distribution that used here is a non-informative prior distribution refers to [12] that consists of parameters $N'_j, \beta, \alpha, \Psi, \sigma^2, \phi$ as follows.

$$N'_j \sim \text{Poisson}\left(e^{W_j - \beta T^\alpha}\right)$$

$$\beta \sim \text{Gamma}(a_\beta, b_\beta), \quad a_\beta, b_\beta \text{ are known}$$

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha), \quad a_\alpha, b_\alpha \text{ are known}$$

$$\Psi \sim N(\mathbf{m}, \mathbf{v}), \quad \mathbf{m}, \mathbf{v} \text{ are known}$$

$$\sigma^2 \sim \text{Gamma}(a_{\sigma^2}, b_{\sigma^2}), \quad a_{\sigma^2}, b_{\sigma^2} \text{ are known}$$

$$\phi \sim \text{Gamma}(a_\phi, \eta), \quad \text{with } a_\phi = \frac{-2 \log(0.05)}{\max(|s_i - s_j|)}$$

Here, we use the latent variable, $N'_j = N_j - n_j$, that used to facilitate sampling the joint posterior distribution [1], [9], [12]. Using the definition of Gaussian process, we get that $\mathbf{W} \sim N(X\Psi, \Sigma)$, where X is the observed covariate matrix, here we denote that the rows are $\mathbf{x}_j = (1, x_j, y_j)$, where $s_j = (x_j, y_j)$ are the geographic coordinate of monitoring stations and Σ is



the covariance matrix with the element are known, i.e. $\Sigma_{ij} = \sigma^2 \rho_\phi(s_i, s_j)$, $i, j = 1, 2, \dots, n$. And consider the independence of prior distribution each other, we have the prior distribution Φ as follows.

$$\pi(\Phi) = \pi(\beta)\pi(\alpha)\pi(\Psi)\pi(\mathbf{W}|\phi, \sigma^2)\pi(\phi)\pi(\sigma^2) \quad (9)$$

Let $\mathbf{W} = (W_1, W_2, \dots, W_n)$, $\Phi = (\beta, \alpha, \Psi, \mathbf{W}, \phi, \sigma^2)$ is a set of parameter, and for $j = 1, 2, \dots, n$, a set of data, $D_j = (t_{1j}, t_{2j}, \dots, t_{n_j j})$, $D = (D_1, D_2, \dots, D_n)$. We have the likelihood function of SNHPP that can be formulated in the following equation.

$$L(\Phi | D) = \left(\beta \alpha^{\sum_{j=1}^n n_j} \right) \left(\prod_{j=1}^n \prod_{i=1}^{n_j} t_{ij}^{\alpha-1} \right) \exp \left(-\beta \sum_{j=1}^n \sum_{i=1}^{n_j} t_{ij}^\alpha + \sum_{j=1}^n n_j W_j - \sum_{j=1}^n e^{W_j} (1 - e^{-\beta T^\alpha}) \right) \quad (10)$$

Based on (9) and (10), for $j = 1, 2, \dots, n$, the latent variable, N_j , $N' = (N_1, N_2, \dots, N_n)$, we have the joint posterior distribution Φ as follows.

$$\pi(N', \Phi | D) \propto L(\Phi | D) \pi(\Phi) \prod_{j=1}^n \pi(N'_j) \quad (11)$$

And we have the conditional posterior density functions for any parameters use the following equation.

$$\pi(\theta | \mathbf{N}, \Theta_{-\theta}) = \frac{\pi(\theta) \pi(\mathbf{N}, \Theta_{-\theta} | \theta)}{\pi(\mathbf{N}, \Theta_{-\theta})} \quad (12)$$

So, by using Box Tiao writing we obtain the following equations.

$$\pi(\beta | \mathbf{N}, \Phi_{-\beta}) \propto \beta^{\sum_{j=1}^n n_j + a_\beta - 1} \exp \left(-\beta \left(\sum_{j=1}^n \sum_{i=1}^{n_j} t_{ij}^\alpha + \sum_{j=1}^n b_\beta + \sum_{j=1}^n N'_j T_j^\alpha \right) \right) \quad (13)$$



$$\pi(\alpha | \mathbf{N}, \Phi_{-\alpha}) \propto \alpha^{\sum_{j=1}^n n_j} \pi(\alpha) \prod_{j=1}^n \prod_{i=1}^{n_j} t_{ij}^{\alpha} \exp \left(-\beta \left(\sum_{j=1}^n \sum_{i=1}^{n_j} t_{ij}^{\alpha} + \sum_{j=1}^n N_j' T_j^{\alpha} \right) \right) \quad (14)$$

$$\pi(\mathbf{W} | \mathbf{N}, \Phi_{-\mathbf{W}}) \propto \exp \left(\sum_{j=1}^n n_j W_j + \sum_{j=1}^n e^{N_j W_j} - \sum_{j=1}^n e^{W_j} - \frac{1}{2} (\mathbf{W} - \mathbf{X}\Psi)^T \Sigma^{-1} (\mathbf{W} - \mathbf{X}\Psi) \right) \quad (15)$$

$$\pi(\Psi | \mathbf{N}, \Phi_{-\Psi}) \propto \exp \left(\frac{1}{2} (\Psi - \mathbf{m})^T \mathbf{v}^{-1} (\Psi - \mathbf{m}) - \frac{1}{2} (\mathbf{W} - \mathbf{X}\Psi)^T \Sigma^{-1} (\mathbf{W} - \mathbf{X}\Psi) \right) \quad (16)$$

$$\pi(\phi | \mathbf{N}, \Phi_{-\phi}) \propto \phi^{a_{\phi}-1} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{W} - \mathbf{X}\Psi)^T \Sigma^{-1} (\mathbf{W} - \mathbf{X}\Psi) - b_{\phi} \phi \right) \quad (17)$$

$$\pi(\sigma^2 | \mathbf{N}, \Phi_{-\sigma^2}) \propto (\sigma^2)^{a_{\sigma^2}-1} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{W} - \mathbf{X}\Psi)^T \Sigma^{-1} (\mathbf{W} - \mathbf{X}\Psi) - b_{\sigma^2} \sigma^2 \right) \quad (18)$$

where $\Theta_{-\theta}$ is a set of parameter Θ except parameter θ (12).

Estimating Parameter and Goodness Of Fit Snhpp Model

To estimate parameter model of SNHPP of (3), we use parametric approach, i.e. Bayesian approach by using the MCMC method. And to generate samples of joint posterior distribution in (11), we use Metropolis-Hastings algorithm, where this algorithm can be used when the conditional posterior distributions are not easy to identified. Here also, we use the latent variable to increasing the accuracy of Metropolis-Hastings algorithm in generating the joint posterior distribution samples.

Let the posterior distribution to be sampled in (13) to (18). The estimated parameter is Φ , where $\Phi = (\beta, \alpha, \mathbf{W}, \Psi, \phi, \sigma^2)$. Here we want renewal i -th elements of Φ , then we choose proposal function q , so that.

$$q(\Phi, \Phi^*) = \pi(\Phi_i^* | \Phi_i) \text{ for } \Phi_i^* = \Phi_i \quad (19)$$

and zero for others.

These are the steps of Metropolis-Hastings algorithm [2], [4], i.e. let $\Phi^t, t = 0, 1, \dots, m$. Firstly, as many t -th iterations, we generate $\Phi_i^* \sim q(\Phi^* | \Phi^t)$. Secondly, choose $\Phi^{(t+1)} = \Phi_i^*$ for the probability equal to $\rho(\Phi^t, \Phi_i^*)$, and $\Phi^{(t+1)} = \Phi_i$ for the probability equal to $(1 - \rho(\Phi^t, \Phi_i^*))$ with $\rho(\Phi, \Phi^*) = \min \left\{ \frac{f(\Phi^*)q(\Phi | \Phi^*)}{f(\Phi)q(\Phi^* | \Phi)}, 1 \right\}$.



After we had generated MCMC samples, we have to check the convergence of MCMC by using Gelman-Rubin diagnostic. There are 7 steps to do Gelman-Rubin diagnostic method [6] as follows.

First, we stimulate sequence $m \geq 2$, each sequence has a length $2n$, then did reduction of first iterations n , and focus on last iterations n .

Second, For each scalar parameter, we calculate $\frac{B}{n} = \sum_{i=1}^m \frac{(\bar{x}_i - \bar{x}_{..})^2}{(m-1)}$ = variance between sequence

mean m , \bar{x}_i = each means against n value of x . Then we calculate $W = \sum_{i=1}^m \frac{s_i^2}{m}$ = variance mean in sequence m , s_i^2 = each variance against degree of freedom $(n-1)$.

Third, estimating the target mean, $\mu = \int xP(x)dx$, where $\hat{\mu} = \bar{x}_{..}$ = sample mean of mn the stimulated value of x .

Fourth, estimating the target variance, $\sigma^2 = \int (x - \mu)^2 P(x)dx$, with the weighted mean W and B , so that $\hat{\sigma}^2 = \frac{n-1}{n}W + \frac{1}{n}B$.

Fifth, we estimate x that can improve estimation value of target distribution. by allowing for the sampling variability of the estimates, $\hat{\mu}$ and $\hat{\sigma}^2$. By using $\hat{\mu}$ as the center, with scale $\sqrt{\hat{V}} = \sqrt{\hat{\sigma}^2 + \frac{B}{mn}}$ and the degree of freedom $df = \frac{2\hat{V}^2}{\text{var}(\hat{V})}$, where

$$\text{var}(\hat{V}) = \left(\frac{n-1}{n}\right)^2 \frac{1}{m} \text{var}(s_i^2) + \left(\frac{m+1}{mn}\right)^2 \frac{2}{m-1} B^2 + 2 \frac{(m+1)(n-1)}{mn^2} \frac{n}{m} \left(\text{cov}(s_i^2, \bar{x}_{i..}) - 2\bar{x}_{..} \overline{\text{cov}}(s_i^2, \bar{x}_{i..}) \right) \quad (20)$$

Sixth, we monitoring the convergence of the simulation estimated by the currently variance ratio, $\sqrt{\hat{R}} = \sqrt{\frac{(\hat{V}/W)df}{df-2}}$, converging to 1 when $n \rightarrow \infty$.

Seventh, according to [3], when \hat{R} is less than 1.1 for all scalar estimations, we decided that convergence has been achieved.

Finally, since we have achieved convergence of MCMC, we should measure goodness of fit the models, and [10] use a goodness of fit criterion which is widely used in Bayesian modeling is Deviance Information Criterion (DIC) that can be formulated as follows.

$$DIC = 2E_{\theta|x}(D) - D(E_{\theta|x}(\theta)) \quad (21)$$



with $D(\cdot)$ is the model deviance, and x is the observed data. And for model that have DIC value smaller than other, it will be good model, then we choose it.

Procedure

We use SNHPP to applied to analysis the semester data of NO_2 and SO_2 concentrations in 12 fixed observed points which is observed in Special Region of Yogyakarta (DIY) in 2012 to 2015. Data obtained from Environmental Agency of Special Region of Yogyakarta which is observed in March and September for each year using active method. It measured using $\mu g / m^3$ unit. The steps of analysis of this study as follows.

First, we use the threshold set by WHO of NO_2 is $40 \mu g / m^3$ and SO_2 is $20 \mu g / m^3$, and we get the number of concentration data of NO_2 and SO_2 that exceeds the threshold by utilizing the counting process. Second, by using the data, we testing the spatial autocorrelation assumption wield Global Moran's I. Third, we determine the fixed geographic coordinates which is be the fixed observed point. Fourth, calculating Euclidean distance between fixed observed points based on the coordinates (u_i, v_i) . Fifth, determining spatial weighted matrix, power distance weights. Sixth, simulating MCMC to generate samples of parameter Φ using Metropolis-Hastings algorithm. Seventh, estimating parameter $\lambda(\cdot)$. Eighth, checking the convergence of MCMC wield Gelman-Rubin diagnostic. Ninth, measuring goodness of fit SNHPP model using DIC. In this analysis, we use software R 3.2.5.

RESULTS AND DISCUSSION

The result of testing spatial autocorrelation assumption using Global Moran's I, i.e. test statistics Z_I , obtained the values of Z_I for NO_2 and SO_2 are 2.267, and 2.314. So, for the confidence interval 95%, $Z_{0.05/2} = 1.96$, we have $Z_I > Z_{0.05/2}$, for each NO_2 and SO_2 concentrations data. It means that there are spatial autocorrelation for each data.

Here, we use the model parameters that proposed by Morales (2016), based on the following prior distribution to applied in this data as follows.

$$\beta \sim \text{Gamma}(0.001, 0.001)$$

$$\alpha \sim \text{Gamma}(0.001, 0.001)$$

$$\Psi \sim N(\mathbf{0}, 1000\mathbf{I})$$

$$\sigma^2 \sim \text{Gamma}(0.1 \times 0.005, 0.005)$$

$$\phi \sim \text{Gamma}\left(4 \times \left(\frac{4}{7}\right), \frac{3}{7}\right)$$

where the model parameters of β, α, Ψ are considered non-informative.



We generate 130,000 samples of the conditional posterior distributions, with burn-in 30,000 by Metropolis-Hastings algorithm. We also do repetition twice for each parameters. And we obtain the parameter estimation Φ as the Table 1.

Table 1. Parameter Estimation Mean Φ

Conc.	Par.	Rep.	Mean	50%	2.5%	97.5%
SO_2	β	1	$2.525 \times 10^{-4} *$	2.27×10^{-4}	1.022×10^{-4}	4.998×10^{-4}
		2	$2.543 \times 10^{-4} *$	2.253×10^{-4}	1.019×10^{-4}	5.017×10^{-4}
	α	1	0.1849	0.1822	0.1379	0.2413
		2	0.1851	0.1820	0.1381	0.2417
	ϕ	1	12.8914	12.89	12.55	13.25
		2	12.8952 *	12.89	12.55	13.26
	σ^2	1	0.1527 *	0.1494	0.1021	0.2157
		2	0.1523	0.1488	0.1020	0.2157
NO_2	β	1	$8.915 \times 10^{-5} *$	7.292×10^{-5}	2.52×10^{-5}	2.11×10^{-4}
		2	$8.858 \times 10^{-5} *$	7.285×10^{-5}	2.531×10^{-5}	2.104×10^{-4}
	α	1	0.1708 *	0.13199	0.03832	0.4479
		2	0.1684 *	0.1298	0.03839	0.4436
	ϕ	1	12.8940 *	12.89	12.54	13.26
		2	12.8966 *	12.90	12.55	13.26
	σ^2	1	0.1525 *	0.1489	0.1020	0.2159
		2	0.1521 *	0.1482	0.1020	0.2160

*: the parameter value which has smallest DIC value.

Then, we earn from checking the convergence of MCMC samples using Metropolis-Hastings diagnostic based on estimation value of \hat{R} for each parameter as follows.

Table 2. Estimation Value of \hat{R}

Cons.	Par.	Est.	U.L. of C.I.
SO_2	$\beta, \alpha, \phi, \sigma^2$	1	1
NO_2			

U.L. of C.I.: Upper Limit of Confidence Interval

Since, the each value of estimation of \hat{R} are less than 1.1, we can conclude that the MCMC sampling has converged. The result of measurement of goodness of fit based on the DIC values for each models in repetition as shown on the Table 3.

**Table 3. DIC's Value Difference of Parameter $\beta, \alpha, \phi, \sigma^2$.**

Par.	DIC Value Diff. (Repetition)	
	Conc. of NO_2	Conc. of SO_2
	R1-R2	R1-R2
β	0	0
α	-0,05664062	
ϕ	0,2617188	
σ^2	-0,1699219	

Based on the DIC value difference, if the plus we have got, it is mean the DIC value of first repetition is greater than second repetition each parameter, and otherwise. If we have the DIC value is zero, it is mean the DIC value of first repetition and second repetition is same. By its values, we choose repetition each parameter for the smallest DIC value, we have got the best parameter in Table 1. By it, we can establish some of the best models shown in Table 4.

Table 4. The best of SNHPP model

Cons.	M	Equations
NO_2	1	$\text{Log } \Lambda_j^{(11)} = \exp(W_j) \beta_1 \alpha_1 t_{ij}^{\alpha_1 - 1} \exp(-\beta_1 t_{ij}^{\alpha_1})$
	2	$\text{Log } \Lambda_j^{(12)} = \exp(W_j) \beta_2 \alpha_1 t_{ij}^{\alpha_1 - 1} \exp(-\beta_2 t_{ij}^{\alpha_1})$
SO_2	1	$\text{Log } \Lambda_j^{(21)} = \exp(W_j) \beta_1 \alpha_1 t_{ij}^{\alpha_1 - 1} \exp(-\beta_1 t_{ij}^{\alpha_1})$
	2	$\text{Log } \Lambda_j^{(22)} = \exp(W_j) \beta_1 \alpha_2 t_{ij}^{\alpha_2 - 1} \exp(-\beta_1 t_{ij}^{\alpha_2})$
	3	$\text{Log } \Lambda_j^{(23)} = \exp(W_j) \beta_2 \alpha_1 t_{ij}^{\alpha_1 - 1} \exp(-\beta_2 t_{ij}^{\alpha_1})$
	4	$\text{Log } \Lambda_j^{(24)} = \exp(W_j) \beta_2 \alpha_2 t_{ij}^{\alpha_2 - 1} \exp(-\beta_2 t_{ij}^{\alpha_2})$

$\text{Log } \Lambda_j^{(11)}$ is the 1-st model of NO_2 concentration.

The $\log \Lambda_j$ value in the NO_2 concentration data is closer to the actual data, but the $\log \Lambda_j$ value of SO_2 concentration is not close to the actual data. This possibility is influenced by the presence of parameter values that do not fit the model that can represent an area. The results of mapping the actual $\log \Lambda_j$ values and PPNHS models for each NO_2 and SO_2 concentration data can be seen in Figure 2 (a) and 2 (b) respectively.

In Figure 2(a) we can see the location with the highest level of concentration of NO_2 exposed in the air both in Sleman Regency. For the second highest position, there are differences in districts that have a concentration of NO_2 . For the third position, the highest concentration of NO_2 is often found in Bantul Regency. And followed by Kulonprogo Regency. At location points that showed a concentration of NO_2 more than the threshold set by WHO than other Regencies. It means that the position of these points can be said to be higher air pollution which is the exposure of NO_2 gas in the air. The cause of the frequent concentration of NO_2 exceeding the threshold can be due to gas released from the engine in the vehicle.

The highest level regency which has a concentration of SO_2 exposed in the air in Figure 2(b) is not obtained the same result between the actual data and the SNHPP model. However, for the regency with the next highest level in the second position are both in Gunung Kidul Regency, then they will continue with the Kulonprogo Regency. At these locations in this regency the concentration of

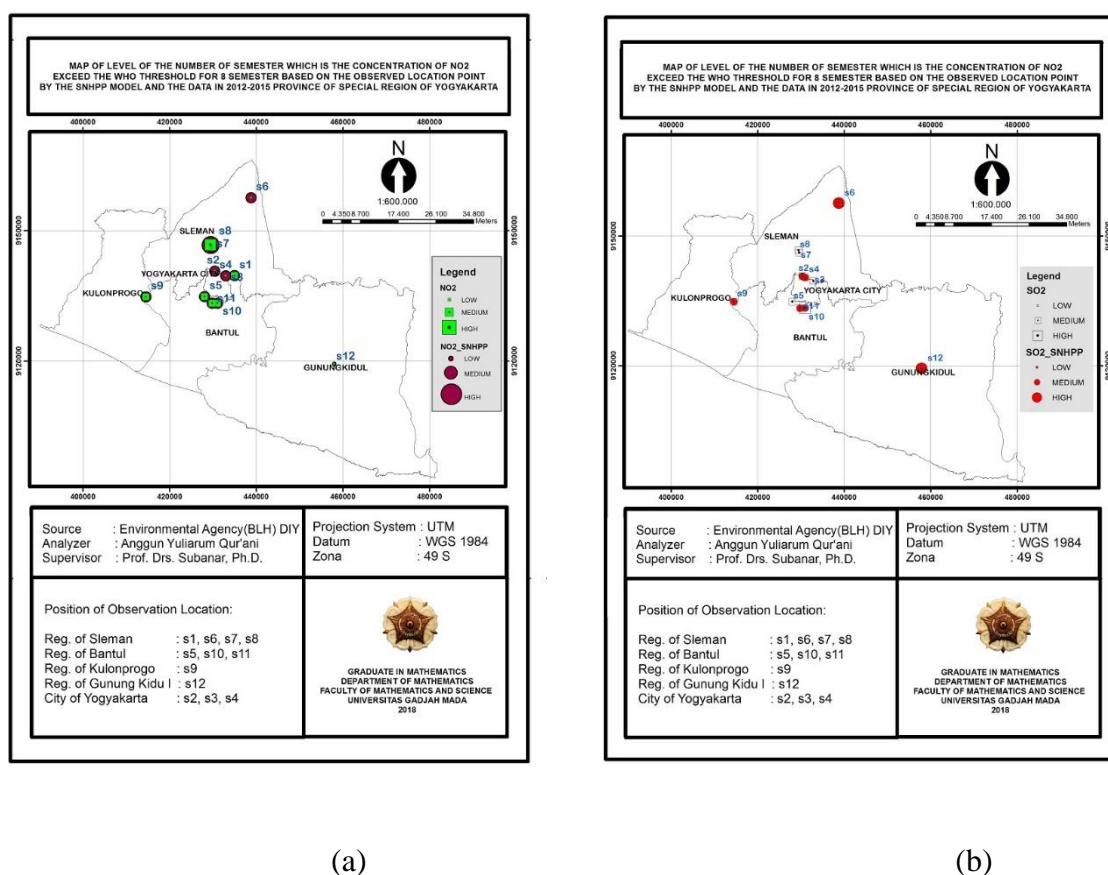


Figure 2: Map of Level of the Number of Semesters which is NO_2 and SO_2 Concentration Value Exceeds WHO Threshold Based on Observation Location Points Based on SNHPP Model and Data in 2012-2015 in the DIY Province

SO_2 often exceeds the threshold compared to other Regencies. The frequent causes of this occurrence are because the combustion process on motorized vehicles. And by paying attention to the observation location points, the observations is carried out at congestion points with the number of motorized vehicles operated.



CONCLUSION

In this research, we propose the nonhomogeneous Poisson process on geostatistical data by adding time component to applied in the study case of air pollution in Special Region of Yogyakarta. From the results that Metropolis-Hastings has a good convergence that is shown by the result of Gelman-Rubin diagnostic. In this case, through the values of DIC, we also obtain the fit model of SNHPP that involves the repetition therein.

The conclusions obtained from the results of the discussion and implementation of the SNHPP model are:

1. The value of the Markov Chain Monte Carlo (MCMC) simulation results with the Gibbs Sampling algorithm with Metropolis-Hastings steps is obtained almost the entire parameter value in the concentration data of NO_2 greater than the parametric value of the concentration data SO_2 .
2. By the results of the parameter values obtained for each data, it obtained
 - a) The best Spatial Non-Homogeneous Poisson Process Model (SNHPP) for concentration data of NO_2 .
 - b) The best Spatial Non-Homogeneous Poisson Process Model (SNHPP) for concentration data SO_2 .
 - c) Regency/cities with the highest level of the number of semesters that have a concentration of NO_2 exceeding the threshold set by WHO are Sleman Regency, Bantul Regency, then followed by Kulonprogo Regency.
 - d) Regency/city with the highest level of the number of semesters that have a concentration value of SO_2 exceeds the threshold set by WHO is Bantul Regency, Gunung Kidul Regency, then followed by Kulonprogo Regency.

Future Research

The results of the research that has been done, still need depth research, for example by adding interpolation for unobserved location points to overcome the limitations of the existing data. In addition, there are still parameter values that cannot represent every region. Supported by research results [Morales, et al., 2016] there are also parameter values that cannot represent each region.

From these shortcomings, it can be used as a theory development regarding Spatial Non-Homogeneous Poisson Process to get parameter values that can be represent each location. Implementation of the SNHPP model that raises the issue of air pollution in the Special Regions of Yogyakarta (DIY) can give you a little picture of the environment especially in DIY, which is starting to be exposed to gases that cause pollution air. We are who live in the DIY area, especially at points that are often exposed to the gas, can increase tree planting on the edges street. In addition, it can reduce the use of motorized vehicles gradually, for example by cycling or walking.



REFERENCES

- Achcar, J.A., Dey, D.K., & Niverthi, M. (1996). A Bayesian Approach Using Nonhomogeneous Poisson Process for Software Reliability Models. CiteSeer, 1-20. DOI: https://doi.org/10.1142/9789812816580_0001
- Albert, J. (2009). Bayesian Computation with R Second Edition. New York: Springer.
- Brook, S., Gelman, A., Jones, G.L., & Meng, Xiao-Li. (2011). Handbook of Markov Chain Monte Carlo. Florida: CRC Press.
- Brook, S.P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics. 7(4), 434-455. DOI: <https://doi.org/10.1080/10618600.1998.10474787>
- Cressie, N.A.C. (1993). Statistics for Spatial Data. New York: John Wiley and Sons.
- Gelman, A. & Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequence. Journal of Statistical Science, 7(4), 457-511. Available at: <https://www.jstor.org/stable/2246093>
- Goel, A.L. (1983). Technical Report: A Guidebook for Software Reliability Assessment Technical Report, University of Syracuse, New York, pp. 1-240. Available at: <https://apps.dtic.mil/sti/pdfs/ADA139240.pdf>
- Griffith, D.A. (2009). Spatial Autocorrelation. Elsevier Inc., 1-10. Available at: <https://booksite.elsevier.com/brochures/hugy/SampleContent/Spatial-Autocorrelation.pdf>
- Kuo, L., & Yang, T. (1996). Bayesian Computation for Nonhomogeneous Poisson Processes in Software Reliability. Journal of American Statistical Association, 91, 763-773. DOI: <https://doi.org/10.1080/01621459.1996.10476944>
- Lawson, A.B. (2009). Bayesian Disease Mapping Hierarchical Modeling in Spatial Epidemiology. North Western: CRC Press.
- Lentz, J. (2009). Teaching Modul: Spatial Autocorrelation Statistics, 1-10. Available at: <http://jenniferalentz.info/Teaching/StudyGuides/SpatialAutocorrelation.pdf>
- Morales, F.E.C., Vicini, L., Hotta, L.K., & Achcar, J.A. (2016). A Nonhomogeneous Poisson Process Geostatistical Model. Journal of Stoch Environ Res Risk Assess, Springer-Verlag, 31(2), 493-507. DOI: <https://doi.org/10.1007/s00477-016-1275-x>
- Szabo, J.G., Bogachev, L.V., & Chen, H. (2010). Multiple Change-point Poisson Model for Threshold Exceedances of Air Pollution Concentrations. Cornell University Library, 1-39. Available at: <https://arxiv.org/ftp/arxiv/papers/1012/1012.1879.pdf>
- Vicini, L., Hotta, L.K., & Achcar, J.A. 2012. Non-homogeneous Poisson Process in the Presence of One or More Change-points: an Application to Air Pollution Data. Journal of Environmental Statistics, 1-27. Available at: <http://www.jenvstat.org/v05/i03/paper>