



## **Covid-19 Data Analysis in Tarakan with Poisson Regression and Spatial Poisson Process**

**A'yunin Sofro\*, Ika Nurwanitantlya Wardani, Khusnia Nurul Khikmah**

Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, East Java, Indonesia 60231

Email: [ayuninsofro@unesa.ac.id](mailto:ayuninsofro@unesa.ac.id)

### **ABSTRACT**

Coronavirus is included in the virus family that causes diseases ranging from mild to severe symptoms entered Indonesia in March 2020 and included North Kalimantan Province, Tarakan. That caused many implications in every aspect, but the outspread and the patterns still needed to be discovered. One approach was to use Generalized Linear Models. The two methods are Poisson Regression and Stochastic with Spatial Poisson Process. The variables used were rainfall, population density, and temperature in each village in Tarakan. The Poisson Regression analysis finds that only one factor affected temperature. Then, the results were refined with the Spatial Poisson Process, where in addition to the influencing factors also, the distribution patterns are obtained. The analysis showed that the pattern of case distribution was included in the non-homogeneous Poisson process criteria. Then the case density intensity model was obtained using regression from the model known that the covariate variables significantly influence rainfall and temperature. Compared with general Poisson regression analysis results, only the average temperature variables had a significant effect. Thus, a better method was used, namely the Spatial Poisson Process. The two models' AIC values show it. The AIC value of the Spatial Poisson Process model is 89.742, and it was smaller than the Poisson Regression.

Copyright © 2023 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0/>)

**Keywords:** covid-19; deterministic; generalized linear models; Poisson regression; spatial Poisson process; stochastic.

### **INTRODUCTION**

Coronavirus (CoV) is included in the virus family that causes diseases ranging from mild to severe symptoms. Previously, there were two types of coronaviruses that were known to cause illness with severe symptoms, such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). Research said SARS is transmitted from civet cats to humans and MERS from camels to humans. However, in the last few years, around the end of 2019, the international community was shocked by a virus outbreak originating from Wuhan City, China, called Coronavirus Disease or better known as COVID-19 [1]. COVID-19 is a new type of virus that has never been identified before in humans. Coronavirus is a zoonosis, which means it is transmitted between animals and humans.

On December 31, 2019, the WHO China Country Office reported an unknown aetiology pneumonia case in Wuhan City, Hubei Province, China. On January 7, 2020,

China identified this case as a new type of coronavirus (COVID-19) [2]. On January 30 2020, WHO designated the Public Health Emergency of International Concern (PHEIC). The increase in the number of COVID-19 cases took place quite quickly, and there has been a spread outside the Wuhan region and other countries. As of June 8 2020, globally, there were 6,931,000 confirmed cases in 214 countries with 400,857 deaths (CFR 5.78%). With no exception, Indonesia included. As of June 8, 2020, the number of confirmed cases was 32,033 positive cases with 1,883 deaths.

According to the official website covid19.go.id owned by the National Disaster Management Agency (BNPB), on the map of the spread of COVID-19 in Indonesia until June 8, 2020, North Kalimantan Province is currently ranked the 24th and 4th most in Kalimantan after South Kalimantan, Central Kalimantan and East Kalimantan. The map mentioned positive cases of COVID-19 patients in North Kalimantan accumulated as many as 165 cases with 2 cases of death. Tarakan, as of June 8, 2020, became the third contributor to this number, with a total of 46 cases. The information was obtained from the official website of tarakankota.go.id from a source: Press Release Task Force for the Acceleration of COVID-19 Countermeasures of Tarakan, where the case has spread in 14 out of 20 villages in the Tarakan. The increase in the number of cases does not necessarily reveal the pattern of its spread. Reason various factors cause the distribution pattern of COVID-19 cases to be random. Tarakan is elected to be the study location because, in this city, it is known that there has been fairly high population mobility in recent times, which has an effect on population density in some areas. Add to this the erratic weather from March to the present based on the forecast from Tarakan BMKG. These two things are predicted to be able to influence the addition of COVID-19 cases based on previous research. Therefore, researchers are interested in implementing it in Tarakan.

In this article, first, the COVID-19 data will be analyzed with the simplest approach, which is Poisson regression. Only influencing factors that are obtained through this approach [3]–[5]. Thus, the research will be continued with a more relevant method, namely Spatial Poisson Process. The Spatial Poisson Process in this study is used as an approach from Spatial point pattern data. A spatial point pattern is a statistical method for random patterns at points in a dimensional space, where the points represent the location of the research object [6]. COVID-19 outspread, and the variables at risk will be known with this method. The location point data used in this study is one type of spatial data included in the Spatial point pattern, and this study uses a geographical point consisting of each location coordinates where COVID-19 patients live. The location is a random point on the map. This spatial method is used to obtain information from data that is influenced by space or location. COVID-19 cases' outspread patterns can be found by testing clarified data homogeneity with contour plots. The outspread pattern will be tested again using the Poisson Regression method by taking research variables, namely temperature, rainfall, and population density [7]. However, data transformation will be performed before testing with Poisson Regression [6]. The goal is for the data used in regression to have a range that does not differ much, so the analysis can be more effective. After that, a regression analysis can be done to find out what factors can influence COVID-19 outspread in Tarakan. Thus, two results will be obtained, outspread pattern and influential variables.

Previous research related to the Spatial Poisson process method has been conducted [8] regarding the analysis of the mammary gland tumours development in female rats, analysis of the medical centre distribution pattern in Surabaya, the analysis of the gas stations distribution patterns by [9]. Mixed Poisson regression models have also been carried out [10] regarding predictions on heart disease, and also by [11] regarding

Ames salmonella assay data. Meanwhile, research on spatial analysis in COVID-19 cases has been carried out by several researchers, such as [12], which analyze factors of population density, elderly population, urbanization, average temperature, and annual rainfall in each province in Iran using the spatial weight estimation and spatial autocorrelation methods. Other studies were also conducted by [13], who analyzed the number of cases of death and recovery in Wuhan, China using the spatial data panel model, and [14], who analyzed humidity and temperature. Also, regarding the analysis of COVID-19 outbreak with local transmission factors and transportation flows in Brazil. Based on that background, this study will analyze the COVID-19 data in Tarakan using Poisson Regression and the Spatial Poisson Process, which is an approach of Generalized Linear Models (GLM). It also adjusts to the distribution of COVID-19 patients in Tarakan who, after being tested, were distributed to Poisson. This research is expected to provide advice to the government, especially in Tarakan, in optimizing the handling of COVID-19 cases.

## METHODS

### Poisson Distribution

The Poisson distribution belongs to an exponential family. Poisson distribution is one type of distribution of the number of events at certain time intervals. Simeon first announced this distribution - Denis Poisson (1781-1840) published his probability theory in 1838 in his *recherches "sur la probabilit  des jugements en mati re criminelle et en mati re civile work"* (examples of Criminal and Civil Law Opportunities). His work focuses on random variables that count among other discrete amounts. When the expected value occurs at an interval is  $\lambda$ , then the probability of occurring  $x$  times is:

$$p(y; \lambda) := \frac{e^{-\lambda} \lambda^y}{y!}, y = 0, 1, 2, \dots \quad (1)$$

where  $y$  is a non-negative integer stating the number of successes that occur, and  $\lambda$  is a positive actual number which is the average number of successes that occur per unit of time, distance, area, or volume; and  $e = 2.71823$  [15], [16]. The following will be discussed regarding an approach used in this study.

### Generalized Linear Model (GLM)

Generalized Linear Model (GLM) is a general form of the Linear Model. GLM is also a procedure that provides regression analysis and variance analysis for one response variable with one or more explanatory variables, which can also be called a covariate. GLM allows researchers to examine the interactions between responses and each response influence, covariates influence, and the covariate's interaction with responses. For example, it is known that the vector  $y$  has  $n$  components, which is the realization of a  $Y$  response matrix. Each component is independent and is distributed with the mean or  $(Y) = \lambda$ . If the model formed has a predictor of  $x$ , with some unknown parameter  $\beta_1, \dots, \beta_n$ , then the model is a linear combination  $\lambda = \sum_{i=1}^p x_i \beta_i$ . As a transition from the linear model to the generalized linear model, the form is described through three components, namely [17]:

1. Random Component, i.e. values of observing the response of  $Y$  that is independent of specific distributions.
2. Systematic Component, i.e. a linear combination of variable  $x$  with parameter  $\beta$ .
3. Link between random and Systematic/link function, which is a function that explains the expected value of the response variable ( $Y$ ) that connects with explanatory variables through linear equations.

### Poisson Regression

Poisson regression can describe the relationship between response variables, where the response variable is the Poisson distribution with predictor variables. Poisson regression uses GLM to use the model in observations where the response variable does not require normal distribution. The Poisson regression model is a standard model for data whose response variable is abnormally distributed and discrete type, with Poisson distribution as its primary condition. The Poisson distribution provides a realistic model for various random phenomena as long as the value of Poisson's independent variable is a non-negative number. Here is the Poisson regression model:

$$y \sim \text{poisson}(\lambda)$$

$$\lambda_i = e^{x_i' \beta} \tag{2}$$

Where  $\lambda$  is the average value of the event numbers that occur in a specific time interval or area, is a predictor variable notated as follows:

$$x = [1 \quad x_1 \quad x_2 \quad \dots \quad x_k]'$$

Furthermore,  $\beta$  is the Poisson Regression parameter that is noted as follows [18]:

$$\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_k]$$

Poisson Regression analysis can be directly done without doing data transformation first. However, Poisson Regression cannot make a COVID-19 outbreak pattern, which is this study's goal. Because Poisson regression can be done using data count. Meanwhile, spatial data is needed to analyze the distribution patterns, such as the location of patients infected with COVID-19. So that researchers conduct further analysis using the Poisson process method, which will be explained in the following sub.

Parameter estimation in Poisson regression uses the likelihood function and the likelihood equation based on the Poisson distribution. The equation is as follows [19]:

$$L(y_i; \hat{\beta}) = \prod_{i=1}^n \frac{e^{-[\lambda(x_i; \hat{\beta})]} [\lambda(x_i; \hat{\beta})]^{y_i}}{y_i!} \tag{3}$$

### Spatial Point Process

The spatial point process is stochastic (a collection of random variables indexed by space or time) in each realization consisting of a limited and infinite set of points in space, for example,  $S$ , where  $S \subset \mathbb{R}^2$  [20], [21]. The spatial point process is used as a statistical model to analyze the pattern of point distribution, where the point represents the location of an object of research [22]. Spatial point process definitions include processes that are limited and can be calculated. In this study, the discussion will be limited to the point process. For example,  $U$ , whose realization is  $u = \{u_1, \dots, u_n\}$ ,  $n \geq 0$ , is a finite set of parts locally in space or region  $S$ . It is said that the point process is defined at  $S$  to determine the distribution of  $U$ . We can determine the distribution of the number of points ( $U$ ). For each  $n = 1$  depending on  $n(U) = n$ , the combined distribution of  $n$  points on  $U$ . Then, the approximation Equivalent determines the distribution of a variable. for example,  $N(A) = (U \cap A)$  for the subset  $A \subseteq S$ , where  $U \cap A = U \cap A$ . In this study, a spatial point process  $U$  in  $\mathbb{R}^2$  is defined as a set of finite parts locally in  $\mathbb{R}^2$ , i.e. ( $A$ ) is a finite random variable each time  $A \subseteq \mathbb{R}^2$  is a restricted region. Next, the Poisson Regression, which will be the method of approach in this research, will be discussed regarding the COVID-19 case.

### Spatial Poisson Process

In this section, a specific spatial point process model will be discussed by a random intensity function with an analogy using a generalized linear model and a random effect model called the Poisson process. The Poisson process is one of the most widely studied

stochastic processes in the field of point processes and the random phenomena applied science because of their easy-to-use properties as a mathematical model and also mathematically interesting. The Poisson process has two main properties including [23]:

1. The number of Poisson points distribution. In all forms, the Poisson point process is related to the Poisson distribution, which implies that the probability of a random variable Poisson  $X$  is equal to  $x$  given in equation (1), where  $\lambda$  is a single Poisson parameter used to define the Poisson distribution.
2. Complete independence. Another significant trait is that for collections of separate (or non-overlapping) sub-regions of the underlying space, the number of points in each restricted sub-region will be completely independent of the others.

The Poisson process is modelled for complete spatial randomness. Usually, a linear log model of the intensity function is considered as follows:

$$\log \lambda(u) = \mathbf{z}(u)' \beta, u \in S \tag{4}$$

Where  $\beta = (\beta_1, \dots, \beta_n)$  is an unknown parameter and  $\mathbf{z}(u) \equiv (z_1(u), \dots, z_p(u))'$  is a covariate function, the covariate function can be obtained from the pixel image transformation function explained in the following subsection.

The Poisson process is divided into two types, homogenous and non-homogeneous [6], [21], [24]. The Poisson process is homogenous or stationary when it has a constant or single parameter, for example,  $\lambda$ . This parameter is called levels or intensities related to the number of Poisson points expected in some restricted areas. The characteristics of the homogeneous Poisson process are as follows:

1. Number of  $N(U \cap A)$  from the point inside region  $B$  has the Poisson distribution.
2. Expectation value from a point inside region  $B$  is  $E(U \cap A) = \lambda \cdot \text{area}(A)$ .
3. If  $B_1$  and  $B_2$  disjoint or are different regions from space, then  $N(U \cap A_1)$ , and  $N(U \cap A_2)$ , are random variables independent of each other.
4. If  $N(U \cap A) = n$ ,  $n$  points were independent and distributed evenly in region  $A$ .

While the Poisson process is said to be non-homogeneous if the intensity function ( $\lambda$ ) is not constant and varies according to changes in time or area. The non-homogeneous Poisson process with the intensity function ( $u$ ) depends on the  $u$  parameter. The characteristics of the non-homogeneous Poisson process are as follows:

1. Number of  $N(U \cap A)$  from the point inside region  $B$  has the Poisson distribution.
2. Expectation value from a point inside region  $B$  is  $E[N(U \cap A)] = \int_B \lambda(u) du$ . If  $B_1$ ,  $B_2$  disjoint or are different regions from space, then  $N(U \cap A_1)$ , and  $N(U \cap A_2)$ , are random variables which are independent to each other.
3. If  $N(U \cap A) = n$ ,  $n$  points were independent and distributed evenly in region  $B$ , with probability as follows:

$$f(u) = \frac{\lambda(u)}{I}$$

where  $I = \int_B \lambda(u) du$ .

A homogeneity test on the Poisson process will be carried out to determine whether the observed intensity of the point pattern is included in the homogeneous point pattern or non-homogeneous point pattern. Thus, when estimating parameters, obtain a model that matches the characteristics or characteristics of the observed point patterns. The homogeneity test of the Poisson process can be done using the chi-square test. The test hypotheses are as follows:

$H_0$ : The intensity of homogeneous COVID-19 cases

$H_1$ : The intensity of the COVID-19 case is not/ non-homogeneous

The intensity referred to in this study is a large number of confirmed cases in each

location grid. Where the rejection of the hypothesis can be done using the following test statistics:

$$\chi^2 = \sum_{j=1}^r \frac{(n_j - \bar{\lambda}t_j)^2}{\bar{\lambda}t_j}$$

Where  $\bar{\lambda} = \frac{n}{t}$ ,  $n$  is the total number of points, and  $t$  is the total number of regions. The test result will reject  $H_0$  if the p-value  $< \alpha$ , which is  $\alpha = 0.05$  [6].

### Spatial Poisson Process Intensity Function Estimation

In this study, data used is the outspread cases of COVID-19 in Tarakan. MLE (Maximum Likelihood Estimator) is used to find intensity function  $\lambda$  that maximizes data occurring possibilities. Suppose you have an  $n$  sample size,  $u = \{u_1, \dots, u_n\}$  from Spatial Poisson Process as described above. Then the likelihood function  $L(\lambda; u = \{u_1, \dots, u_n\}) = f(u = \{u_1, \dots, u_n\}; \lambda)$  is the probability obtained from sample  $u$ . so the intensity function  $\lambda$  becomes:

$$\Lambda = \int_S \lambda(u) du \tag{5}$$

The observed point  $n$  probability is  $e^{-\Lambda} \frac{\Lambda^n}{n!}$  and the observation probability density is  $\frac{\lambda(u_i)}{\Lambda}$ . Because the observations are independent of each other, then,  $P(u = \{u_1, \dots, u_n\}) = \prod_{i=1}^n \frac{\lambda(u_i)}{\Lambda}$ . The likelihood function to obtain a sample  $u = \{u_1, \dots, u_n\} = P(n) \cdot P(u = \{u_1, \dots, u_n\} | n)$  is:

$$L(\lambda; u = \{u_1, \dots, u_n\}) = e^{-\Lambda} \frac{\Lambda^n}{n!} \prod_{i=1}^n \frac{\lambda(u_i)}{\Lambda} \tag{6}$$

Suppose the sample is ordered  $u_1 < u_2 < \dots < u_n$ . So, likelihood function from an ordered sample is a sample of the likelihood function unordered multiple by  $n!$ . The likelihood function from an ordered sample is:

$$L(\lambda) = e^{-\Lambda} \frac{\Lambda^n}{n!} \cdot \prod_{i=1}^n \frac{\lambda(u_i)}{\Lambda} \cdot n!$$

$$L(\lambda) = e^{-\Lambda} \cdot \prod_{i=1}^n \frac{\lambda(u_i)}{\Lambda} = e^{\int_S \lambda(u) du} \cdot \prod_{i=1}^n \lambda(u_i) \tag{7}$$

So, its likelihood function is as follows [22], [25], [26]:

$$\ell(\beta) = \log(L(\lambda)) = \sum_{i=1}^n \log \lambda_{\beta}(u_i) - \int_S \lambda_{\beta}(u) du \tag{8}$$

In general, numeric integration is needed to calculate integrals [26]. In this study, covariate function data  $Z(u)$  that definite all spatial locations, shown in pixel image or contour plot form, which will be explained in the following sub.

### Poisson Regression in Spatial Poisson Process

Poisson regression analysis was carried out in the spatial Poisson process because general Poisson Regression could not include spatial analysis to find COVID-19 distribution patterns in Tarakan. Regression analysis on the Poisson Process is carried out with the same steps as in general Poisson regression. However, before doing so, the data transformation is performed first. Thus, the model obtained is also different. The following is a Poisson process model with covariate functions [10], [24].

$$\lambda_i(u) = e^{(\beta_0 + \beta_i Z(u)')}, i = 1, 2, \dots, n \tag{9}$$

Where  $\beta = (\beta_1, \dots, \beta_n)$  is an unknown parameter and  $Z(u)' \equiv (z_1(u), \dots, z_p(u))'$  covariate function.

In this section, covariate variables are displayed as pixel images. Nadaraya-Watson smoother was used for spatial smoothing of the mark value at the point distribution. Spatial functions can be written as follows:

$$\tilde{m}(q) = \frac{\sum_{j=1}^n m_j k(q - r_j)}{\sum_{j=1}^n k(q - r_j)}$$

Where  $k$  is a Gaussian kernel function that can be formulated as  $k(n) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(-n^2)}$ ,  $n \in \mathbb{R}$ ;  $q$  is the spatial location.  $r_i$  is the  $i$  data location.  $m_i$  is the  $i$  data mark value, and  $\tilde{m}(q)$  is the mean value for the spatial location of  $u$  [6].

The value range of each covariate variable in this study is different. So, after pixel image transformation, there will be standardization variables to overcome differences in data that are too far away, then standardization to reduce the massive range of covariate variables used in the study. Within the scope of regression, standardization is used to reduce colinearity due to interactions in the regression model [27]. After the data transformation is obtained, the next step is to conduct a regression analysis as usual. Then we conduct the hypothesis testing. Hypothesis testing is a test conducted on all parameter coefficients of the independent variable on the response variable. This test is carried out simultaneously and partially. The test statistic used in the simultaneous test to test the effect of the overall parameter coefficient in the model is the G or likelihood ratio test. Concurrent test hypotheses are as follows:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$  and  $H_1: \text{at least there is } \beta_p \neq 0$ .

Concurrent test statistics used for the model are [24]:

$$G = 2 \log \frac{L(\hat{\beta})}{L(\hat{\lambda})} \tag{10}$$

Furthermore,  $L(\hat{\beta})$  and  $L(\hat{\lambda})$  values are maximum likelihood values for each model. Statistical  $G$  is the approach of the chi-square distribution, so the test criteria are rejected  $H_0$  if  $G > \chi^2_{(\alpha, n)}$  where  $n$  is the number of predictor variables or the number of parameters in the model or  $p\text{-value} < \alpha$ , where the level of significance is  $\alpha = 0.05$ . In comparison, a partial test is performed to determine the effect of the independent variable parameters individually by comparing the standard error. The partial test hypothesis for  $i = 1, 2, \dots, k$  is as follows:

$H_0: \beta_i = 0$  and  $H_1: \beta_i \neq 0$  for  $i = 1, 2, \dots, k$ .

The partial test statistics used for the model are:

$$W = \left[ \frac{\hat{\beta}_i}{SE \hat{\beta}_i} \right]^2 \tag{11}$$

Where  $\hat{\beta}_i$  is estimator  $\beta_i$ .  $SE \hat{\beta}_i$  is the standard error estimator  $\beta$ ,  $W$  is the Wald test following the distribution of  $\chi^2$  with a freedom degree one.  $H_0$  is rejected if  $W > \chi^2_{1, \alpha}$  or  $p\text{-value} < \alpha$ , so it can be concluded that the independent variable influences the response variable.

### Akaike Information Criteria (AIC)

The AIC method is a method that can be used to choose the best regression model found by Akaike and Schwarz. The method is based on the maximum likelihood estimator (MLE) method. Let  $L$  be the maximum value of the likelihood function of a model, and  $k$  is

the number of parameters estimated in the model. Then to calculate the value of AIC can be used the following formula:

$$AIC = 2k - 2 \ln L(\hat{\beta}) \tag{12}$$

Where  $L(\hat{\beta})$  is the likelihood value, and  $k$  is the number of estimated parameters. If multiple models are given for a data set, a better model is a model with a smaller AIC value [28], [29].

**Data Analysis Sources and Techniques**

The data used in this study are secondary data on COVID-19 cases confirmed in Tarakan from 30 March to 8 June 2020 obtained from the official website of the City Government of Tarakan, namely tarakankota.go.id. The data consisted of 46 cases of COVID-19 patients spread across 12 out of 20 villages in Tarakan. The data will be modelled using the Poisson Regression model to determine the predictor variables that affect the response variable, where the response variable ( $Y$ ) is the number of sufferers in each village, and the predictor variable is rainfall ( $X_1$ ) in each village with millimetres, population density ( $X_2$ ) in each village with the unit of person/km, and temperature ( $X_3$ ) in each village with the unit °C, where the rainfall and temperature data are obtained from the BMKG website in Tarakan, while the population density data is obtained from the Tarakan Central Statistics Agency website.

Poisson Regression analysis has not met the objectives of this study, so the analysis was continued with the spatial Poisson process method using data on the number of patients with COVID-19, as many as 46 cases where the location of each patient will be displayed in a plot using latitude and longitude data on google maps. After that, it is divided into 20 grids which are restricted to the number of villages in Tarakan. Then, a homogeneity test can be performed and also displays a contour plot to reinforce the results obtained. Then, performing pixel image transformation and Poisson regression testing can also be done. The covariates used were the same as the predictor variable data in Poisson regression.

**RESULTS AND DISCUSSION**

This section will explain how to find out the distribution patterns of COVID-19 and also the variables that influence it. The first method used is Poisson regression, then proceed with the spatial Poisson process. It aims to compare the results of variables that significantly influence the two methods.

**Modeling and Testing Poisson Regression Parameters**

The data used for the application of Poisson Regression is data on the number of COVID-19 cases per village in Tarakan with predictor variables such as rainfall, population density and temperature. Estimated parameters used in Poisson regression using the Maximum likelihood method. The results obtained from the parameter estimation using the Maximum likelihood method in the following table:

Table 1. Result of Poisson regression parameter estimates.

Parameter	Estimation	Standard Error	z-value	p-value
$\beta_0$	-5,331	3,868	-1,378	0,578
$\beta_1$	1,738	1,386	1,254	0,210
$\beta_2$	$-6,245 \times 10^{-6}$	$1,718 \times 10^{-5}$	-0,363	0,716
$\beta_3$	0,015	$6,484 \times 10^{-3}$	2,305	0,021

Based on Table 1, an estimation result is obtained from the covariate. The estimated



value of the average rainfall is  $-1.738$ , the population density is  $-0.000006245$ , and the average temperature is  $0.01495$ . After obtaining the estimated parameters, the next step is testing the parameters simultaneously to determine whether there is an influence of the predictor variables on the response variable by testing the hypothesis as follows:

$H_0: \beta_1 = \beta_2 = \beta_3$  and  $H_1$ : at least one  $\beta_i \neq 0, i = 1,2,3$ .

The  $G$  value in this analysis is  $65.306$ , and the degree of freedom is  $19$ , while  $\chi^2_{(3)(0.05)} = 7.815$ . Because the value of  $G > \chi^2_{(3)(0.05)}$ , reject  $H_0$ , which means at least one predictor variable that influences the response variable, namely the number of cases of COVID-19.

To find out the effect is by each of these predictor variables, a partial parameter test is performed using the following hypothesis:

$H_0: \beta_i = 0$  and  $H_1$ : at least one  $\beta_i \neq 0, i = 1,2,3$ .

Based on Table 1, shows that the test results failed to reject  $H_0$ , which means that the parameter  $\beta_0$  is not significant because the p-value of  $0.578$  is greater than  $\alpha = 0.05$ . Tests on Parameter  $\beta_1$  obtained the test results failed to reject  $H_0$ , that the p-value of  $0.21$  is greater than  $\alpha = 0.05$ , which means that the variable  $X_1$  does not affect the response variable. Tests on Parameter  $\beta_2$  obtained the test results failed to reject  $H_0$ , that the p-value of  $0.716$  is greater than  $\alpha = 0.05$ , which means that the variable  $X_2$  does not affect the response variable. Tests on parameter  $\beta_3$  obtained the test results reject  $H_0$  that the p-value of  $0.021$  is smaller than  $\alpha = 0.05$ , which means that the  $X_3$  variable influences the response variable. From the results of hypothesis testing that has been done, it is found that the predictor variable that significantly influences temperature. So, the Poisson regression model is obtained as follows:

$$\lambda = e^{(0.01495X_3)}$$

### Spatial Poisson Process Modelling

In this discussion, the distribution pattern analysis will be conducted first. However, before that, the data is pre-processed. The data used to model the Poisson process cannot be used directly, so pre-processing of the data is needed first. The next step of data pre-processing is to group the plots above into 20 grids of the same size. The number of grids was chosen to approach the many urban villages in Tarakan. With the R program's help, the results of the grid division are presented in Figure 1. below.

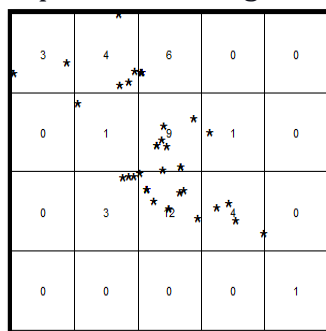


Figure 1. Result of data pre-processing.

### Exploration of Location Data and Covariate Variables

In this discussion, the distribution pattern analysis will be conducted first. However, before that, the data is pre-processed. The data used to model the Poisson process cannot be used directly, so pre-processing of the data is needed first. In this study, the coordinates of each infected village were plotted and plotted with the following display.

In the COVID-19 data, there are 46 observation points obtained from the official website of the City Government of Tarakan. The covariate variables used in the study are

average rainfall, population density magnitude, and the average temperature in each village. In this section, the distribution pattern analysis will be carried out by doing a homogeneity test. The results of R software show that the homogeneity test of the data fulfils the non-homogeneous Poisson process, where the  $p$ -value  $< 0.05$ . It will be clarified again with the results of the contour plots shown in Figure 2. With the R program's help, the resulting drawings where the contour plots formed show the peak density of the spread of COVID-19 in Tarakan.

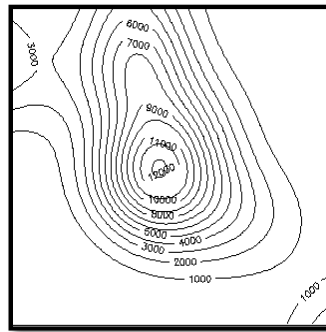


Figure 2. The contour of COVID-19 distribution sites.

Based on the picture above, it is known visually that the data distribution is uneven. The centre of distribution is seen at one point, which is known to be the East Tarakan, Mamburungan Village. This is estimated because the level of population mobility in the East Tarakan region is higher than in other Tarakan areas. Thus, this initial assumption is that the homogeneity of COVID-19 cases in Tarakan is not homogeneous. The characteristics of each covariate variable are illustrated below in Figures 3 (a) to 3 (c).

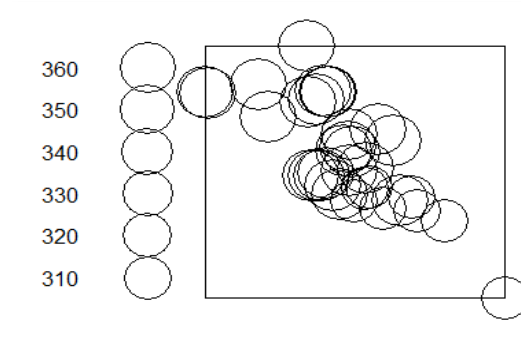


Figure 3(a). Rainfall mark point pattern for each region

Based on Figure 3 (a), the level of rainfall in each village infected by the virus is almost the same. It can be seen that many circles accumulate, and some adjacent intersects patient locations.

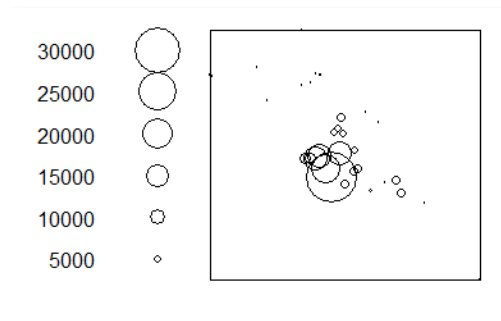


Figure 3(b). Population density Mark point pattern for each region

Based on Figure 3(b), it is known that the level of population density in each village infected by the virus is different. It is seen that several variations of the circle formed.

However, some are similar, so several circles are seen piling up, and some adjacent intersect patient locations.

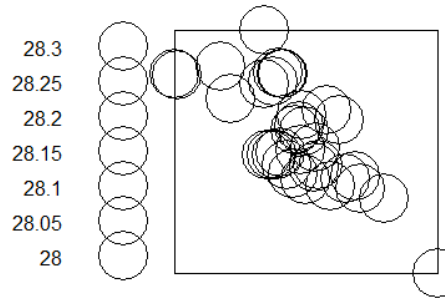


Figure 3(c). Mark point pattern average temperature of each region

Figure 3(c) shows that the temperature level in each village infected by the virus is almost similar. It is seen that many circles accumulate, and some adjacent patient locations intersect. The covariate variables in this study were also pre-processed by transforming them into pixel images, as presented in the following image.

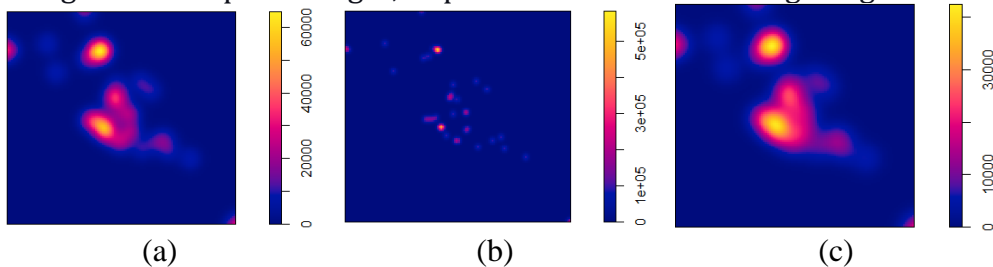


Figure 4. Variable covariate (a) rainfall, (b) population density, and (c) temperature in the form of pixel image

The data analysis step is carried out, beginning with the visualization of density and contour with the help of R software. Then, the homogeneity test of the data is carried out to determine the intensity of the pattern of distribution of data to meet the homogeneous Poisson process or non-homogeneous Poisson process. Then the Poisson process model parameter estimation is performed using the Maximum Likelihood method. The results obtained are then analyzed and interpreted.

### Poisson Regression Model in the Spatial Poisson Process

Data distribution of COVID-19, which has been explored visually, then performed Poisson regression modelling. Estimated parameters used in Poisson regression using the Maximum likelihood method. The results obtained from the estimated parameters using the Maximum likelihood method with the help of the R program are presented in the following table:

Table 2. Results of Estimated Regression Parameters in the Poisson Process

Parameter	Estimation	Standard Error	z-value	p-value
$\beta_0$	-0.023	0.248	-0.093	$8.215 \times 10^{-3}$
$\beta_1$	$-4.589 \times 10^{-3}$	$2.339 \times 10^{-3}$	-1.962	0.0498
$\beta_2$	$1.294 \times 10^{-3}$	$8.394 \times 10^{-4}$	1.542	0.123
$\beta_3$	$3.459 \times 10^{-3}$	$1.535 \times 10^{-3}$	2.253	0.024

Based on table 2, the estimation results from the covariate are obtained. The average value of average rainfall is  $-0.004589$ , the estimated value of the population is  $0.001294$ , and the average estimated value is  $0.003459$ . After obtaining the estimated parameters,

then the simultaneous parameter testing is to determine whether there is a covariate effect on the response variable by conducting the following hypothesis test:

$H_0: \beta_1 = \beta_2 = \beta_3$  and  $H_1: \text{at least one } \beta_i \neq 0, i = 1,2,3.$

The  $G$  value in this analysis is 86.669, and the degree of freedom is 19, while  $X\chi^2_{(3)(0.05)} = 7.815$ . Because the value is  $G > \chi^2_{(3)(0.05)}$ , it rejects  $H_0$ , which means that at least one free variable is against the COVID-19 case. To determine the effect of each covariate, partial parameters were tested using the following hypotheses:

$H_0: \beta_i = 0$  and  $H_1: \text{at least one } \beta_i \neq 0, i = 1,2,3.$

Based on Table 2, it was shown that the test results were rejected by  $H_0$ , which means that the parameter  $\beta_0$  is significant because the p-value of 0.008215 is smaller than  $\alpha = 0.05$ . Tests on Parameter  $\beta_1$  obtained the rejected  $H_0$  test results, namely the p-value of 0.0498 is smaller than  $\alpha = 0.05$ , which means that the variable  $X_1$  affects the response variable. Tests on Parameter  $\beta_2$  obtained the rejected  $H_0$  test results. The p-value of 0.1231 is greater than  $\alpha = 0.05$ , which means that the  $X_2$  variable does not conflict with the response variable. Tests on Parameter  $\beta_3$  obtained the rejected  $H_0$  test results. The p-value of 0.0242 is smaller than  $\alpha = 0.05$ , which means that the  $X_3$  variable applies to the response variable. From the results of hypothesis testing that has been done, the significant covariates needed are rainfall and temperature. So, the following model is obtained:

$$\lambda = e^{(-0,022987-0,00459X_1-0,00346X_3)}$$

### **Selection of The Best Model**

Based on the data processing results of several variables used in this study, AIC values will be obtained to find the best model. The AIC value is obtained based on the log-likelihood value previously described. AIC calculation results can be seen in the following table:

Table 3. Models obtained from Both Methods.

Method	AIC
Poisson Regression	101.38
Regression on Poisson process	89.742

In Table 3, the AIC values for each model are obtained with the help of *R* software. Furthermore, from this table, it can be seen that the Spatial Poisson Process model is better than the usual Poisson Regression model, namely with the AIC value is 89.742. So, the best model is obtained from the regression method in the Spatial Poisson Process. From the results obtained in the distribution analysis of COVID-19 cases in Tarakan, differences were found from previous studies. In [12] study, the results showed that the factors that determine the high cases of COVID-19 in Iraq, namely the high rate of urbanization and the high cases of the elderly, the equation of the average ambient temperature could also determine the level of additional cases.

The results of this study support previous research in Shokouhi [14], which found eight countries, such as China, Japan, South Korea, and others. COVID-19 cases are affected by temperature, humidity, and latitude. Therefore in this study, improvisation can be done is that researchers can prove that weather factors that can affect COVID-19 cases, not only environmental temperature but the level of rainfall, can also affect the increase in COVID-19 cases in an area. However, everything is inseparable from the characteristics of each region is different. Each country has its environmental conditions.

Meanwhile, temperature and rainfall can be tested, and many other environmental factors are also predicted to influence the increasing cases of COVID-19 in an area. As in

the [14] study, the results show that the humidity and temperature factors significantly influence the increasing cases of COVID-19 in each country. As for this study, it did not use the humidity factor because the data obtained were the same for each village, so further analysis could not be done.

## CONCLUSIONS

Based on the conducted analysis results, direct modelling using Poisson regression obtained the results of only temperatures that have a significant effect on the response variable. Then, after further analysis using the Spatial Poisson Process, another result was obtained from the distribution pattern of COVID-19 in Tarakan. Where the distribution pattern is visually not homogeneous or included in the non-homogeneous Poisson process criteria, this can be seen from the results of the contour plot, where there is a peak spread of cases in the East Tarakan area. The average rainfall and temperature significantly affect the COVID-19 intensity model in Tarakan using the regression method in the Spatial Poisson Process. The AIC test results also showed that the Spatial Poisson Process model is better than the regular Poisson Regression. All the analysis results are expected to be information for the city government regarding handling COVID-19 cases in the City of Tarakan going forward. The research proves that temperature and rain significantly affect the spread of COVID-19 in Tarakan. The results of this discovery can be used to provide directions to the community in related seasons to be more vigilant. In future studies, please combine COVID-19 case data with a map of the area under study. This will give better and more accurate results than just using a grid because each region has a different character.

## REFERENCES

- [1] A. Hafeez, S. Ahmad, S. A. Siddqui, M. Ahmad, and S. Mishra, "A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention," *Ejmo*, vol. 4, no. 2, pp. 116–125, 2020.
- [2] R. Djalante *et al.*, "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," *Progress in disaster science*, vol. 6, p. 100091, 2020.
- [3] L. Kuhn, L. L. Davidson, and M. S. Durkin, "Use of Poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program," *Am J Epidemiol*, vol. 140, no. 10, pp. 943–955, 1994.
- [4] D. L. Preston, "Poisson regression in epidemiology," *Encyclopedia of biostatistics*, vol. 6, 2005.
- [5] R. Bender, "Introduction to the use of regression models in epidemiology," in *Cancer Epidemiology*, Springer, 2009, pp. 179–195.
- [6] E. Gabriel, "A. Baddeley, E. Rubak, R. Turner: Spatial Point Patterns: Methodology and Applications with R." Springer, 2017.
- [7] Z. Sun, H. Zhang, Y. Yang, H. Wan, and Y. Wang, "Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China," *Science of The Total Environment*, vol. 746, p. 141347, 2020.
- [8] J. F. Lawless, "Regression methods for Poisson process data," *J Am Stat Assoc*, vol. 82, no. 399, pp. 808–815, 1987.
- [9] A. N. Syaifulloh, N. Iriawan, and P. P. Oktaviana, "Analisis Pola Persebaran Stasiun Pengisian Bahan Bakar Umum (SPBU) Wilayah Surabaya Menggunakan Spatial Poisson Point Process," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, pp. D57–D64, 2020.

- [10] C. Mufudza and H. Erol, "Poisson mixture regression models for heart disease prediction," *Comput Math Methods Med*, vol. 2016, 2016.
- [11] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Min Knowl Discov*, vol. 26, no. 2, pp. 275–309, 2013.
- [12] R. Ramírez-Aldana, J. C. Gomez-Verjan, and O. Y. Bello-Chavolla, "Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level," *PLoS Negl Trop Dis*, vol. 14, no. 11, p. e0008875, 2020.
- [13] H. Guliyev, "Determining the spatial effects of COVID-19 using the spatial panel data model," *Spat Stat*, vol. 38, p. 100443, 2020.
- [14] M. D. Shokouhi, F. Miralles-Wilhelm, M. D. A. Amoroso, and M. M. Sajadi, "Temperature, Humidity, and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19.," *Working paper*, 2020.
- [15] W. M. Meredith, "THE POISSON DISTRIBUTION AND POISSON PROCESS IN PSYCHOMETRIC THEORY 1," *ETS Research Bulletin Series*, vol. 1968, no. 2, pp. i–81, 1968.
- [16] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and statistics for engineers and scientists*, vol. 5. Macmillan New York, 1993.
- [17] J. K. Lindsey, "Applying Generalized Linear Models. Springer, New York,," 1997.
- [18] S. Yang and G. Berdine, "Poisson regression," *The Southwest Respiratory and Critical Care Chronicles*, vol. 3, no. 9, pp. 61–64, 2015.
- [19] J. A. Santos and M. M. Neves, "A local maximum likelihood estimator for Poisson regression," *Metrika*, vol. 68, no. 3, pp. 257–270, 2008.
- [20] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of spatial statistics*. CRC press, 2010.
- [21] T. D. Johnson, "Introduction to spatial point processes," *Www-Ljk.Imag.Fr*, 2008.
- [22] J. Møller and R. P. Waagepetersen, "Modern statistics for spatial point processes," *Scandinavian Journal of Statistics*, vol. 34, no. 4, 2007, doi: 10.1111/j.1467-9469.2007.00569.x.
- [23] H. P. Keeler, "Notes on the Poisson point process," *Weierstrass Inst., Berlin, Germany, Tech. Rep*, 2016.
- [24] A. Baddeley, "Analysing spatial point patterns in R," in *Workshop notes version*, 2008.
- [25] M. Berman and T. R. Turner, "Approximating point process likelihoods with GLIM," *J R Stat Soc Ser C Appl Stat*, vol. 41, no. 1, pp. 31–38, 1992.
- [26] L. C. Drazek, "Intensity estimation for Poisson processes," *The University of Leeds, School of Mathematics*, 2013.
- [27] W. H. Finch and J. E. Bolin, *Multilevel modeling using Mplus*. Chapman and Hall/CRC, 2017.
- [28] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.
- [29] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *Wiley Interdiscip Rev Comput Stat*, vol. 11, no. 3, p. e1460, 2019.