International Journal on Advanced Science Engineering Information Technology

Vol.9 (2019) No. 3
ISSN: 2088-5334

Comparison of Nelder Mead and BFGS Algorithms on Geographically Weighted Multivariate Negative Binomial

Yuliani.S.Dewi[*], Purhadi[#1], Sutikno[#2], Santi.W.Purnami[#3]

[*]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Jember, Jember, 68121 Indonesia*
E-mail: yulidewi.fmipa@unej.ac.id

[#]*Department of Statistics, Faculty of Mathematics, Computation and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya 60111 Indonesia*
E-mail: [1]purhadi@statistika.its.ac.id; [2]sutikno@statistika.its.ac.id; [3]santi_wp@statistika.its.ac.id

***Abstract*— Geographically Weighted Negative Binomial Regression (GWNBR) was proposed related to univariate spatial count data with overdispersion using MLE via Newton Raphson algorithm. However, the Newton Raphson algorithm has the weakness, it tends to depend on the initial value. Therefore, it can have false convergence if the initial value is mistaken. In this research, we derive estimating the mean of dependent variables of multivariate spatial count data with overdispersion, Geographically Weighted Multivariate Negative Binomial (GWMNB) and compare it to the global method, multivariate negative binomial (MNB). We use MLE via Nelder Mead and Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithms. We conduct the simulation study and application of mortality data to find out the characteristics of the methods. They show that GWMNB performs better than global method (MNB) in estimating the means of dependent variables of the spatial data. The Nelder Mead tends to be more successful in estimating the means for all locations than BFGS algorithm. Although BFGS is a stable algorithm in MNB related to the initial value, it tends to have false convergence in GWMNB. The mortality rate of infant is larger than it of toddler and preschool and also maternal. The highest deaths of infant, toddler, and preschool and also maternal tend to happen in east parts of East Java.**

***Keywords*— spatial data; over dispersion; GWMNB, MLE; nelder mead; BFGS.**

I. INTRODUCTION

Poisson regression is usually used for modeling count data. Related to multivariate count data, an early overview of multivariate regression for count data is described [1], [2]. Some studies have been done related to bivariate Poisson data. Some of them are simulated maximum likelihood estimation of bivariate count data with unrestricted correlation pattern of unobserved heterogeneity [3], EM algorithm for estimating the parameters of bivariate and diagonal inflated bivariate Poisson regression models [4], and bivariate generalized Poisson for healthcare data [5]. Moreover, the researches about multivariate Poisson have been proposed, a multivariate Poisson regression using the MLE method via the EM algorithm and Bayesian [6], and multivariate generalized Poisson regression [7].

Behind the popularity of Poisson regression, it has a weakness regarding equidispersion assumption. It restricts the use of the method because many data in real applications are under or overdispersed data. One of the methods for overcoming overdispersion is by using a negative binomial model. There are some studies related to the model for more than one dependent variables. Those are Bivariate Negative Binomial Regression (BNBR) using MLE via Newton Raphson algorithm [8], the comparison of two bivariate negative binomial regression models which come from the different distributions derivation [9], the Seemingly Unrelated Negative Binomial (SUNB) using Generalized Nonlinear Least Square (GNLS) [10], the multivariate negative binomial model using copula and MLE method [11], and a robust likelihood approach for the overdispersed correlated count data analysis based on a multivariate negative binomial model using MLE via iterative Newton Raphson method [12].

Global regression assumes that the relationships being measured are stationary over space. The parameter estimates are applied equally over the whole region or the relationships being measured are assumed to be stationary over space. Therefore, if there is a variation of the relationships between locations, the global model is not suitable to the reality. One of the spatial effects is spatial heterogeneity [13]. This happens when dealing with spatial data as the processes generating them might vary across space. This gives effect to the parameters of the model varying or not homogeneous between locations. Those led to the development of the methods to model spatial data. One of them is the method based on spatial heterogeneity that is Geographically Weighted Regression (GWR) model.

Regarding to the spatial count data with over dispersion, Geographically Weighted Negative Binomial (GWNB) to model univariate count data [14] and the comparison of zero-inflated Poisson (ZIP) and Geographically Weighted Zero-Inflated Poisson (GWZIP) using the MLE method for modeling excess zero univariate count data [15] have been proposed. On the other side, many data in real applications are multivariate count data.

Therefore, in this research, we propose estimating the means of multivariate spatial count data with over dispersion using spatial weights and negative binomial distribution (Geographically Weighted Multivariate Negative Binomial). The model can be written as $Y_{ij} \sim \text{MNB}\left[\mu_{ij}\left(\boldsymbol{\beta}_j\left(u_i,v_i\right)\right),\tau\left(u_i,v_i\right)\right]$ where Y_{ij} is the dependent variable-*j* on location-*i*, μ_{ij} is the mean of dependent variable-*j* on location-*i* with the coefficients at equation point *i*, $\tau\left(u_i,v_i\right)$ is dispersion on location-*i*, $E\left(Y_{ij}\right)=\mu_{ij}\left(u_i,v_i\right)=t_{ij}e^{\mathbf{x}_{ij}^T\boldsymbol{\beta}_j\left(u_i,v_i\right)}$, $\mathbf{x}_{ij}^T=\left(x_{ij0},x_{ij1},...,x_{ij(p-1)}\right)$, $\boldsymbol{\beta}_j^T\left(u_i,v_i\right)=\left(\beta_{j0}\left(u_i,v_i\right),\beta_{j1}\left(u_i,v_i\right),...,\beta_{j(p-1)}\left(u_i,v_i\right)\right)$, t_{ij} is exposure on location-*i*, dependent variable-*j*, $\mathbf{x}_{ij}$ is the independent variable of dependent variable-*j* on location-*i*, and $\boldsymbol{\beta}_j\left(u_i,v_i\right)$ is vector coefficient of independent variable for dependent variable-*j* on location-*i*, *i* =1,2,...,*n*, *j*=1,2,..,*m*.

GWZIP uses the EM algorithm for estimating the parameters [15]. GWNB uses the Newton Raphson algorithm for the estimation [14]. However, the Newton Raphson algorithm has a weakness in dealing with false convergence due to the improper initial value. Besides that, if there are issues in obtaining second partial derivative matrix of likelihood function or the second partial derivative of the likelihood function is singular, the Newton Raphson algorithm cannot be implemented regularly. The formation of the stiff information matrix, especially the calculation of its inverse are computationally expensive.

On the other hand, when analyzing more than one dependent variables, things are more complicated, especially the multivariate count regression models are less developed. Based on those, Focusing on multivariate Poisson data with overdispersion, geographically weighted multivariate negative binomial (GWMNB) using MLE via BFGS and Nelder Mead algorithm is considered and derived in this research. BFGS algorithm needs the first derivative of the likelihood function for its process. Nelder Mead algorithm is a free derivative method. The algorithms are more robust than the Newton Raphson algorithm [16], [17]. In this paper, geographically weighted regression is more used as prediction tools than inference [18], and also as exploratory data [19]. Therefore, we focus on estimating the mean of dependent variables then assess it using the Goodness of Fit and represent them into the maps rather than modeling and testing the hypothesis of the parameter estimates.

The paper is organized as follows. In section II we describe the multivariate negative binomial distribution and provide estimation the means of dependent variables of count data by using the geographically weighted multivariate negative binomial method and MLE via BFGS and Nelder Mead algorithms. The simulation study using the two algorithms and the application of the methods are discussed in section III. We use mortality data of East Java, Indonesia, 2014 as an application.

II. Materials and Methods

A. Multivariate Negative Binomial Distribution

GWMNB method in this research is built based on the marginal probability function of the multivariate negative binomial distribution

$$f\left(\mathbf{y}_i;\boldsymbol{\mu},\delta\right)=\frac{\left(\prod_{j=1}^{m}\mu_{ij}^{y_{ij}}\right)\delta^{\delta}\Gamma\left(\delta+y_{i+}\right)}{\left(\prod_{j=1}^{m}y_{ij}!\right)\Gamma\left(\delta\right)\left(\delta+\mu_{i+}\right)^{\delta+y_{i+}}} \quad (1)$$

where $\delta=\tau^{-1}$, *i* = 1,2,...,*n* and *j* = 1,2,...,*m*, $y_{i+}=\sum_{j=1}^{m}y_{ij}$ and $\mu_{i+}=\sum_{j=1}^{m}\mu_{ij}$.

The mean, variance, covariance, and correlation of Y_{ij} can be written as [20]

$$E\left(y_{ij}\right)=\mu_{ij},\quad \text{var}\left(y_{ij}\right)=\tau\mu_{ij}^2+\mu_{ij} \quad (2)$$

$$\text{cov}\left(y_{ij},y_{ij'}\right)=\tau\mu_{ij}\mu_{ij'},\text{for } j\neq j' \quad (3)$$

$$\text{corr}\left(y_{ij},y_{ij'}\right)=\frac{\tau\sqrt{\mu_{ij}\mu_{ij'}}}{\sqrt{1+\tau\mu_{ij}}\sqrt{1+\tau\mu_{ij'}}},\text{for } j\neq j' \quad (4)$$

for *i*=1,2,…,*n* and $j,j'=1,2,\ldots,m$. From the equation(4), the correlation is always positive and for small τ, count data y_{ij} close to the observations that have independent Poisson distribution and each observation has mean and variance μ_{ij} .

B. Estimation The Means of Dependent Variables of Geographically Weighted Multivariate Negative Binomial

Based on GWNB [14] and equation (1), the loglikelihood function for GWMNB to estimate the equation coefficients and dispersion index on location-*l* can be written as follows:

$$\ell\left(\boldsymbol{\theta}^*\right)=\sum_{i=1}^{n}\left\{\sum_{j=1}^{m}y_{ij}\ln\mu_{ij}\left(u_l,v_l\right)+\tau^{-1}\left(u_l,v_l\right)\ln\tau^{-1}\left(u_l,v_l\right)-A_i+B_i\right\}w_{il} \quad (5)$$

where

$$A_i=\sum_{j=1}^{m}\ln y_{ij}!+\left(\tau^{-1}\left(u_l,v_l\right)+y_{i+}\right)\ln\left(\tau^{-1}\left(u_l,v_l\right)+\mu_{i+}\left(u_l,v_l\right)\right),$$

$$B_i = \ln\left(\frac{\Gamma\left(y_{i+} + \tau^{-1}(u_l,v_l)\right)}{\Gamma\left(\tau^{-1}(u_l,v_l)\right)}\right); \; i, l = 1,2,3,\ldots,n, \; y_{ij} = 0,1,2,\ldots,$$

for $j = 1,2, \ldots,m$, $y_{i+} = \sum_{j=1}^{m} y_{ij}$ and $\mu_{i+}(u_l,v_l) = \sum_{j=1}^{m} \mu_{ij}(u_l,v_l)$.

$\boldsymbol{\theta}^* = \boldsymbol{\theta}(u_l,v_l) = \left(\boldsymbol{\beta}_1^T(u_l,v_l), \boldsymbol{\beta}_2^T(u_l,v_l), \ldots, \boldsymbol{\beta}_m^T(u_l,v_l), \tau(u_l,v_l)\right)^T$ and w_{il} is the geographical weight. In this research, we use the Fixed Bisquare Kernel weight, [21]

$$w_{il} = \begin{cases} \left(1 - \left(\frac{d_{il}}{b}\right)^2\right)^2; & \text{untuk } d_{il} \le b \\ 0 \; ; & \text{untuk } d_{il} > b \end{cases} \quad (6)$$

where w_{il} is the weight of the observation on location-i for estimating the coefficients and index of dispersion on location- l, b is the bandwidth, and d_{il} is Euclidean distance between point-i and l.

The optimum bandwidth is gotten by cross-validation using the formula

$$CV(b) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij} - \hat{y}_{\neq ij}(b)\right)^2 \quad (7)$$

where $\hat{y}_{\neq ij}(b)$ is the prediction value of y_{ij} which is predicted without the observation on location-i and n is the number of location. The optimum bandwidth is the bandwidth with CV minimum. We use the golden section algorithm to find optimum bandwidth. We consider BFGS and Nelder Mead algorithms for estimating the means of dependent variables of MNB and GWMNB.

1) BFGS Algorithm for GWMNB: BFGS is an optimization algorithm in the family of quasi-Newton methods. The algorithm needs a gradient of the likelihood function that we maximize in each iteration. Hessian Matrix is approximated by iteration of gradient evaluation [16], [22]. The gradient is obtained by deriving log likelihood function with respect to the coefficient $\boldsymbol{\beta}_j(u_l,v_l)$ and $\tau(u_l,v_l)$.

$$\mathbf{U}_{\boldsymbol{\beta}_j(u_l,v_l)} = \frac{\partial \ell(\boldsymbol{\theta}(u_l,v_l))}{\partial \boldsymbol{\beta}_j(u_l,v_l)} = \sum_{i=1}^{n}\left\{ y_{ij}\mathbf{x}_{ij} - \frac{\left(\tau^{-1}(u_l,v_l) + y_{i+}\right)\mathbf{x}_{ij}\mu_{ij}(u_l,v_l)}{\tau^{-1}(u_l,v_l) + \mu_{i+}(u_l,v_l)} \right\} w_{il} \quad (8)$$

for i, l = 1,2,...,n and j = 1,2,...,m; m is the number of dependent variables, $\mathbf{x}_{ij}$ is vector with its elements, 1 and p-1 independent variables in each location-i, while $\mathbf{y}_j$ and $\boldsymbol{\mu}_j(u_i,v_i)$ are n x 1 vectors. A Derivative of the likelihood function with respect to the index $\tau(u_l,v_l)$ is given by

$$U_{\tau(u_l,v_l)} = \frac{\partial \ell(\boldsymbol{\theta}(u_l,v_l))}{\partial \tau(u_l,v_l)} = \sum_{i=1}^{n}\left(-\ln\left(1+\tau(u_l,v_l)\mu_{i+}(u_l,v_l)\right) + \sum_{k=0}^{y_{i+}-1}\left(\tau^{-1}(u_l,v_l)+k\right)^{-1} + C_i\right)\tau^{-2} w_{il} \quad (9)$$

where $C_i = \frac{\left(\mu_{i+}(u_l,v_l) - y_{i+}\right)}{\left(\tau^{-1}(u_l,v_l) + \mu_{i+}(u_l,v_l)\right)}$ and the summation $\sum_{k=0}^{y_{i+}-1}\left(\tau^{-1}(u_l,v_l)+k\right)^{-1} = 0$ for $y_{i+} < 1$ [23].

Let $R^p \to R$ be continuously differentiable. Consider the following unconstrained optimization problem:

$$\min \; -\ell(\boldsymbol{\theta}(u_l,v_l)), \boldsymbol{\theta}(u_l,v_l) \in R^p,$$

BFGS method generates a sequence $\{\boldsymbol{\theta}_r(u_l,v_l)\}$ iteratively [22].

Starting with an initial value $\boldsymbol{\theta}_0$ and Hessian matrix $\mathbf{H}_0$, the algorithm repeats these steps until $\boldsymbol{\theta}_r(u_l,v_l)$ is converged.

1. The direction $\mathbf{q}_r$ is obtained by solving the equation $\mathbf{H}_r\mathbf{q}_r = -\mathbf{U}(\boldsymbol{\theta}_r(u_l,v_l))$
   where $\mathbf{U}(\boldsymbol{\theta}_r)$ is gradient function in iteration r^{th} or

$$\mathbf{U}(\boldsymbol{\theta}_r(u_l,v_l)) = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta}(u_l,v_l))}{\partial \beta_{1,0}(u_l,v_l)} \\ \vdots \\ \frac{\partial \ell(\boldsymbol{\theta}(u_l,v_l))}{\partial \beta_{m,p-1}(u_l,v_l)} \\ \frac{\partial \ell(\boldsymbol{\theta}(u_l,v_l))}{\partial \tau(u_l,v_l)} \end{bmatrix}_r, \quad \mathbf{q}_r = \begin{bmatrix} q_1 \\ \vdots \\ q_{mp+1} \end{bmatrix} \text{ and}$$

   $\mathbf{H}_r$ is a Hessian matrix in iteration r^{th}
2. Generate $\boldsymbol{\theta}_{r+1}(u_l,v_l) = \boldsymbol{\theta}_r(u_l,v_l) + \alpha_r\mathbf{q}_r$, r = 0, 1, 2,…
3. The matrix $\mathbf{H}_r$ is updated using the formula

$$\mathbf{H}_{r+1} = \mathbf{H}_r + \frac{\Delta\mathbf{U}_r\Delta\mathbf{U}_r^T}{\Delta\mathbf{U}_r^T\mathbf{s}_r} - \frac{\mathbf{H}_r\mathbf{s}_r\mathbf{s}_r^T\mathbf{H}_r}{\mathbf{s}_r^T\mathbf{H}_r\mathbf{s}_r}$$

   where $s_r = \alpha_r\mathbf{q}_r$ and $\Delta\mathbf{U}_r = \mathbf{U}(\boldsymbol{\theta}_{r+1}(u_l,v_l)) - \mathbf{U}(\boldsymbol{\theta}_r(u_l,v_l))$.

   The matrix $\mathbf{H}_r$ is positive definite if $\Delta\mathbf{U}_r^T s_r > 0$. The condition is guaranteed to hold if step length α_r is chosen by the exact line search

$$-\ell(\boldsymbol{\theta}_r(u_l,v_l) + \alpha_r\mathbf{q}_r) = \min_{\alpha>0} -\ell(\boldsymbol{\theta}_r(u_l,v_l) + \alpha\mathbf{q}_r)$$

   or the Wolfe-type inexact line search

$$-\ell(\boldsymbol{\theta}_r(u_l,v_l) + \alpha_r q_r) \le -\ell(\boldsymbol{\theta}_r(u_l,v_l)) + c_1\alpha_r\mathbf{U}_r^T\mathbf{q}_r,$$

$$\mathbf{U}(\boldsymbol{\theta}_r(u_l,v_l) + \alpha_r\mathbf{q}_r)^T\mathbf{q}_r \le c_2\mathbf{U}_r^T\mathbf{q}_r,$$

   for $c_1, c_2 \in (0,1)$, $c_1 < c_2$.

As stated in [24], Newton's method often fails to converge because of the poor initial estimate, therefore the convergence is checked from the Euclidean norm of the gradient, $\|\Delta\mathbf{U}_r\|$.

2) Nelder Mead Algorithm for GWMNB: Nelder Mead is a free derivative algorithm. The algorithm finds the coefficients and index of dispersion estimate using reflection, expansion, contraction and shrinkage process. Let $\mathbf{\Theta}$ is the p dimension parameter space and $\ell\left(\boldsymbol{\theta}\left(u_l,v_l\right)\right)$ is the log likelihood weighted by geographical weight or $y=-\ell\left(\boldsymbol{\theta}^*\right)=-\ell\left(\boldsymbol{\theta}\left(u_l,v_l\right)\right)$ is the objective function which be minimized, where $\boldsymbol{\theta}(u_l,v_l)\in\mathbf{\Theta}$ and $\boldsymbol{\theta}^*=\boldsymbol{\theta}\left(u_l,v_l\right)=$ $[\beta_{1,0}\left(u_l,v_l\right)\ \beta_{1,1}\left(u_l,v_l\right)\ ...\beta_{1,p-1}\left(u_l,v_l\right)...\ \beta_{m,0}\left(u_l,v_l\right)\ \beta_{m,1}\left(u_l,v_l\right)$ $...\ \beta_{m,p-1}\left(u_l,v_l\right)\ \tau\left(u_l,v_l\right)]$.

Each iteration of the algorithm is started from simplex in $\mathbf{\Theta}$, that is the structure formed by p +1 vertex, $\widehat{\boldsymbol{\theta}}_{(1)}\left(u_l,v_l\right)$ $,\ldots,\widehat{\boldsymbol{\theta}}_{(p+1)}\left(u_l,v_l\right)$, where $p=mp+1$. If the optimum is minimum, the algorithm replaces the worst vertex in a simplex with the new point that has a smaller value of minus log likelihood using one of the reflection, expansion, or contraction operation. If those fail to find the new point to replace the worst point on the simplex, then the shrinkage operation is carried out. See [17], [25], [26], [27], and [28] for Nelder Mead algorithm in general. Fig. 1 represents the workflow of Nelder Mead algorithm for GWMNB in this research.

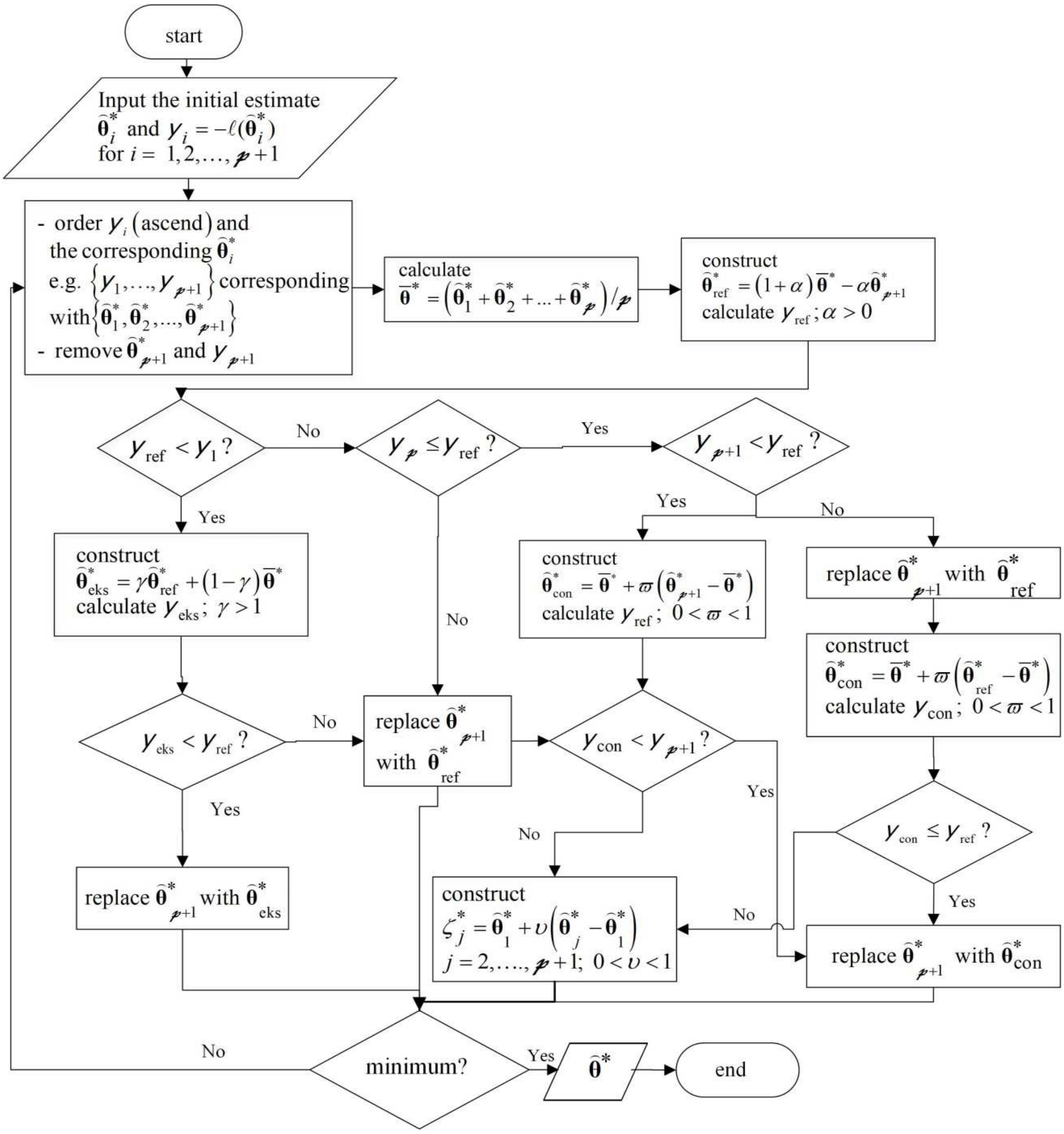


Fig. 1 The flowchart of Nelder Mead algorithm for GWMNB

III. Results and Discussion

In this section we assume $\mathbf{x}_{i1}^T = \mathbf{x}_{i2}^T = ... = \mathbf{x}_{im}^T = \mathbf{x}_i^T = \begin{bmatrix} 1 & x_{i1} & x_{i2} & ... & x_{i(p-1)} \end{bmatrix}$, then $\ln \mu_{ij}(u_i, v_i) = \ln\left(t_{ij} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j(u_i,v_i)}\right)$ for i = 1,2,...,n and j = 1,2, …,m. By considering the spatial effect we estimate the means of the dependent variables, where $Y_{ij} \sim \text{MNB}\left[\mu_{ij}\left(\boldsymbol{\beta}_j(u_i,v_i)\right), \tau(u_i,v_i)\right]$, j =1,2,…,m. and $\mu_{ij}(u_i,v_i) = t_{ij} \exp\left(\sum_k \beta_{jk}(u_i,v_i) x_{ik}\right)$, t_{ij} is exposure variable on dependent variable-j on location-i. The coordinates of latitude and longitude represent a geographical factor of locations. We use the latitude and longitude of the regencies in East Java Indonesia as a geographical reference, where there are 38 regencies. We use Euclid distance to measure the distance between regencies/towns. The optimum bandwidth is the bandwidth that has a minimum CV.

A. Simulation Study

We conduct the simulation to find out the performance of MNB and GWMNB using MLE method via Nelder Mead and BFGS algorithms in estimating the means of dependent variables. The simulation is conducted using the R software based on the procedure presented in [19] and [14]. We generate two independent variables based on equation (12) and (13), and trivariate dependent variables which have a positive correlation between them.

For generating the dependent variables, we set some coefficients, $\beta_{11}(u_i,v_i)$, $\beta_{12}(u_i,v_i)$ and $\tau(u_i,v_i)$ as a function of locations to reckon spatial heterogeneity in the data [14]. We use the latitude and longitude of East Java as a geographical factor. It consists of 38 locations. Therefore we use the sample size n= 38 in this simulation study. We set the coefficients $\left(\boldsymbol{\beta}(u_i,v_i)\right)$ and index of dispersion $\left(\tau(u_i,v_i)\right)$ as follows:

$$\beta_{10}(u_i,v_i) = 1,\ \beta_{11}(u_i,v_i) = \sqrt{[u_i - \bar{u}]^2 \times [v_i - \bar{v}]^2},$$

where $\bar{u} = \dfrac{\sum_{i=1}^{n} u_i}{n}$ and $\bar{v} = \dfrac{\sum_{i=1}^{n} v_i}{n}$.

$$\beta_{12}(u_i,v_i) = 26\left\{-\left[(u_i - \bar{u})/5\right]^2 - \left[(v_i - \bar{v})/5\right]^2 + 0{,}13\right\},$$

$$\beta_{20}(u_i,v_i) = 2,\ \beta_{21}(u_i,v_i) = -0.5,\ \beta_{22}(u_i,v_i) = 1,$$

$$\beta_{30}(u_i,v_i) = 0.5,\ \beta_{31}(u_i,v_i) = -0.5,\ \beta_{32}(u_i,v_i) = -1 \text{ and}$$

$$\tau(u_i,v_i) = 20\left\{10^{-11}\left[|v_i|\right]^5\right\}^2;\ t_{ij} = 1.$$

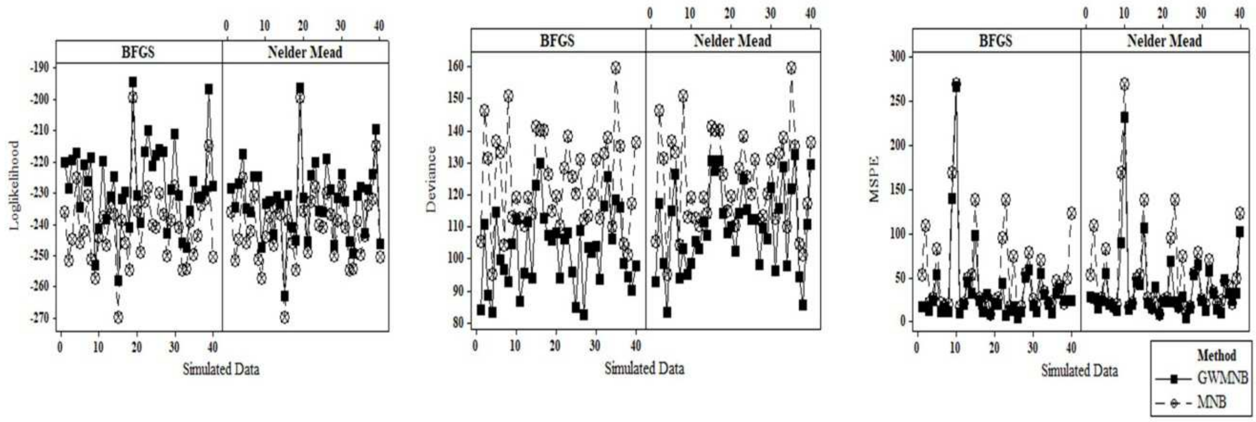


Fig. 2 Goodness of fit of GWMNB and MNB via BFGS and Nelder Mead in estimating the means of dependent variables

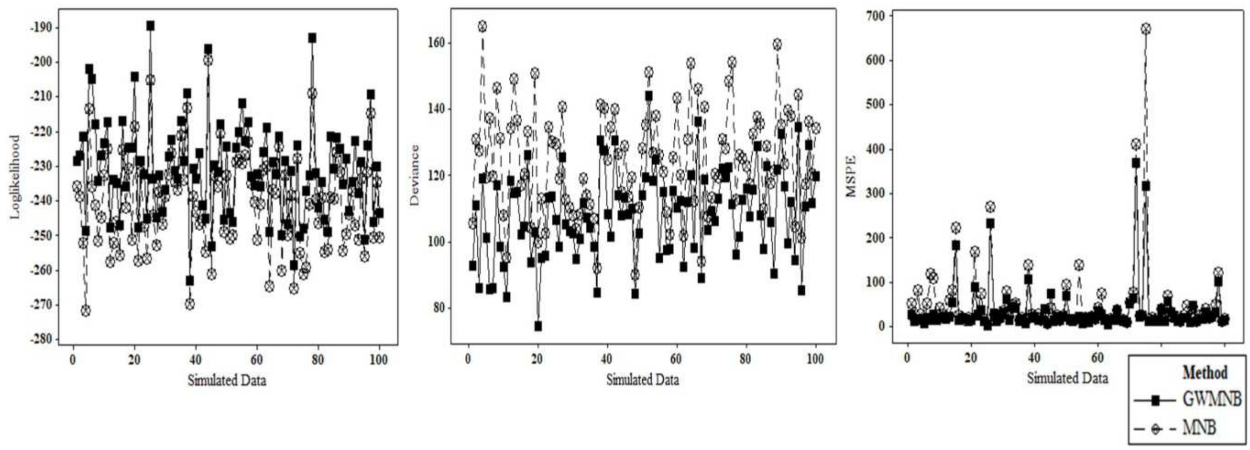


Fig. 3 The goodness of fit of estimating the mean of dependent variables using GWMNB and MNB via Nelder Mead algorithm

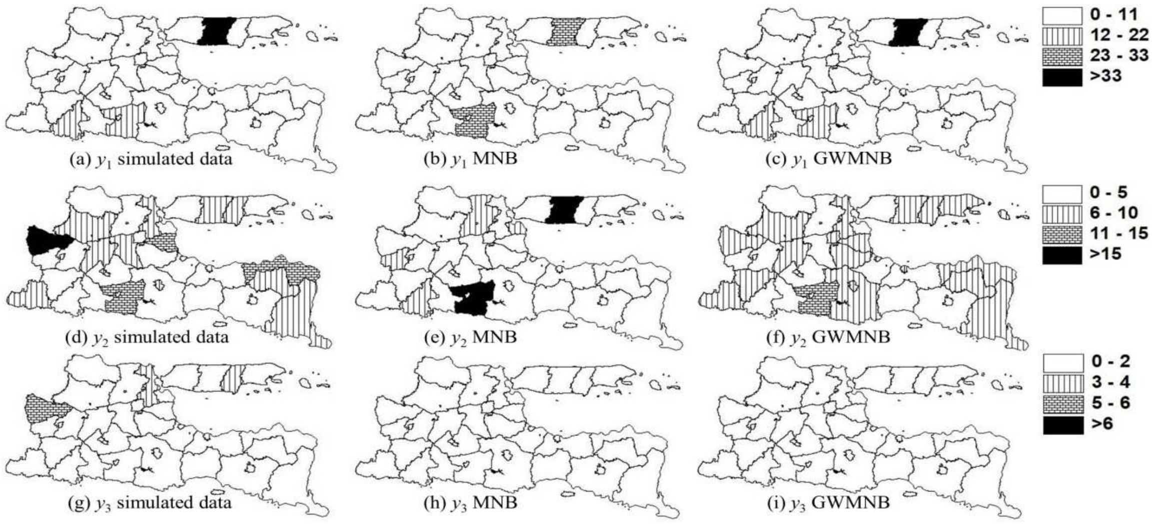


Fig. 4 The simulated data of dependent variables of multivariate spatial count data with over dispersion and their estimates using MNB and GWMNB methods

The independent variables $x_{i1}*$ and $x_{i2}*$ are initially simulated from a uniform (0,1) distribution, then the variable x_1 is updated to be spatially dependent by using the equation.

$$x_{i1} = \left(\frac{x_{i1}^{*} \left| u_i \right|}{1000} \right)^2 + \left(\frac{x_{i1}^{*} \left| v_i \right|}{90} \right)^2 \tag{12}$$

and x_2 is updated based on the equation

$$\mathbf{x}_2^{\phi} = \sin(\phi)\,\mathbf{x}_1 + \cos(\phi)\,\mathbf{x}_2 \tag{13}$$

where ϕ is the correlation level. In this research we set ϕ = 0.1.

By using those coefficients, index of dispersion and independent variables, we generate the dependent variables based on the procedure in [29] for generating negative binomial data, where the distribution of the dependent variables (negative binomial) come from Poisson and gamma distribution.

We use the coefficients and index of dispersion estimate of MNB as an initial estimate of GWMNB in estimating the means of dependent variables. Optimum bandwidth is gotten by using golden section algorithm. We make the replication of simulation 100 times.

The result of the simulation is given in Table 1. Although BFGS is a stable algorithm in MNB, it tends to be unsuccessful estimating the means in GWMNB. It is only 40% of 100 simulations, BFGS can be successful in estimating the means of multivariate spatial count data with over dispersion. However, the results of BFGS tend to have the slight better goodness of fit than them of the Nelder Mead algorithm. On the other hand, the Nelder Mead tends to be successful in estimating the means of dependent variables than BFGS algorithm. Nelder Mead is successful in estimating the means in all 100 simulations.

By using BFGS and Nelder Mead algorithms, the GWMNB tends to have the better goodness of fit than MNB method. Fig. 2 describes the goodness of fit of the methods for 40 replications of simulation where GWMNB using BFGS succeeds in estimating the means of dependent variables while Fig.3 describes the goodness of fit of the methods in estimating the means using Nelder Mead algorithm for all replications. Based on Table 1, Fig. 2 and Fig. 3, the log-likelihood value of GWMNB tends to be higher and its deviance tends to be smaller than them of MNB method. We use Mean Square Prediction Error (MSPE) to measure the distance between the means and simulated data.

TABLE I
GOODNESS OF FIT (IN AVERAGE) OF 100 SIMULATED DATA

	MNB		GWMNB	
Indicator	**BFGS**	**N.Mead**	**BFGS**	**N.Mead**
The percentage of the method success	100%	100%	40%	100%
Bandwidth			1.81	2.23
CV			4547.57	11934.04
Log likelihood	-240.18	-240.18	-227.56	-230.96
Deviance	124.52	124.52	102.74	108.41
MSPE	50.28	50.28	34.48	35.19

$$\text{MSPE} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \hat{\mu}_{ij} \right)^2$$

MSPE of GWMNB tends to be smaller than it of MNB method (see Table 1, Fig. 2 and Fig. 3). Fig.4 illustrates the comparison between the value of dependent variables of simulated data and their means of MNB and GWMNB methods using Nelder Mead algorithm. Based on Fig.4, the means of GWMNB and simulated data are more similar than the means of MNB and simulated data, as indication GWMNB estimates the means better than MNB.

MNB and GWMNB give the different results of $\boldsymbol{\beta}(u_i, v_i)$ and $\tau(u_i, v_i)$ estimates. It is caused by those of spatial data have the local character, depending on the location where the data are observed. As illustration, the coefficients and index of dispersion estimate of MNB and GWMNB for location 1 and 2 are presented in Table 2.

TABLE II
COEFFICIENTS AND INDEX OF DISPERSION ESTIMATE OF MNB AND GWMNB

Coefficients	MNB	GWMNB	
		1	2
β_{10}	1.059	1.184	1.142
β_{11}	0.263	0.048	0.138
β_{12}	1.917	1.854	1.846
β_{20}	1.891	2.171	2.149
β_{21}	-0.530	-0.750	-0.811
β_{22}	0.768	0.819	0.931
β_{30}	0.238	0.514	0.475
β_{31}	-0.905	-0.987	-0.805
β_{32}	0.495	0.310	0.266
τ	0.479	0.465	0.439
Loglikelihood	-229.178	-211.901	
Deviance	126.177	95.112	
MSPE	14.533	8.342	

B. Application

The mortality rate is one of the health status indicators in the society. The mortality rate of infant, child and maternal, the life expectancy, morbidity rate, and nutritional status are health status indicators in East Java Indonesia [30]. The accessibility and affordability of health services are the factors which influenced the mortality rate of infant and child [31]. The good quality health services can prevent high mortality rate.

We use the data from the East Java Provincial health office [30] as an application. The data consist of 38 regencies/towns in East Java Province, Indonesia. The dependent variables are the number of infant deaths (Y_1), toddler and preschool deaths (Y_2) and pregnant and childbirth mother (maternal) deaths (Y_3). There are seven independent variables, the percentage of villages that always take care their health (X_1), households that have healthy behavior (X_2), handling obstetric complications (X_3), prenatal visits to the health worker minimum four times (X_4), childbirth with health workers help (X_5), the ratio of mini-hospitals and the number of population in each regency (X_6) and the percentage of integrated health posts giving service actively (X_7). Exposure variables in this research are the population of infant (N_1), toddler and preschool (N_2), and maternal (N_3). The descriptive statistics of the data is given in Table 3.

TABLE III
DESCRIPTIVE STATISTICS OF MORTALITY DATA VARIABLES

Variable	Mean	Variance	Minimum	Median	Maximum
Y_1	137	6375.6	11	131	298
Y_2	11	146.8	0	8	70
Y_3	15	81.7	1	14	39
N_1	15226	9.70E+07	2066	14837	41581
N_2	63578	1.70E+09	8206	62586	176241
N_3	15955	1.10E+08	2154	15488	43822
X_1	96.2	52.7	65.4	100	100
X_2	46.3	209.5	20.1	43.3	68.7
X_3	92.3	96.4	64.6	94.6	105.8
X_4	88.3	32.5	75.4	88.8	98.2
X_5	92.2	17.8	83.6	91.6	101
X_6	0.3	0	0.1	0.3	0.4
X_7	67.9	295.3	34.2	69.7	96.3

The highest mortality rate of infant, toddler, and preschool and also maternal tend to happen in the east parts of East Java (Fig. 5). There are 428 deaths from 2415952 toddlers and preschool population, or the mortality rate of toddler and preschool is 0.018%. The mortality rate of infant and maternal are 0.902% and 0.094% respectively.

The mean value of Y_1, Y_2, and Y_3 are less than the variance. The Pearson correlation Y_1 and Y_2, Y_1 and Y_3, and Y_2 and Y_3 are 0.516, 0.841 and 0.643 respectively. Therefore we use the multivariate negative binomial model for estimating the means of dependent variables. We exclude the independent variable that has a high positive correlation with others for handling multicollinearity (Table 4). There is a high correlation, 0.868 between X_4 and X_5. Therefore we omit X_4 from data analysis to conquer the multicollinearity of independent variables. We select the independent variables by adding one by one of them into the model. The best combination for each number of independent variables is presented in Table 5. From that process, the model with two independent variables (X_1 and X_2) gives the best performance than others based on AIC indicator. Therefore we use those independent variables to build MNB and GWMNB.

TABLE IV
PEARSON CORRELATION OF MORTALITY DATA VARIABLES

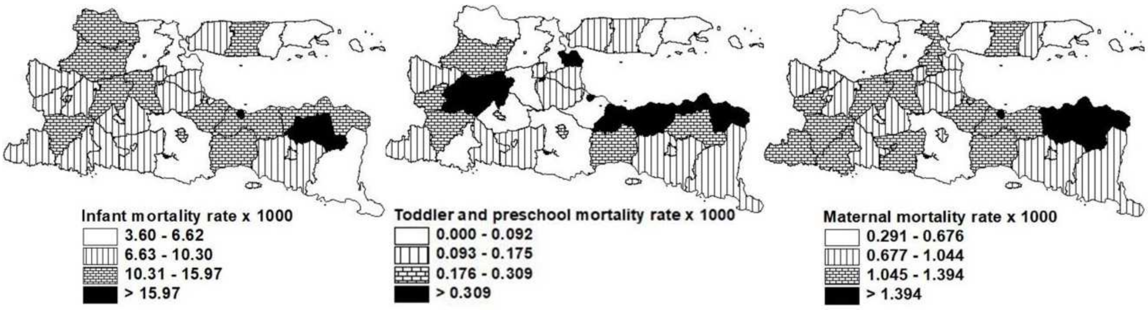
	Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4	X_5	X_6
Y_2	0.52								
Y_3	0.84	0.64							
X_1	0.15	-0.01	-0.03						
X_2	-0.04	0.14	0.05	0.23					
X_3	0.06	0.11	0.07	-0.16	0.02				
X_4	-0.03	0.07	-0.07	0.40	0.31	-0.20			
X_5	0.05	0.02	-0.11	0.39	0.21	-0.18	**0.87***		
X_6	-0.57	-0.27	-0.57	-0.01	0.19	0.21	0.01	-0.07	
X_7	-0.24	-0.16	-0.23	0.08	0.33	0.21	0.00	0.07	0.25

*: $|r| \geq 0.7$

TABLE V
INDEPENDENT VARIABLES SELECTION

Independent Variables	Loglikelihood	AIC
Model 1 : $\mathbf{X}(X_2)$	-473.5	964.7
Model 2 : $\mathbf{X}(X_1, X_2)$	-463.8	**955.8**
Model 3 : $\mathbf{X}(X_1, X_2, X_3)$	-459.2	959.6
Model 4 : $\mathbf{X}(X_1, X_2, X_3, X_7)$	-456.5	970.9
Model 5 : $\mathbf{X}(X_1, X_2, X_3, X_5, X_7)$	-454.0	988.1
Model 6 : $\mathbf{X}(X_1, X_2, X_3, X_5, X_6, X_7)$	-453.0	1017.4

The results of the analysis using MNB and GWMNB are given in Table 6 and Fig. 6. By using the BFGS algorithm, GWMNB fails in cross-validation process to find $\left(\boldsymbol{\beta}(u_i, v_i)\right)$ and $\left(\tau(u_i, v_i)\right)$ estimates for estimating the means. Cross-validation process using golden section with range bandwidth (0.6503076, 3.150016) give the result optimum bandwidth 3.150, CV= -. There is not CV resulted using MLE via the BFGS algorithm. This might be caused by BFGS is a derivative-based method. The estimation in one iteration is influenced by the derivative of the previous iteration, so that if in a certain iteration the condition in BFGS is not fulfilled then the BFGS fails to get the $\boldsymbol{\beta}(u_i, v_i)$ and $\tau(u_i, v_i)$ estimate for estimating the means of dependent variables.



(a) (b) (c)

Fig. 5 The description of the mortality rate of infant (a), toddler and preschool (b) and maternal (c).

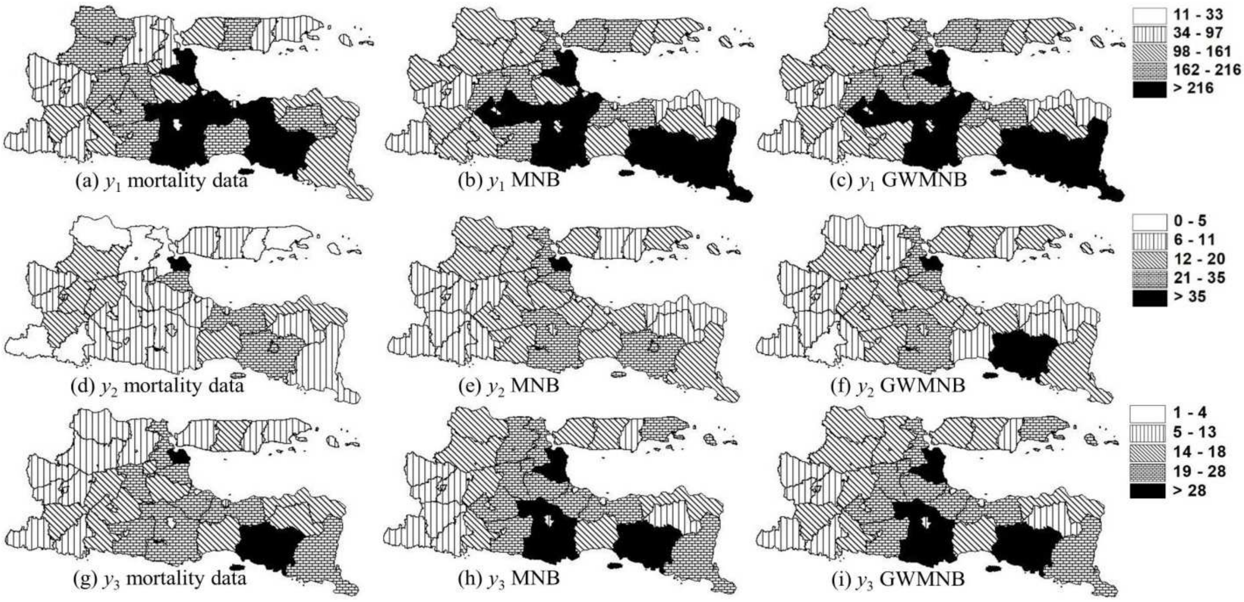


Fig. 6 The mortality data and their estimates using MNB and GWMNB methods

On the other hand, Nelder Mead (free derivative) algorithm can be used successfully estimating the means of dependent variables. GWMNB using the Nelder Mead algorithm gives better goodness of fit than others. It has a higher log likelihood, smaller deviance, and MSPE. The means of dependent variables from GWMNB tend to be more similar to the observations than the means from MNB (Fig. 6).

TABLE VI
THE GOODNESS OF FIT MNB AND GWMNB

	MNB		GWMNB	
Indicator	**BFGS**	**Nelder Mead**	**BFGS**	**Nelder Mead**
Bandwidth			3.15	1.61
CV			-	148008.40
Loglikelihood	-463.79	-463.83	-460.92	-443.36
Deviance	259.15	259.09	252.56	217.78
MSPE	1107.02	1093.30	1109.28	1069.36

The same results as simulation study are shown by the results of mortality data analysis, $\boldsymbol{\beta}(u_i,v_i)$ and $\tau(u_i,v_i)$ estimates of GWMNB vary depending on the location observed. In Table 7, we present the results of those estimate for the two regencies of East Java province, those are Jombang and Madiun regencies.

TABLE VII
COEFFICIENTS AND INDEX OF DISPERSION ESTIMATE OF MNB AND GWMNB FOR MORTALITY DATA

Coefficients	MNB	GWMNB	
		Jombang	**Madiun**
β_{10}	-4.864	-2.384	-4.417
β_{11}	0.005	-0.020	-0.001
β_{12}	-0.005	-0.006	-0.003
β_{20}	-7.166	-8.273	-8.119
β_{21}	-0.019	-0.020	-0.021
β_{22}	0.009	0.026	0.030
β_{30}	-5.201	-5.267	-5.653
β_{31}	-0.018	-0.015	-0.012
β_{32}	0.001	-0.005	-0.003
τ	0.136	0.080	0.090
Deviance	259.091	217.783	
MSPE	1093.303	1069.364	

IV. CONCLUSIONS

In this research, we derive estimating the means of dependent variables of multivariate spatial count data with overdispersion using Geographically Weighted Multivariate Negative Binomial (GWMNB) and multivariate negative binomial (MNB) methods. We use MLE via two algorithms, based on derivative (BFGS) and free derivative (Nelder Mead) algorithms. Based on the goodness of fit, GWMNB performs better than MNB for estimating the means of dependent variables in the existence of spatial heterogeneity in the data.

The Nelder Mead tends to be successful rather than BFGS algorithm in estimating the means of dependent variables count data with overdispersion. BFGS algorithm tends to have false convergence in estimating the means of them. GWMNB is complex and very time consuming, especially for the simulation study. Nelder mead is more robust related to the initial value, but it is slower than the BFGS algorithm. Therefore, this research encourages to find out the fast and robust algorithms for future studies of geographically weighted multivariate. The Bayesian or modification of BFGS can be alternative methods for the next studies. The highest deaths of the infant, toddler, and preschool and also maternal tend to happen in east part of East Java province. The mortality rate of the infant is larger than it of toddler and preschool and also maternal.

ACKNOWLEDGMENT

The authors thank East Java Provincial Health Department for providing access to the mortality data. We are grateful to the referees for careful checking of the details and for helpful comments which improve this paper.
*Corresponding author : *purhadi@statistika.its.ac.id*

REFERENCES

[1] S. Gurmu and P. K. Trivedi, *Recent developments in models of event counts: a survey*. University of Virginia, Thomas Jefferson Center for Political Economy, 1994.

[2] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*, vol. 53. Cambridge university press, 2013.

[3] R. E. Society and T. E. Journal, "Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application Author (s): Murat K. Munkin and Pravin K. Trivedi Published by : Wiley on behalf of the Royal Economic Society Stable URL : http://www.jstor.org," vol. 2, no. 1, pp. 29–48, 2017.

[4] D. Karlis and I. Ntzoufras, "Bivariate Poisson and Diagonal Inflated Bivariate Poissn Regression Models in R," *J. Stat. Softw.*, vol. 30, no. April, pp. 1–3, 2009.

[5] H. Zamani, P. Faroughi, and N. Ismail, "Bivariate generalized Poisson regression model : applications on health care data," *Empir. Econ.*, 2016.

[6] D. Karlis and L. Meligkotsidou, "Multivariate Poisson regression with covariance structure," *Stat. Comput.*, vol. 15, no. 4, pp. 255–265, 2005.

[7] F. Famoye, "A multivariate generalized poisson regression model," *Commun. Stat. - Theory Methods*, vol. 44, no. 3, pp. 497–511, 2015.

[8] F. Famoye, "On the bivariate negative binomial regression model," *J. Appl. Stat.*, vol. 37, no. 6, pp. 969–981, 2010.

[9] Y. S. Dewi, Purhadi, Sutikno and S. W. Purnami, "Comparison of Bivariate Negative Binomial Regression Models for Handling Over dispersion," *Int. J. Appl. Math. Stat.*, vol. 56, no. 5, pp. 53–62, 2017.

[10] R. Winkelman, "Seemingly unrelated negative binomial regression," *Oxford Bull. Econ. Stat.*, vol. 62, no. 4, pp. 553–560, 2000.

[11] P. Shi and E. A. Valdez, "Multivariate negative binomial models for insurance claim counts," *Insur. Math. Econ.*, vol. 55, no. 1, pp. 18–29, 2014.

[12] I. L. Solis-Trapala and V. T. Farewell, "Regression analysis of overdispersed correlated count data with subject specific covariates," *Stat. Med.*, vol. 24, no. 16, pp. 2557–2575, 2005.

[13] L. Anselin, "Spatial econometrics," *A companion to Theor. Econom.*, pp. 310–330, 2001.

[14] A. R. da Silva and T. C. V. Rodrigues, "Geographically Weighted Negative Binomial Regression-incorporating overdispersion," *Stat. Comput.*, vol. 24, no. 5, pp. 769–783, 2014.

[15] Purhadi, Y. S. Dewi, and L. Amaliana, "Zero Inflated Poisson and Geographically Weighted Zero- Inflated Poisson Regression Model: Application to Elephantiasis (Filariasis) Counts Data," *J. Math. Stat.*, vol. 11, no. 2, pp. 52–60, 2015.

[16] J. Nocedal and S. Wright, *Numerical Optimization*. Springer New York, 2006.

[17] J. C. Nash, *Nonlinear Parameter Optimization Using R Tools*. John Wiley and Sons Ltd, 2014.

[18] P. Harris, A. S. Fotheringham, and S. Juggins, "Robust Geographically Weighted Regression : A," *Ann. Assoc. Am. Geogr.*, vol. 100, no. December 2014, pp. 37–41, 2010.

[19] A. Páez, S. Farber, and D. Wheeler, "A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships," *Environ. Plan. A*, vol. 43, no. 12, pp. 2992–3010, 2011.

[20] R. Winkelmann, *Econometric Analysis of Count Data*, 5th ed. Springer Publishing Company, Incorporated, 2008.

[21] A. S. Fotheringham, C. Brunsdon, and M. Charlton, "Geographically Weighted Regression-The Analysis of Spatially Varying Relationships." Chichester, UK: John Wiley & Sons, 2002.

[22] D.-H. Li and M. Fukushima, "On the Global Convergence of the BFGS Method for Nonconvex Unconstrained Optimization Problems," *SIAM J. Optim.*, vol. 11, no. 4, pp. 1054–1064, 2001.

[23] J. F. Lawless, "Negative binomial and mixed poisson regression," *Can. J. Stat.*, vol. 15, no. 3, pp. 209–225, 1987.

[24] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Math. Comput.*, vol. 19, no. 92, pp. 577–593, 1965.

[25] J. C. Nash, *Compact numerical methods for computers: linear algebra and function minimisation*. CRC press, 1990.

[26] J. C. Nash, "On best practice optimization methods in R," *J. Stat. Softw.*, vol. 60, no. 2, pp. 1–14, 2014.

[27] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[28] C. G. Small and J. Wang, *Numerical methods for nonlinear estimating equations*, vol. 29. Oxford University Press on Demand, 2003.

[29] J. M. Hilbe, *Negative Binomial Regression*, Second. Cambridge: Cambridge University Press, 2011.

[30] Dinas Kesehatan Provinsi Jawa Timur, "Profil Kesehatan Provinsi Jawa Timur 2014," 2015.

[31] Bappenas, *Faktor-faktor yang Mempengaruhi Kelangsungan Hidup Anak*. Badan Perencanaan Pembangunan Daerah, 2009.