



Distance and Areas Weighting of GWR Kriging for Stunting Cases In East Java

Deby Ardianti¹, Henny Pramoedyo², Nurjannah³

^{1,2,3} Departement of Statistics Faculty of Mathematic and Natural Sciences, Brawijaya University, Indonesia

Email: ard.dianti@gmail.com, hennyp@ub.ac.id, nj_anna@ub.ac.id

ABSTRACT

Spatial heterogeneity shows the characteristic location from one location to others location and it is the main assumption in Geographically Weighted Regression. The location becomes a weight on GWR model, there are two groups of location weight namely based on distance and area. The weight considers the closeness between the location. The accuracy weighted is needed because the weighting represents the data location. The aim of this research was to get a suitable weighting method for stunting data. This research used secondary data about stunting and the influence factors of stunting such as coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, immunization coverage, and clean & health behaviour. Those data obtained from the Healthy Ministry of East Jawa. Based on the results of this research show that the goodness weighting for GWR model is Adaptive Bisquare Kernel (distance weighting). The predicted mapping stunting is showed by interpolation Kriging with a range of 27% to 49,5%.

Keywords: GWR; Weighting; Stunting

INTRODUCTION

Multiple analysis regression model was introduced in 1886 by Sir Francis Galton [1], is it called global model because it does not involve the location model. Multiple regression model had the assumed normal distribution, no multicollinearity, no autocorrelation and homogeneous in variance [2]. But in fact, it often the characteristic of location influence the model/spatial heterogeneity. Thus multiple analysis regression is not suitable for data that has spatial heterogeneity. The statistical method that takes into account this spatial aspect is Geographically Weighted Regression (GWR) [3]. Geographically Weighted Regression (GWR) was introduced by Brunson et al. (1996), to study the potential for relationships in a regression model to vary in geographic space, or what is called parametric non-stationarity[3]. In this method, each parameter is calculated at each location so that each observation location has different regression parameters [4].

The weights used in the GWR model are used to describe the closeness of the relationship between regions. The accuracy of the weighting method is needed because the weighted value will represent the location of the origin of the data [5]. Weights are grouped into two, namely based on distance (distance) and based on the area (contiguity) [6]. In this study, the distance weighting used is the Adaptive Bisquare

Kernel, which is a kernel weighting function assumes that the bandwidth adapts itself in size relatives to variations in the density of the data [7]. Meanwhile, the weighting based on area (contiguity) uses the queen contiguity weight, which can represent the intersection of the sides and the intersection of the corners [8].

The GWR model is used to predict at each observation location, and observations outside the study sample are not well predicted by the GWR model [9]. To solve this problem, the Kriging method will be used, which is a geostatistical technique used to predict and interpolate data at unsampled locations [10]. Some of research carried out by GWR and Kriging [11][12][13]. This study will use the GWR Kriging Spatial Interpolation method to analyze the factors that affect stunting in East Java with a distance and area weighted matrix. Stunting is a condition in which toddlers have a length or height that is less than their age [14]. Stunting is influenced by several factors such as coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, immunization coverage, and clean & health behaviour [15].

METHODS

Data in this research were data in East Java Health Profile Publication Book in 2018. This data consist of one dependent variable and five independent variable. In this research, there is 38 district of East Java province. The data analyzed by Interpolation GWR Kriging. The first step is testing the spatial heterogeneity that conducted by Breusch Pagan Test. The next step is to determine the distance and area weighting that calculated with equation 1:

The distance weighting (Adaptive Bisquare Kernel) [4]

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b_i} \right)^2 \right]^2, & \text{if } d_{ij} < b \\ 0, & \text{if } d_{ij} \geq b \end{cases} \quad (1)$$

With [4]

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$$

And b is the optimum bandwidth based on *Cross-Validation* (CV) method [4]

$$CV(b) = \sum_{i=1}^n (y_i - \hat{y}_i(b))^2$$

The area weighting (Queen Contiguity) [16]

$$c_{ij} = \begin{cases} 1, & \text{if region } i \text{ dan } j \text{ intersect} \\ 0, & \text{another} \end{cases} \quad (2)$$

With the standard form of Queen Contiguity [16]

$$w_{ij(IDM)} = \begin{cases} \frac{c_{ij}}{\sum_j c_{ij}}, & \text{untuk } i \neq j \\ 0, & \text{untuk } i = j \end{cases}$$

The next step is estimating the GWR model parameter using the formula [1]

$$\hat{\beta}(u_i, v_i) = [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \quad (3)$$

And do the simultaneous and partial hypothesis test for GWR model. The best model between two weighting is chosen by R^2 . The predictive mapping of the best model is described by Kriging with exponential semivariogram. The research data analysis process was carried out using R and Arcgis Software.

RESULTS AND DISCUSSION

Heterogeneity Spatial with Breuch Pagan test are presented in Table 1

Table 1. P value of Heterogenity Spatial Test

Testing	<i>p-value</i>	
	Distance Weighting (Adaptive Bisquare Kernel)	Area Weighting (Queen Contiguity)
Heterogenity Spatial (Breuch Pagan)	0.000174	0,000488

Based on Table 1, *P* value for two weighting are 0,000174 and 0,00488 less than $\alpha = 0,05$, so reject H_0 and the conclusion is that there is heterogeneity spatial in stunting cases data. Next step is the simultaneous and partial hypothesis test of the parameter GWR model. Table 2 shows the result of simultaneous of GWR model for both weight.

Table 2. Simultaneously testing

GWR Model	<i>F</i>	<i>F_{table}</i>
GWR Model with distance weighting	2,4875	2,06
GWR Model with area weighting	0,5362	2,85

Table 2 shows that the statistic test $F = 2,4875 > F_{(0,1;19,431;12,569)} = 2,06$, so reject H_0 , the conclusion is the distance weight influences estimates parameter of GWR model and all predictor variable affecting simultaneously with response variable. While the area weighting show that the statistic test $F = 0,5362 < F_{(0,05;0,09;37,3)} = 2,85$, so accept H_0 , the conclusion is area weighting has not Influence in estimates parameter of GWR model and at least one predictor variables not affecting simultaneously with response variable.

Table 3. District Grouping with distance weighting of GWR Model

District	Significant Variable
Bangkalan, Gresik, Sampang, Kota Pasuruan City, Surabaya City	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding
Banyuwangi	consumption of FE tablet, exclusive of breastfeeding, immunization coverage
Blitar, Lamongan, Mojokerto, Pasuruan, Sidoarjo, Tuban, Tulungagung, Mojokerto City	exclusive of breasfeeding
Bojonegoro,Nganjuk	consumption of FE tablet coverage visiting of pregnant women (K1), exclusive of breastfeeding, immunization coverage, and clean & health behaviour
Bondowoso, Jember	
Madiun, Magetan, Ngawi, Situbondo, Pacitan, Ponorogo, Trenggalek, Madiun City	consumption of FE tablet, exclusive of breasfeeding

District	Significant Variable
Pamekasan	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, and clean & health behaviour
Probolinggo, Probolinggo City	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, immunization coverage, and clean & health behaviour
Sumenep	coverage visiting of pregnant women (K1), exclusive of breastfeeding, and clean & health behaviour
Jombang, Kediri, Lumajang, Malang, Batu City, Blitar City, Kediri City, Malang City	No Significant Variable

The partial test of GWR model with distance weighting (Adaptive Bisquare kernel) presented in Table 3. 10 groups were obtained shows that there are differences a significant variables that influence prevalence stunting. In comparison the partial test of GWR model with area weighting (Queen Contiguity) are presented in Table 4.

Table 4. District Grouping of area weighting of GWR Model

District	Variable significant
Bangkalan, Banyuwangi, Bondowoso, Jember, Kediri, Lamongan, Lumajang, Madiun, Magetan, Malang, Mojokerto, Nganjuk, Ngawi, Pamekasan, Pacitan, Trenggalek, Pasuruan	consumption of FE tablet, exclusive of breastfeeding, clean & health behaviour
Blitar, Sumenep	consumption of FE tablet, exclusive of breastfeeding, immunization coverage, coverage visiting of pregnant women
Bojonegoro, Gresik, Probolinggo, Sampang, Sidoarjo, Situbondo	(K1), consumption of FE tablet, exclusive of breastfeeding, immunization coverage, and clean & health behaviour
Jombang, Pasuruan, Ponorogo, Kediri, City, Malang City, Mojokerto City, Surabaya City	consumption of FE tablet, exclusive of breastfeeding
Tuban, Tulungagung	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, clean & health behaviour
Batu City	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding
Blitar City, Madiun City	coverage visiting of pregnant women (K1), consumption of FE tablet, exclusive of breastfeeding, immunization coverage

Examination of the accuracy GWR Model was done by calculating R^2 values in GWR model with distance and are weighting. The higher R^2 value the better the accuracy of the model. Following is the examination of the accuracy GWR Model.

Table 5. Comparison Model

GWR Model	R^2
GWR Model with distance weighting	0,97
GWR Model with area weighting	0,32

Based on the result in Table 5 , GWR model with distance weighting has a higher R^2 than GWR model with area weighting. So, it can concluded that distance weighting is better used to form GWR model for stunting cases data.

Then, the value of Y prediction of GWR model of distance weighting will use to form to mapping GWR Kriging with the exponential semivariogram. The mapping of stunting prediction presented in Figure 1.

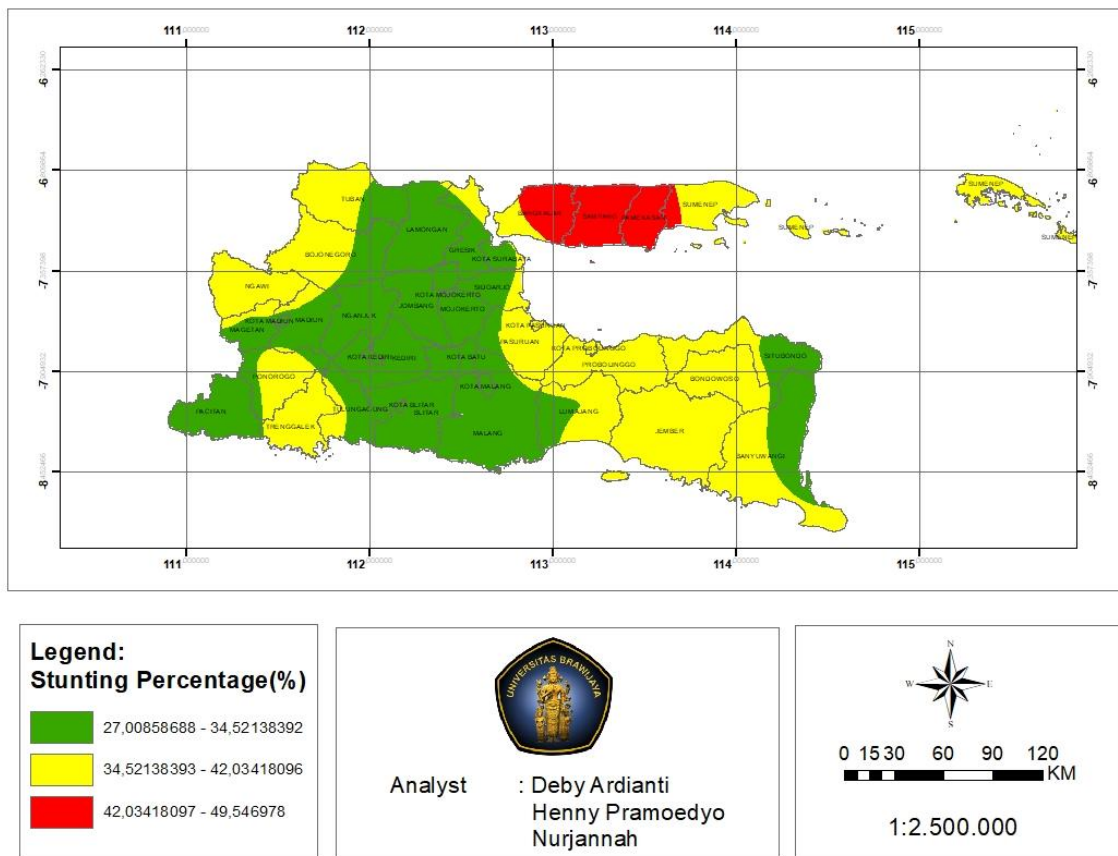


Figure 1. Prediction Map of Distribution Stunting in East Java

Based on the predicted map in Figure 1 shows in 2019 the percentage stunting predicted high marked on that map in red, which is Bangkalan, Sampang, and Pamekasan District. On the contrary, in the regional section the middle namely Nganjuk District has a percentage of stunting low enough that marked on the map in green. The predicted value of prevelance stunting by using KWing GWR interpolation ranging from 27.008% to 49.547%.

CONCLUSIONS

GWR model with distance weighting (Adaptive Bisquare Kernel) produces a better accuracy model in terms of high R^2 values compared to GWR model with area weighting (Queen Contiguity). The predicted mapping stunting is showed by interpolation Kriging with a range of 27% to 49,5%, which Sampang has a high percentage of stunting. The predicted model approach the actual data in the field because it has included location weighting in variables.

REFERENCES

- [1] N. R. Draper and H. Smith, *Applied regression analysis*, vol. 326. John Wiley & Sons, 1998.
- [2] D. N. Gujarati, *BASIC ECONOMETRICS*, FOURTH. New York: McGraw-Hill Companies, INC, 2003.
- [3] M. Fischer and A. Getis, *Handbook of Applied Spatial Analysis*. New York: Springer, 2010.
- [4] A. S. Fotheringham, C. Brunson, and M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.
- [5] C. Chasco, "Modeling spatial variations in household disposable income with Geographically Weighted Regression," no. January, 2007.
- [6] L. Anselin, I. Syabri, and Y. Kho, "GeoDa : An Introduction to Spatial Data Analysis," vol. 38, pp. 5–22, 2006.
- [7] J. A. Yacim and D. G. B. Boshoff, "A Comparison of bandwidth and kernel function selection in geographically weighted regression for house valuation," *Architecture*, vol. 10, no. 1, 2019.
- [8] X. Li, W. Li, L. Anselin, S. Rey, and J. Koschinsky, "A MapReduce algorithm to create contiguity weights for spatial analysis of big data," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, 2014, pp. 50–53.
- [9] J. Walter and W. L. Jeremy, "Local and global approaches to spatial data analysis in ecology Spatial autocorrelation," pp. 97–98, 2005.
- [10] M. Armstrong, *Basic linear geostatistics*. Springer Science & Business Media, 1998.
- [11] B. Selby and K. M. Kockelman, "Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression," *J. Transp. Geogr.*, vol. 29, pp. 24–32, 2013.
- [12] C. Zeng *et al.*, "Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method," *Geoderma*, vol. 281, pp. 69–82, 2016.
- [13] S. Kumar, R. Lal, and D. Liu, "A geographically weighted regression kriging approach for mapping soil organic carbon stock," *Geoderma*, vol. 189, pp. 627–634, 2012.
- [14] Kementerian Kesehatan RI, *Situasi Balita Pendek (Stunting) di Indonesia*. Jakarta: Kemenkes RI, 2019.
- [15] TNP2K, *100 Kabupaten/Kota Prioritas Untuk Intervensi Anak Kerdil (Stunting)*. Jakarta: TNP2K RI, 2017.
- [16] H. Pramoedyo, *Analisis Spasial Dasar*. Malang: Universitas Negeri Malang Press, 2017.