

Estimation of Percentage on Malnutrition Occurrences in East Java using Geographically Weighted Regression Model

Ida Mariati Hutabarat^{1,2*}, Asep Saefuddin², Hardinsyah³, Anik Djuraidah²

1. Department of Mathematics, Faculty of Mathematics and Sciences, Universitas Cenderawasih, Jayapura 99358, Indonesia
2. Department of Statistic, Faculty of Mathematics and Sciences, Institut Pertanian Bogor, Bogor 16680, Indonesia
3. Department of Public Health, Faculty of Human Ecology, Institut Pertanian Bogor, Bogor 16680, Indonesia

*e-mail: ida_mariati@yahoo.com

Abstract

The Province of East Java has its characteristics that differentiate it from any other regions. Dissimilarities in characteristics of a region may encompass issues such as social, economic, cultural, parenting, education, and the environment, so as to cause the difference in the case of severe under nutrition between one region to another. Sufferers of malnutrition in one region may be linked and influenced by the surrounding regions. Therefore, we need a statistical modeling that can take into account the spatial factor. Statistical methods that can be used to analyze the data and also takes into account the spatial factor are the Geographically Weighted Regression (GWR). This study is aimed to determine the case of malnutrition models in East Java Province using GWR model with kernel adaptive bi-square weighting and comparing it to the conventional linear regression model. The data used in the study are secondary data obtained from the National Socio-Economic Survey and Basic Health Research (2010) conducted in 38 districts in East Java. Estimation is done by using the Weighted Least Squares method that provides different weighting values to each region. The result showed that there are 38 models of the malnutrition case that is different from each district in East Java. The GWR model with bi-square kernel weighting function is better in modelling the case of malnutrition in East Java compared to the conventional linear regression models that are based on the criteria of goodness that is the R-square, Mean Square Error and the Akaike Information Criterion.

Abstrak

Pendugaan Persentase Kejadian Gizi Buruk di Jawa Timur Menggunakan Model Regresi Terboboti Geografis. Provinsi Jawa Timur memiliki karakteristik wilayah yang berbeda. Perbedaan karakteristik wilayah dapat berupa masalah sosial, ekonomi, budaya, pola asuh, pendidikan, dan lingkungan, sehingga dapat menyebabkan adanya perbedaan kejadian balita gizi buruk antara wilayah satu dengan wilayah lainnya. Penderita gizi buruk dari satu wilayah diduga dipengaruhi oleh wilayah sekitarnya. Oleh karena itu, diperlukan suatu pemodelan statistik dengan memperhitungkan faktor spasial. Metode statistik yang dapat digunakan untuk menganalisis data dengan memperhitungkan faktor spasial adalah regresi terboboti geografis (RTG). Penelitian ini bertujuan untuk menentukan model kejadian gizi buruk di Provinsi Jawa Timur menggunakan model RTG dengan pembobot kernel adaptif kuadrat ganda dan membandingkan model RTG dengan model regresi linear konvensional. Data yang digunakan adalah data sekunder dari survei sosial ekonomi nasional (SUSENAS) dan Riset Kesehatan Dasar (RISKESDA) 2010 pada 38 kabupaten/kota di Jawa Timur. Pendugaan parameter model dilakukan dengan menggunakan metode *weighted least squares* (WLS) yaitu dengan memberikan pembobot yang berbeda untuk setiap wilayah. Hasil penelitian diperoleh terdapat 38 model kejadian gizi buruk yang berbeda untuk setiap kabupaten/kota di Jawa Timur. Model RTG dengan menggunakan pembobot fungsi kernel kuadrat ganda lebih baik digunakan untuk memodelkan kejadian gizi buruk di Jawa Timur dibandingkan dengan regresi linear konvensional berdasarkan kriteria kebaikan model yaitu *R-square*, *mean square error* (MSE) dan *Akaike information criterion* (AIC).

Keywords: geographical weighted regression, malnutrition, weighted least square

Introduction

The number of people suffering from malnutrition is still quite high. In the year 2010, the figure reached 17.9

percent. Since 2015, the Government has taken the efforts to lower the aforementioned figure down to 15.1 percent in line with the target laid down by the Millennium Development Goals (MDGs). Data from *Badan Penelitian*

dan Pengembangan Kesehatan (Health Development and Research Board) revealed that some provinces have quite a high number of malnutrition cases. Among the provinces in the island of Java, the Province of East Java takes the lead with 14,720 cases with the highest prevalence of malnutrition at 4.8 percent.¹

Some endeavors have been taken by the Government, particularly those carried out the Health Department, by employing both strategic and tactical approaches. A strategic approach entails the efforts to optimize the health service operations for pregnant mothers and children under the age of five years old, including, among others, optimizing the function of *Posyandu* (integrated health service station, i.e., centers for pre- and postnatal health care and information for women and for children under the age of five). Whereas, a tactical approach covers the efforts to anticipate the rising prevalence among children under the age of five years old who suffer from malnutrition and measures to reduce such an occurrence through some studies or research about malnutrition cases in children under the age of five years old.

Malnutrition sufferers from any specific region may be influenced by the cases occurring in the surrounding regions. Obviously, aspects of geographic, socio-cultural and economic conditions will be different from one region to another; therefore, it requires a specific statistical modeling that takes into account the spatial factor. Literature related to spatial issues have been developed many times before, including spatial analysis of variance,² a structural change model for spatial discrete data,³ spatial adaptive filtering model,⁴ which is used for continuous spatial data between their spaces. The last model is a development of the non-parametric regression model, wherein the estimation for each region will be calculated using the calibration model. In response to the difficulties when using calibration model when estimating the parameters,⁵ a model for non-stationer spatial data is developed, this model is known as the Geographically Weighted Regression (GWR).

The GWR method is relatively simpler in its operation and is more effective than any other methods.⁶ Case in point, in the health sector, numerous studies have been carried out utilizing the GWR method,⁷ using the GWR model to reveal the factors that influence the level of food poverty. Using the GWR method to uncover the correlation between the level of poverty and economic growth and government's policy.⁸ The weighting matrices used in GWR are based on the position or distances between one region of observation to another observation location. The closer the proximity between the regions the stronger the influence will be.

The GWR model is a development of the global regression model, in which the basic ideas are taken

from the non-parametric regression model.⁹ This model is a locally linear regression that produces local parameter estimation for any specific point or region from where the data is collected. The GWR parameters are calculated for each and every region observed. Therefore, each region will have a different parameter values. The dependent variable in GWR model is a continuous random variable, which is predicted using independent variables wherein each of its regression coefficients is dependent on the region where the data is observed. Unlike the conventional linear regression, the parameter values are assumed to be the same for each point or region under observation. Therefore, the resulting parameter estimation will have a single value that valid for each region observed.

The province of East Java has its unique regional characteristics, resulting in dissimilarities between the occurrence of malnutrition among children under the age of years old in one region to the other. Dissimilarities in characteristics of a region may encompass issues such as social, economic, cultural, parenting, education, and the environment. The location where people live, whether it is in the city, village or rural or coastal areas will influence their thought pattern, culture and their socio-economic perception. The geographic factor will also contribute to their mobility and access toward food, nutrition services as well as health services both primary and referral. Consequently, sufferers of malnutrition are variables with spatial effect. For that reason, they require a statistical modeling that can take into account the spatial factor. The aim of this study is to determine the model for the occurrence of malnutrition in the province of East Java using Geographically Weighted Regression model with kernel adaptive bi-square weighting.

Methods

The data used in the study are secondary data collected by BPS (*Badan Pusat Statistik*) (Central Bureau of Statistics), i.e. the data obtained from the National Socio-Economic Survey (*Survei Sosial Ekonomi Nasional/ SUSENAS*) and the Basic Health Research (*Riset Kesehatan Dasar*) conducted in 38 districts/cities in the East Java in 2010.

Also, the study also employs data on astronomical positions (position of latitude and longitude). GWR method requires data on coordinate's points of the region observed (latitude and longitude). These coordinates are used to determine the distance from one point of observation to the next. In this study, the observation unit is the districts/cities in the province of East Java. Data collection was carried out from July to August 2010 of the households selected as samples. The samples consisted of 29,952 households spread all over the districts/cities in East Java, ranging between 640 to 1,120 households per district/city. The sampling technique used in the study was the

probability proportional to size (PPS)–Systematic, wherein the size of the total households were taken from the listing of each block from the result of voters registration and continuous population registration P4B (*Pendaftaran Pemilih dan Pendataan Penduduk Berkelanjutan*). The data was collected through face-to-face interviews with surveyor officials and respondents, i.e. head of the households, husband/ wife who head the household or member of the households who were familiar with the characteristics of questions. Based on the census population carried out by the East Java BPS in 2010, the total number of population in East Java were 39 million.¹⁰

The variables used in this study are based on the study.¹¹ These variables include the percentage of children under the age of five years old suffering from malnutrition as dependent variable (Y), percentage of households with access to clean water (X_1), percentage of poor people (X_2), percentage of children under the age of five years old suffering from infections (X_3), percentage of expenses per capita for food in a month (X_4), percentage of the total *Posyandu* (X_5), percentage of children under the age five years old who are breastfed exclusively (X_6).

Stages of data analysis in acquiring the GWR model for modelling the percentage of children under the age five years old suffering from malnutrition are as follows: a) Description of data on children under the age five years old suffering from malnutrition in the Province of East Java; b) Obtaining the linear regression model between the dependent and independent variables using the OLS method (ordinary least square); c) Performing linear parameter test simultaneously using ANOVA and partial test using t- test; d) Performing heterogeneity test using Breusch-Pagan (BP) test; e) Determining up and v_i according to the south latitude and east longitude for each district/city in East Java province; f) Calculating the Euclidian distance between the regions observed based on their geographical positions. The Euclidian distance between regions i that are located in the coordinate (u_i, v_i) against the regions j that are located in the coordinate (u_j, v_j) . These calculations are carried out for the entire regions of observation; g) Determining the optimum bandwidth according to the criteria of minimum CV values; h) Calculating the weighting matrices using the i bi-square kernel. Esti-mating the GWR model parameters using the optimum bandwidth; i) Performing Goodness of Fit test on the GWR model; j) Performing parameter significance test; k) Obtaining the GWR model for the occurrence of malnutrition on children under the age five years old; l) Comparing the results obtained from the linear regression and GWR with R-Square criteria (R^2), mean squares error (MSE) and Akaike information criterion (AIC).

The method used for parameter estimation model is the Weighted Least Squares (WLS) by assigning different

weight for each region. The analysis is performed using R.3.0.2 program.

Results and Discussion

In general, the region of East Java can be divided into two major areas, the mainland East Java and the Island of Madura. The area of mainland East Java covers almost 90% of the entire area of East Java Province while the remaining 10% is the area of Madura Island. Administratively the province of East Java consisted of 29 districts and 9 cities.¹²

The average percentage of children under the age of five years old suffering from malnutrition in East Java in the year 2010 is 4.48%, wherein the lowest percentage of children under the age of five years old suffering from malnutrition is in the Municipality of Madiun at 1%, whereas the highest number of children under the age of five years old suffering from malnutrition is in the District of Sampang at 16.2%. Based on their distribution pattern, the occurrence of children under the age of five years old suffering from malnutrition is spread between districts/cities. Categorically, the highest percentage (more than 9%) of children under the age of five years old suffering from malnutrition occurs in the District of Jember and Sampang. The high number of poor households and the lack of access to *Posyandu* may have been the cause for the high numbers of children under the age of five years old suffering from malnutrition in the District of Sampang.

The percentage of populations who are poor is calculated based on the ability to fulfill their basic needs (basic needs approach). In this approach, poverty is seen as the inability as seen from the economic aspect to fulfill the basic needs to get food instead of measuring the need for food as seen from the spending aspect. Members of the population whose monthly per capita expenses are below the Poverty Line (PL) are categorized as poor. According to the 2010 survey of the National Socio-Economic Survey in the Province of East Java, the District(s) of Sampang and Jember are two regions whose per capita spending is lower than the Poverty Line (PL), i.e. 0.16% and 0.22% respectively. Further it could be discerned that in the District of Sampang, the percentage of the poor population is in line with the percentage of children under the age of five suffering from malnutrition.

The linear regression model developed using six independent variables are as follows:

$$Y = -4.96 - 0.04X_1 + 0.18X_2 + 0.14X_3 + 0.10X_4 + 0.21X_5 - 0.05X_6$$

The above model explains that the occurrence of children under the age of five suffering from malnutrition will be reduced by 0.04% if the variable of households that can access clean water (X_1) increase by

one percent on the condition that the other variables remain constant. The occurrence of children under the age of five suffering from malnutrition will also decrease by 0.05% if the number of children under the age five years old who are breastfed exclusively (X_6) increases by one percent on the condition that the other variables remain constant. Nevertheless, the occurrence of children under the age of five suffering from malnutrition may increase by 0.18% if the number of poor households (X_2) increases by one percent and 0.14%, if the percentage of children under the age of five suffering from infection (X_3) increases by one percent on the condition that the other variables remain constant. Similar interpretation will also be valid for each variable in the regression model.

In Table 1, of the six independent variables, only one of them has a significant impact on the dependent variable at $\alpha=5\%$, namely the variable of the percentage of the poor population (X_2). Hence, for the next analysis, the variables used for developing the GWR model is the X_2 variable.

The result of Breusch-Pagan test reveals a BP value of 9,4642; meanwhile, the Chi-Square table at a significant level (α) at 10% is 9.24%. Therefore, the decision taken in this case is to reject H_0 , which means that the linear regression has a spatial heterogeneity effect. The occurrence of spatial heterogeneity case on children under the age of five suffering from malnutrition in East Java indicates that the regression model parameters are influenced by the factors about region observed, in this case, the geographic location of a district/city. To overcome the effect of spatial heterogeneity, it is necessary to use a local regression model, namely the geographically weighted regression (GWR).

The first step in obtaining GWR model is to establish the region of observations, in this case the geographical location (latitude and longitude) of each district/city in East Java Province. Next is calculating the Euclidian distances based on the geographical location of each district/city in East Java Province. Then, determining the optimum bandwidth for each district/city with kernel adaptive bi-square function using R 2.15.0 software.

Table 1. Test of Regression Model Parameters

Estimation	Coeff	Std Error	t-value	p-value	VIP
intercept	-4.955	7.240	-0.68	0.499	
X_1	-0.043	0.037	-1.15	0.259	1.4
X_2	0.175	0.070	2.48	0.019*	2.7
X_3	0.138	0.109	1.27	0.215	1.4
X_4	0.101	0.119	0.85	0.403	2.8
X_5	0.206	0.292	0.70	0.487	1.2
X_6	-0.049	0.103	-0.47	0.638	1.6

Remarks: *) significant at $\alpha=5\%$

Bandwidth calculation is based on the distance of one region to its nearest neighbor that affecting the region. The bandwidth value with kernel bi-square function for the Province of East Java was obtained from the result of iteration of 0.483 with CV=204.98.

Parameter estimation for GWR model was obtained by assigning weight to each region of observation in the calculation using the weighted least squares (WLS) method resulted in 38 different models of occurrences of malnutrition for each district/ city in East Java. The value parameter estimation for GWR model with kernel adaptive bi-square weighting function for each region (u_i, v_i); 1, 2, ..., to 38 is presented in Table 2.

Table 2. The GWR Model with Kernel Adaptive bi-square Weighting Function

District/City	Model
Pacitan	2.49 + 0.02X ₂
Ponorogo	2.10 + 0.07 X ₂
Trenggalek	2.51 + 0.02 X ₂
Tulungagung	2.54 + 0.02 X ₂
Blitar	2.42 + 0.05 X ₂
Kediri	2.25 + 0.10 X ₂
Malang	1.90 + 0.17 X ₂
Lumajang	1.62 + 0.23 X ₂
Jember	1.57 + 0.24 X ₂
Banyuwangi	1.29 + 0.24 X ₂
Bondowoso	1.33 + 0.24 X ₂
Situbondo	1.29 + 0.25 X ₂
Probolinggo	1.42 + 0.23 X ₂
Pasuruan	1.44 + 0.23 X ₂
Sidoarjo	1.78 + 0.18 X ₂
Mojokerto	2.28 + 0.13 X ₂
Jombang	2.44 + 0.10 X ₂
Nganjuk	2.05 + 0.07 X ₂
Madiun	2.09 + 0.06 X ₂
Magetan	2.11 + 0.05 X ₂
Ngawi	2.06 + 0.06 X ₂
Bojonegoro	2.07 + 0.07 X ₂
Tuban	2.26 + 0.09 X ₂
Lamongan	2.50 + 0.11 X ₂
Gresik	2.49 + 0.11 X ₂
Bangkalan	1.72 + 0.18 X ₂
Sampang	1.23 + 0.24 X ₂
Pamekasan	1.21 + 0.25 X ₂
Sumenep	1.14 + 0.25 X ₂
City of Kediri	2.23 + 0.10 X ₂
City of Blitar	2.24 + 0.08 X ₂
City of Malang	1.86 + 0.17 X ₂
City of Probolinggo	1.39 + 0.24 X ₂
City of Pasuruan	1.69 + 0.20 X ₂
City of Mojokerto	2.34 + 0.12 X ₂
City of Madiun	2.09 + 0.06 X ₂
City of Surabaya	1.73 + 0.18 X ₂
City of Batu	2.13 + 0.15 X ₂

household (X2) variable has a significant influence on most district/city. It means that the cause of malnutrition occurrences in such a region is poverty, i.e. the inability as seen from the economic aspect to fulfill the basic needs to get food, where a person can only meet the daily requirement for nutrition less than 2,100 calories per day.

Discussion. The result of this study uncovered 38 different models of malnutrition for each district/city in East Java. The finding was in line with *Bitter et al.* (2007) that states the developed GWR model will be unique, i.e. the regression model will be different from one point to the other.

The spatial weighting matrices are the important part in a GWR regression model since spatial parameter estimation is dependent upon the spatial weighting matrix **W**. The weighting matrix used in the study is the kernel adaptive bi-square weighting function. This is based on studies¹³ that revealed the kernel adaptive bi-square weighting function as better in modelling the occurrences of malnutrition in East Java. As for the value of the weighting used in developing model is dependent on the distance between the region observed. The closer the proximity, the greater the influence and the value of the weighting are closer to one. On the other hand, the farther the distance between regions, the smaller the influence, hence the value of the weighting is closer to zero.¹⁴ The above-mentioned weighting matrix is used to estimate the parameter in the region (u_i, v_i). To make an estimation of the parameter in the region (u_2, v_2) it is necessary to calculate the weighting matrix **W**(u_2, v_2) using the same method as above, until the weighting matrix for the final observation is obtained. The weighting calculated in this study up to (u_{3g}, v_{3g}), in the end there are 38 weighting calculated for the study from different points.

The best model was then selected in accordance with the criteria R^2 , MSE and AIC. The best model is the model with the highest R^2 value and the lowest MSE and AIC values.¹⁵ The GWR model is the best model for modelling the percentage of malnutrition occurrences in East Java Province compared to the conventional regression model since the GWR model can increase the value of determinant coefficient (R^2) from 42.17% to 93.72%, and lowering the MSE value from 6.223 to 2.142 and the AIC value from 185.5761 down to 109.5714.

The GWR modelling using bi-square kernel weighting function resulted in 5 groups of districts/cities according to the variables that have significant influences on the occurrences of malnutrition in the Province of East Java. The region of district/city that are near to each other tend to form into the same group on the variables that influencing the occurrences of malnutrition in children

under the age of five years old in a significant manner in East Java. This fact is certainly supported by the first law of geography as proposed by Tobler that states: "*Everything is related to everything else, but near things are more related than distant things*".¹⁶

Conclusions

There 38 GWR models developed that correspond to the total number of districts/cities in East Java (Table 2), wherein each of this district/city will have a different model. The conclusion that can be drawn from this study shows that GWR method is better in modelling the percentage of children under the age of five years old suffering from malnutrition in East Java along with all the independent variables as opposed to the conventional linear regression models. The GWR model can increase the value of determinant coefficient (R^2) from 42.17% to 93.72%, and lower the MSE value from 6.223 to 2.142 and the AIC value from 185.5761 down to 109.5714. The GWR modelling using bi-square kernel weighting function reveals five groups of districts/ cities with similar variables with significant impact on the occurrence of malnutrition in East Java Province.

References

1. Badan Penelitian dan Pengembangan Kesehatan. *Laporan Hasil Riset Kesehatan Dasar* (Riskesdas) Nasional. Jakarta: Departemen Kesehatan Republik Indonesia; 2008. [In Indonesia]
2. Griffith DA. A spatially adjusted ANOVA model. *Geogr. Anal.* 1978;10:296-301.
3. Anselin L. *Spatial Economics: Methods and Models*. Dordrecht: Academic Publishers; 1988.
4. Foster SA, Gorr WL. An adaptive filter for estimating spatially varying parameters: Application to modeling police hours spent in response to calls for service. *Manag. Sci.* 1986;32(7):878-889.
5. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationary. *Geogr. Anal.* 1996;28(4):281-298.
6. Huang Y, Leung Y. Analyzing regional industrialization in Jiangsu Province using geographically weighted regression. *J. Geogr. Syst.* 2002;4:233-249.
7. Farrow A, Larreab C, Hyman G dan Lemaa G. *Exploring the spatial variation of food poverty in Ecuador*. [cited 2013 May 27]. Available from: <http://www.sciencedirect.com/science/article/pii/S0306919205000783>.
8. Duval-Diop D. *Rediscovering the delta a reassessment of the linkages between poverty, economic growth and public policy using geographically weighted regression analysis* [Dissertation]. Louisiana State University and Agricultural & Mechanical College, United State-Louisiana. [cited 2013 Oct 27]. Available from: http://etd.lsu.edu/docs/available/etd-11102006-112931/unrestricted/Duval-Diop_dis.pdf.
9. Mei CL. Geographically weighted regression technique for spatial data analysis, School of Science Xi'an

- Jiaotong University; 2005. Mei CL. Geographically weighted regression technique for spatial data analysis, School of Science Xi'an Jiaotong University; 2005.
10. Badan Pusat Statistik. *Jawa Timur dalam angka*. Jakarta: Badan Pusat Statistik; 2010. [In Indonesia].
 11. Hutabarat IM, Saefuddin A, Hardinsyah, Djuraidah A. *Estimation in measurement error models on cases of malnutrition in the province of East Java*. Proceedings the 3rd Basic Science International Conference. 2013; (3): M21-1-2.
 12. Badan Perencanaan Pembangunan Daerah. [homepage on the Internet]. [update 2013 Mar 18; cited 2013 Oct 15]. Available from: <http://bappeda.jatimprov.go.id/>.
 13. Hutabarat IM, Saefuddin A, Hardinsyah, Djuraidah A, Mangku IW. *Geographically Weighted Regression Modeling on Cases of Malnutrition in East Java Province*. Proceeding of The International Conference on Applied Statistics. 2013; 139-148.
 14. Fotheringham AS, Brunsdon C, Charlton M. *Geographically weighted regression: the analysis of spatially varying relationships*. England: John Wiley & Sons, Ltd., West Sussex; 2002.
 15. Nakaya T, Fotheringham AS., Brunsdon C, and Charlton M. Geographically Weighted Poisson Regression for disease association mapping. *Statistics in Medicine*. 2005; 24: 2695-2717.
 16. Schabenberger O, Gotway CA. *Statistical methods for spatial data analysis*. Chapman & Hall/CRC; 2005.