# Development of Geographically Weighted Regression Using Polynomial Function Approach and Its Application on Life Expectancy Data

Toha Saifudin, Suliyanto, Elly Ana

Department of Mathematics

Faculty of Science and Technology

University of Airlangga

Email: tohasaifudin@fst.unair.ac.id

## Abstract

Geographically Weighted Regression (GWR) is a varying coefficient model. However, as an extension of Ordinary Linear Regression (OLR), it models a dependent variable at each location as a linear function of a set of independent variables. In real life, one or more independent variables involved in the model may have nonlinear relationships with the dependent variable. For this case, the GWR model is no longer realistic to use since the resulted analysis lead to be misleading. To overcome the problem, we develop the GWR by using a polynomial function approach. Here, the model is called Geographically Weighted Polynomial Regression (GWPolR). This paper aims to provide an algorithm, based on Akaike Information Criterion (AIC), for finding the optimal bandwidth and polynomial degrees. Furthermore, this paper aims to analyze life expectancy data in East Java province, Indonesia based on human development index and per capita expenditure. Compared with OLR and GWR models, GWPolR gave a significant improvement of goodness of fit measures and a more complete explanation of how each independent variable was related to the life expectancy.

**Keywords:** *Geographically weighted polynomial regression, Life expectancy, Spatial analysis*

**Introduction**

Geographically Weighted Regression (GWR) model is a popular method in spatial analysis (Brunsdon, Fotheringham, & Charlton, 1999, 1996). The GWR parameters are assumed to be functions of the locations (Brunsdon et al., 1996; Fotheringham, Charlton, & Brunsdon, 1997). This means that GWR is a model with spatially varying coefficients. Some procedures relating to the GWR model have been established. This model has been widely studied both in theory and application. Theoretically, many authors have studied the GWR technique, for example, see (Brunsdon et al., 1999; Fotheringham, Brunsdon, & Charlton, 2003; Fotheringham, Charlton, & Brunsdon, 1998; Harris, Fotheringham, Crespo, & Charlton, 2010; Leung, Mei, & Zhang, 2000). Practically, the GWR technique has been also widely applied to different areas, for example, see (Brunsdon, McClatchey, & Unwin, 2001; Fotheringham, Charlton, & Brunsdon, 2001; Han & Gorman, 2013; Lu, Charlton, Harris, & Fotheringham, 2014; Wang, Zhang, & Yan, 2012).

It is important to note that the GWR model is an extension of an Ordinary Linear Regression (OLR) model. By this extension, it can be viewed that the dependent variable in each location is modeled as a linear function of a set of independent variables. If a nonlinear relationship is present in sample, then the linear approach in the GWR model may be unrealistic to use. There are many possibilities of nonlinearity cases in the relationships between one or more independent variable and the dependent variable. For instance, see (Chamidah, Saifuddin, & Rifada, 2014; Chiang, Peng, & Chang, 2015).

Over the years, some extensions of GWR have been proposed. One of the extensions is Geographically Weighted Generalized Linear Models (GWGLM). While the GWR assumes that the dependent variable is a continuous type and the error term follows a Normal distribution, GWGLM fits generalized linear models with geographically local coefficients. The GWGLM accommodates count and binomial dependent variables by using Geographically Weighted

Poisson Regression (GWPR) and Geographically Weighted Logistic Regression (GWLR), respectively (Fotheringham, Brunsdon, & Charlton, 2002). Furthermore, the Mixed Geographically Weighted Poisson Regression (MGWPR) has been developed (Nakaya, Fotheringham, Brunsdon, & Charlton, 2005). Although the GWGLM has been introduced, extension of GWR, which accommodates dependent variable in continuous variable and has nonlinear relationships with one or more involved independent variables, has not been found. Thus, an extension of the GWR model which can overcome the problem is needed.

In global regression, the nonlinear relationships between the dependent variable Y and any independent variable X can be modeled by using polynomial regression model (Rawlings, Pantula, & Dickey, 1998). In this paper, to overcome the problem above we propose a development of GWR model by using a polynomial function approach, namely Geographically Weighted Polynomial Regression (GWPolR) model. The polynomial function is chosen due to its flexibility in following the behaviour pattern of the real data. Then, the purpose of this paper is to provide an algorithm for finding the optimal bandwidth and polynomial degrees in the GWPolR technique. Furthermore, it will be applied to find a life expectancy model, in East Java province of Indonesia, based on human development index and per capita expenditure. This case was taken in this study since each independent variable tends to have a nonlinear relationship with the life expectancy. Based on the sample, the GWPolR model will be compared with the previous models, i.e., OLR and GWR models.

### Related Work: Geographically Weighted Regression

In this section, we briefly describe the basic GWR technique which has been described by some previous authors, such as (Brunsdon et al., 1999, 1996; Fotheringham et al., 2003, 1998; Leung et al., 2000). In the basic GWR, the spatial nonstationarity of a regression relationship is explored by locally varying coefficients model of the form

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i)\, x_{ij} + \varepsilon_i, i = 1, 2, \dots, n,\ j = 1, 2, \dots, p \tag{1}$$

where $\beta_j(u_i, v_i)$, for $j = 1, 2, \dots, p$ are unknown parameters at location $(u_i, v_i)$, and $\varepsilon_i$ is error

independently distributed $N(0, \sigma^2)$ for $i = 1, 2, \dots, n$.

Let $\hat{\beta}_j(u_i, v_i)$ be the estimate of $\beta_j(u_i, v_i)$. The Weighted Least Square (WLS) estimate of

parameter vector $\widehat{\boldsymbol{\beta}}(u_i, v_i) = \left[\hat{\beta}_0(u_i, v_i), \hat{\beta}_1(u_i, v_i), \dots, \hat{\beta}_p(u_i, v_i)\right]^{\mathrm{T}}$ can be obtained in the

form of

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = [\mathbf{X}^{\mathrm{T}}\, \mathbf{W}(u_i, v_i)\, \mathbf{X}]^{-1}\, \mathbf{X}^{\mathrm{T}}\, \mathbf{W}(u_i, v_i)\, \boldsymbol{y} \tag{2}$$

where $\mathbf{X}$ is a matrix of order $n \times (1 + p)$ with the $i$th-row of $\boldsymbol{x}_i^{\mathrm{T}} = (1, x_{i1}, x_{i2}, \dots, x_{ip})$,

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^{\mathrm{T}}, \tag{3}$$

$$\mathbf{W}(u_i, v_i) = \mathrm{diag}[w_1(i), w_2(i), \dots, w_n(i)\,] \tag{4}$$

is a diagonal weighting matrix with the elements $w_j(i)$ for $j = 1, 2, \dots, n$. Further study on GWR

can be found on the references mentioned above.

### *The Proposed Method: Geographically Weighted Polynomial Regression*

We develop a linear relationship in the GWR by the polynomial function approach. The

model and an estimation algorithm is given in the following sub-sections.

## Model and Estimation

In this paper, we propose an extension of equation (1) by using the polynomial function

approach in the form of

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \sum_{j=1}^{d_k} \beta_{k,j}(u_i, v_i)\, x_{ik}^{\ j} + \varepsilon_i\,, \tag{5}$$

where $\varepsilon_i \sim iid\, N(0, \sigma^2)$, $\beta_{k,j}(u_i, v_i)$ is the regression coefficient of the $k$th independent variable

on $j$th order of polynomial function at location $i$. Then we call equation (5) with geographically

weighted polynomial regression (GWPolR) model. In a matrix form, equation (5) can be written as

$$y_i = \boldsymbol{x_i^{*T}} \, \boldsymbol{\beta}_{pol}(u_i, v_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{6}$$

where

$$\boldsymbol{x_i^{*T}} = \left(1 \; x_{i1} \; x_{i1}^2 \; \cdots \; x_{i1}^{d_1} \; x_{i2} \; x_{i2}^2 \; \cdots \; x_{i2}^{d_2} \cdots \; x_{ip} \; x_{ip}^2 \; \cdots \; x_{ip}^{d_p}\right), \tag{7}$$

and

$$\boldsymbol{\beta}_{Pol}^{T}(u_i, v_i) = ( \, \beta_0(u_i, v_i) \, \beta_{1,1}(u_i, v_i) \, \beta_{1,2}(u_i, v_i) \cdots \beta_{1,d_1}(u_i, v_i) \beta_{2,1}(u_i, v_i) \, \beta_{2,2}(u_i, v_i) \dots$$

$$\cdots \beta_{2,d_2}(u_i, v_i) \cdots \beta_{p,1}(u_i, v_i) \, \beta_{p,2}(u_i, v_i) \cdots \beta_{p,d_p}(u_i, v_i). \tag{8}$$

For a given location $(u_0, v_0)$, $\boldsymbol{\beta}_{Pol}(u_0, v_0)$ can be estimated by minimizing the weighted least square function as follows

$$\sum_{i=1}^{n} \left(y_i - \boldsymbol{x_i^{*T}} \, \boldsymbol{\beta}_{pol}(u_0, v_0)\right)^2 K_h(d_{01}), \tag{9}$$

with respect to each elements of $\boldsymbol{\beta}_{pol}(u_0, v_0)$. The solution for this weighted least square problem is

$$\widehat{\boldsymbol{\beta}}_{pol}(u_0, v_0) = \left[\mathbf{X}_{pol}^{T}\mathbf{W}(u_0, v_0)\mathbf{X}_{pol}\right]^{-1}\mathbf{X}_{pol}^{T}\mathbf{W}(u_0, v_0)\mathbf{y} \tag{10}$$

where $\mathbf{y} = (y_1, y_2, \dots, y_n)^{T}$,

$$\mathbf{X}_{pol} = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^{d_1} & \cdots & x_{1p} & x_{1p}^2 & \cdots & x_{1p}^{d_p} \\ 1 & x_{21} & x_{21}^2 & \cdots & x_{21}^{d_1} & \cdots & x_{2p} & x_{2p}^2 & \cdots & x_{2p}^{d_p} \\ & & & & \vdots & & & & & \\ 1 & x_{n1} & x_{n1}^2 & \cdots & x_{n1}^{d_1} & \cdots & x_{np} & x_{np}^2 & \cdots & x_{np}^{d_p} \end{bmatrix}, \tag{11}$$

$$\mathbf{W}(u_0, v_0) = \mathrm{diag}[K_h(d_{01}), K_h(d_{02}), \dots, K_h(d_{0n})] \tag{12}$$

is a diagonal weighting matrix with $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$, $K(\cdot)$ is a kernel function, and $h$ is a bandwidth. The $d_{0j}$ is a distance between location $(u_0, v_0)$ and $(u_j, v_j)$ for $j = 1, 2, \dots, n$. Furthermore, a Euclid distance between location $i$ and $j$ can be expressed as

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}. \tag{13}$$

By taking $(u_0, v_0)$ to be each of the locations $(u_i, v_i)$, $i = 1, 2, \dots, n$, then the GWPolR estimate can be obtained as

$$\widehat{\boldsymbol{\beta}}_{pol}(u_i, v_i) = \left[\mathbf{X}_{pol}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{X}_{pol}\right]^{-1}\mathbf{X}_{pol}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\boldsymbol{y}, \tag{14}$$

where $\mathbf{W}(u_i, v_i)$ is a diagonal weighting matrix at location *i*.

## Residual Sum of Squares and Variance Estimate

Based on the estimate, the vector of the fitted values for the dependent variable *y* at *n* designed locations as

$$\widehat{\boldsymbol{y}}_{Pol} = (\hat{y}_1^*, \hat{y}_2^*, \dots, \hat{y}_n^*)^{\mathrm{T}} = \mathbf{G}\,\boldsymbol{y}, \tag{15}$$

where

$$\mathbf{G} = \begin{bmatrix} \boldsymbol{x}_1^{*T} \left[\mathbf{X}_{pol}^T \mathbf{W}(u_1, v_1)\, \mathbf{X}_{pol}\right]^{-1} \mathbf{X}_{pol}^T \ \mathbf{W}(u_1, v_1) \\ \boldsymbol{x}_2^{*T} \left[\mathbf{X}_{pol}^T \mathbf{W}(u_2, v_2)\, \mathbf{X}_{pol}\right]^{-1} \mathbf{X}_{pol}^T \ \mathbf{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^{*T} \left[\mathbf{X}_{pol}^T \ \mathbf{W}(u_n, v_n)\, \mathbf{X}_{pol}\right]^{-1} \mathbf{X}_{pol}^T \ \mathbf{W}(u_n, v_n) \end{bmatrix} \tag{16}$$

is called a hat matrix of GWPolR model, and $\boldsymbol{x}_i^{*\mathrm{T}}$ is the vector written in equation (8). Based on the hat matrix $\mathbf{G}$, the residual vector of GWPolR is

$$\widehat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y} - \widehat{\boldsymbol{y}}_{Pol} = (\mathbf{I} - \mathbf{G})\boldsymbol{y}, \tag{17}$$

and the residual sum of squares (RSS) for GWPolR is

$$\mathrm{RSS}_{pol} = \widehat{\boldsymbol{\varepsilon}}_{pol}^{\mathrm{T}}\widehat{\boldsymbol{\varepsilon}}_{Pol} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\boldsymbol{y}, \tag{18}$$

where $\mathbf{I}$ is identity matrix of order *n*.

Furthermore, here we assume that $\hat{y}_i^*$ is an unbiased estimate of $E(y_i)$, that is $E(\hat{y}_i^*) = E(y_i)$ for all $i = 1, 2, \dots, n$. So, we have

$$E(\widehat{\boldsymbol{\varepsilon}}_{Pol}) = E(\boldsymbol{y}) - E(\widehat{\boldsymbol{y}}_{Pol}) = \mathbf{0}. \tag{19}$$

On the other hand, based on assumption of error terms we have

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = \sigma^2 \mathbf{I}, \tag{20}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n)^{\mathrm{T}}$.

Based on both conditions, we can express RSS$_{pol}$ as

$$\mathrm{RSS}_{Pol} = \left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right)^{\mathrm{T}} \left(\hat{\boldsymbol{\varepsilon}}_{Pol} - E(\hat{\boldsymbol{\varepsilon}}_{Pol})\right) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\boldsymbol{\varepsilon}. \tag{21}$$

So,

$$
\begin{aligned}
E(\mathrm{RSS}_{Pol}) &= E(\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\boldsymbol{\varepsilon}) \\
&= E\left(tr(\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\boldsymbol{\varepsilon})\right) \\
&= tr\left((\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\right) \\
&= \sigma^2 \gamma_1,
\end{aligned}
\tag{22}
$$

where $\gamma_1 = tr\left((\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\right)$. From equation (22), we can take that

$$\hat{\sigma}^2{}_{pol} = \frac{\mathrm{RSS}_{Pol}}{\gamma_1} \tag{23}$$

is an unbiased estimate of $\sigma^2$. Practically, $\gamma_1$ can be expressed by the following form

$$
\begin{aligned}
\gamma_1 &= tr\left((\mathbf{I} - \mathbf{G})^{\mathrm{T}}(\mathbf{I} - \mathbf{G})\right) \\
&= tr(\mathbf{I}) - tr(\mathbf{G}) - tr(\mathbf{G}^{\mathrm{T}}) + tr(\mathbf{G}^{\mathrm{T}}\mathbf{G}) \\
&= n - 2\, tr(\mathbf{G}) + tr(\mathbf{G}^{\mathrm{T}}\mathbf{G}).
\end{aligned}
\tag{24}
$$

## Spatial Weighting Function

The weighting matrix $\mathbf{W}(u_i, v_i)$ is designed based on the proximity of the observation point $i$ to the data points around $i$ (Fotheringham, Brunsdon, & Charlton, 2002). For a fixed bandwidth $h$ and the distance $d_{ij}$, some functions commonly used to form elements of the weighting matrix at location i are

1. Gaussian Kernel

$$K_h\left(d_{ij}\right) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right), \ \ j = 1, 2, \dots, n, \tag{25}$$

2. Bisquare Kernel

$$K_h\left(d_{ij}\right) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & if \ d_{ij} \leq h \\ 0, & if \ d_{ij} > h. \ j = 1, 2, \dots, n. \end{cases} \tag{26}$$

## Akaike Information Criterion for The Optimal Bandwidth and Polynomial Degrees

We know that the estimated parameters of GWPolR are dependent on the weighting function selected. So, in the GWPolR estimation, the weighting matrix should first be determined. As an illustration, if we use the weighting function in equation (15), the bandwidth $h$ should be determined. If $h$ becomes larger at least equal to the maximum distance between points in the study area, the model solution will be close to the global polynomial regression. Conversely, if $h$ becomes smaller, then the estimated parameters will increasingly depend on observations close to location $i$, and hence it will increase variety.

On the other hand, the estimated parameters of GWPolR depend on the polynomial degrees of independent variables. If we determine that each independent variable has a polynomial degree of one, the GWPolR will become basic GWR. Whilst, as we use higher polynomial degree for each independent variable, the model solution will be complicated.

Based on the description above, we need to select the optimal bandwidth and the optimal number of polynomial degrees of each independent variable involved in the model. A criterion which provides a trade-off between goodness-of-fit and degrees of freedom is to minimize the Akaike Information Criterion (AIC) (Fotheringham, Brunsdon, & Charlton, 2002). Here, we adopt the AIC for GWPolR as

$$AIC_c = 2nLog_e\left(\hat{\sigma}_{pol}\right) + nLog_e(2\pi) + n\left(\frac{n+\text{tr}(\mathbf{G})}{n-2-\text{tr}(\mathbf{G})}\right). \tag{27}$$

where $n$ is the sample size, $\hat{\sigma}_{pol}$ is the estimated standard deviation of the error term by GWPolR

model, that is, square root of equation (23), and tr($\mathbf{G}$) denotes the trace of the hat matrix in

equation (16).

## An Algorithm for Finding The Optimal Condition Based on Akaike Information Criterion

Here we provide an algorithm to select the optimal bandwidth and the optimal number

of polynomial degrees based on the AIC$_c$ approach with the objective function in equation (27).

The algorithm is given as follows:

1) Enumerate the number of independent variables involved in the model, denoted by $p$.

2) Set the maximum polynomial degree of each independent variable, denoted by $d_j$ for

   $j = 1, 2, \ldots, p.$

3) Find all arrays that can be arranged from the possible polynomial degrees of the

   ndependent variables. Suppose that $s$ is the number of arrays, then $s = \prod_{j=1}^{p} d_j$.

4) Find the minimum AIC$_c$ value and accordingly bandwidth based on GWPolR modelling

   in each array.

5) Find the smallest AIC$_c$ value among the minimum AIC$_c$ values resulted from the entire

   arrays in Step 4.

6) Select the bandwidth and polynomial degrees that yield the smallest AIC$_c$ value obtained

   in Step 5 to be the optimal solution.

### Application to Life Expectancy Data

### Data

Life Expectancy data in this research was obtained from The Statistics of East Java,

Indonesia. The data involved 38 observation units consisting of 29 districts and 9 cities in East

Java in 2017. The observed attributes of each district or city are Life Expectancy Rate (LER) in years, Human Development Index (HDI) without units, and Per capita Expenditure (PE) in thousands of Rupiah. In this study, the dependent variable is LER notated by Y. Whereas, the independent variables are HDI and PE which notated by $X_1$ and $X_2$, respectively.

## A Preliminary Modelling: Ordinary Linear Regression

One of the main purposes of regression is to provide a summary of the relationship between the dependent variable Y and any independent variable X (Green & Silverman, 1994). The trend of those relationships can be conventionally approximated by using a scatter diagram (Draper & Smith, 1998).  This scatter diagram can visually display linear or curvilinear trends. The trend of relationships between LER and each independent variable can be seen in Figure 1.

For preliminary modelling, the relationships between LER and the independent variables were evaluated by a linear approach using OLR technique. Two individual models and a simultaneous model were investigated. Firstly, the OLR model of LER based on HDI was performed. Its results are presented in Table 1. This table shows that, by using a significance level of 1%, the HDI significantly affects LER with a positive sign of coefficient estimate (Coef: 0.286 with *p*-value: 0.000).

The second one, LER was regressed on PE. The results are presented in Table 2. Based on Table 2, it can be seen that, by using a significance level of 0.01, PE also significantly affects LER with a positive sign of coefficient estimate (Coef: 0.00049 with *p*-value: 0.001).

Furthermore, LER was simultaneously regressed on HDI and PE. Based on the two independent variables, the simultaneous test for parameters yielded a *p*-value of 0.000. This means that the two independent variables simultaneously affect LER. The OLR model yielded RSS of 43.410 and $R^2$ of 71.8%. The OLR estimates and the corresponding partial tests are given in Table 3. In this OLR model, collinearity between HDI ($X_1$) and PE ($X_2$) was investigated. If

the VIF value > 10, it indicates the presence of multicollinearity among the independent variables (Montgomery, Peck, & Vining, 2012). Based on collinearity statistics in Table 3, it can be concluded that there was no collinearity in this case (VIF < 10). Therefore, the two independent variables can be used simultaneously in the model.

Using significance level of 0.01, Table 3 shows that all parameters are significant (*p*-value < 0.01). This means that HDI and PE significantly affect LER. Nevertheless, an irregularity was found. The estimated coefficient of PE ($X_2$) was negative (-0,00094). Whereas, the scatter plot of $X_2$ in Fig 1(b) visually indicated an increasing trend. Also, the individual model, i.e., the OLR model of LER on PE presented in Table 2, showed that PE significantly affects LER with a positive sign of slope. This is a contradiction. So, we can state that the OLR model is not feasible for this simultaneous model.

## Geographically Weighted Regression Results

This data was spatial type, therefore modelling with GWR model was warranted. For this sample, the GWR technique was performed using the following model

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \varepsilon_i. \tag{28}$$

In this research, GWR modelling was performed on R-3.5.1 software. Each location is marked by a coordinate point consisting of latitude and longitude. By using weighting function of Gaussian Kernel in equation (16), it was obtained an optimal bandwidth of 0.8401029° (or equals to 93.52 km) when the minimum $AIC_c$ was 104.7531. Based on this optimal condition, the GWR model yielded an $R^2$ of 80.31% and RSS of 30.3514. Based on these goodness of fit measures ($AIC_c$, RSS, and $R^2$), GWR modelling showed an improvement from OLR modelling. This model has three parameters that vary in 38 locations. Summary of the GWR estimates on the Life Expectancy data is listed in Table 4. In addition, the GWR model is in fact very robust

to the effects of multicollinearity (Fotheringham & Oshan, 2016). Therefore, in this paper we did not carry out multicollinearity analysis.

Table 4 shows that each coefficient of GWR varies between minimum value (Min) and maximum value (Max). Nevertheless, it can be seen that the signs of GWR estimates are generally similar to those of OLR estimates. In this model, PE ($X_2$) coefficients also generally have a negative sign. This is contrary to the result obtained by the individual model. So, we can state that the GWR model is also not feasible for this case.

## Geographically Weighted Polynomial Regression Results

Visually, Figure 1(a) shows that the relationship between LER and HDI forms a nonlinear pattern, as well as the relationship between LER and PE in Figure 1(b). This is an early reason to extend the linear modelling to be polynomial modelling on Life Expectancy data. Therefore, we continue with the GWPolR modelling. We suspected that a GWPolR technique would give better results and help us to better explain how independent variables were related to Life Expectancy.

In this modelling, we applied the algorithm in section 3.5 for selecting the optimal bandwidth and polynomial degrees of GWPolR model. Based on Life Expectancy data we involved two independent variables in the model, so we have $p = 2$. Then, we set the maximum polynomial degree for each independent variable. By looking at the pattern shown by the scatter plot in Figure 1, we think that each pattern can be sufficiently modeled using a polynomial approach with a maximum degree of 3. Therefore, we set the same value for the maximum polynomial degree of $X_1$ and $X_2$ variables, i.e., $d_1 = d_2 = 3$. Based on this setting, we have a number of $s = 3 \times 3 = 9$ arrays of polynomial degrees. Next, we select the minimum $AIC_c$ value based on GWPolR procedure in each array. By using the weighting function of Gaussian

Kernel in equation (16), it was obtained minimum $AIC_c$ value and the accordingly optimal bandwidth listed in Table 5.

The smallest $AIC_c$ value among nine minimum $AIC_c$ values is 96.1493.  It is found in the row of number five, according to the optimal bandwidth of $0.7259163^o$ (or equals to 80.81 km) and array of (2, 2). This means that the optimal polynomial degree for each independent variable is 2. So, the GWPolR model under optimal condition in this study was performed by using the following model:

$$y_i = \beta_0(u_i, v_i) + \beta_{1,1}(u_i, v_i)x_{i1} + \beta_{1,2}(u_i, v_i)x_{i1}^2 + \beta_{2,1}(u_i, v_i)x_{i2} + \beta_{2,2}(u_i, v_i)x_{i2}^2 +$$

$$\varepsilon_i \tag{29}$$

Based on the optimal condition, the GWPolR model resulted an $R^2$ of 85.90% and RSS of 21.7375. This GWPolR model has five parameters that vary in 38 locations. Furthermore, summary of GWPolR coefficient estimates on the Life Expectancy data is listed in Table 6.

According to Table 6, it can be stated that the optimal GWPolR model on Life Expectancy data has been able to accommodate real data patterns with the presence of second-degree polynomial parameters. Based on the estimation results, it is known that the estimated coefficient of each independent variable generally has a positive sign on the first degree of polynomial and a negative sign on the second degree of polynomial. This means that the relationship between each independent variable and LER is nonlinear, i.e., it forms an increasing pattern but with a deceleration (not constant speed). It shows that the GWPolR model has followed the actual data pattern. Thus, it can be said that this model provided a more complete understanding of how each independent variable was related to LER.

## Model Comparison

In this section, we summarize model comparison based on some indicators. The results are presented in Table 7. From Table 7, the GWPolR model gave the best performance. In

addition, the GWPolR model has reduced the RSS value by 21.6725 and 8.6139 from OLR and GWR, respectively. Also, it increases $R^2$ by 14.10% and 5.59% from GLR and GWR, respectively. Furthermore, the GWPolR model is feasible while the two others are not. This is relatively strong evidence for the improvement of GWPolR modelling on the Life Expectancy data.

To compare visually the model results, actual LER and its estimates were mapped by using quantile classification method which is shown by Figure 2. From Figure 2, it can be seen that the geographical distribution of GWPolR fits is more similar to the actual LER than that of GWR or OLR fits. It shows that the GWPolR estimate is the best for this case.

**Conclusion**

By searching the smallest AIC among the minimum AIC in each array, for all possible bandwidth values, it can be guaranteed that the array and the corresponding bandwidth resulted by the algorithm in this paper, are in optimal condition. The GWPolR will be the GWR model when each independent variable has an optimal polynomial degree of one. Computationally, this algorithm still takes a long time due to its searching in all possible bandwidths and all arrays of polynomial degrees. The time will be longer when the number of variables involved in the model increases. For future research, some algorithms can be developed using a modern approach, for example, genetic algorithm or neural network.

Based on some goodness of fit criteria (inter alia: $AIC_c$, RSS, and $R^2$), it can be empirically concluded that the GWPolR model with a polynomial degree of two for HDI and PE is the best model for life expectancy in East Java, Indonesia. The GWPolR model provides a more complete understanding of how each independent variable was related to LER. However,

hypothesis tests have to be developed to determine whether a GWPolR is statistically better than a GWR or not.

**Acknowledgments**

**Table 1.** Summary of OLR results on LER by HDI ($X_1$)

| Predictor | Coef | SE Coef | T | *p*-value |
|---|---|---|---|---|
| Intercept | 51.019 | 3.016 | 16.92 | 0.000 |
| $X_1$ | 0.286 | 0.0427 | 6.69 | 0.000 |

**Table 2.** Summary of OLR results on LER by PE ($X_2$)

| Predictor | Coef | SE Coef | T | *p*-value |
|---|---|---|---|---|
| Intercept | 65.971 | 1.485 | 44.42 | 0.000 |
| $X_2$ | 0.00049 | 0.00014 | 3.55 | 0.001 |

**Table 3.** Summary of OLR results on Life Expectancy data

| Predictor | Coef | SE Coef | T | *p*-value | Collinearity Statistics | |
|---|---|---|---|---|---|---|
| | | | | | Tolerance | VIF |
| Intercept | 36.936 | 3.954 | 9.34 | 0.000 | - | - |
| $X_1$ | 0.627 | 0.083 | 7.55 | 0.000 | 0.172 | 5.801 |
| $X_2$ | -0.00094 | 0.00021 | -4.52 | 0.000 | 0.172 | 5.801 |

**Table 4.** Summary of GWR coefficient estimates on the Life Expectancy data

| Predictor | Min. | 1st Qu. | Median | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| Intercept | 34.180 | 37.830 | 39.940 | 43.830 | 46.370 |
| $X_1$ | 0.468 | 0.523 | 0.562 | 0.593 | 0.664 |
| $X_2$ | -0.00096 | -0.00085 | -0.00081 | -0.00076 | -0.00064 |

**Table 5.** Optimal bandwidth and minimum $AIC_c$ for each array of polynomial degrees on the

Life Expectancy data

| Number | Array | Opt $h$ | Minimum $AIC_c$ |
|---|---|---|---|
| 1 | (1, 1) | 0.8401029 | 104.7531 |
| 2 | (1, 2) | 0.9039733 | 104.1308 |
| 3 | (1, 3) | 0.7991147 | 100.9174 |
| 4 | (2, 1) | 0.7718448 | 100.2573 |
| 5 | (2, 2) | 0.7259163 | 96.1493 |
| 6 | (2, 3) | 0.7399819 | 96.4080 |
| 7 | (3, 1) | 0.7214168 | 96.6860 |
| 8 | (3, 2) | 0.8869832 | 100.5664 |
| 9 | (3, 3) | 1.2562700 | 106.1920 |

**Table 6.** Summary of GWPolR coefficient estimates on the Life Expectancy data

| Predictor | Min. | 1st Qu. | Median | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| Intercept | -70.120 | -28.850 | -9.884 | 28.040 | 118.400 |
| $X_1$ | -1.708 | 1.084 | 2.303 | 2.875 | 4.003 |

| | | | | | |
|---|---|---|---|---|---|
| $X_1^2$ | -0.0246 | -0.0168 | -0.0122 | -0.0034 | 0.0171 |
| $X_2$ | 0.0013 | 0.0023 | 0.0026 | 0.0029 | 0.0035 |
| $X_2^2$ | -0.000116 | -0.000095 | -0.000082 | -0.000068 | -0.000011 |

**Table 7.** Model comparison on the Life Expectancy data

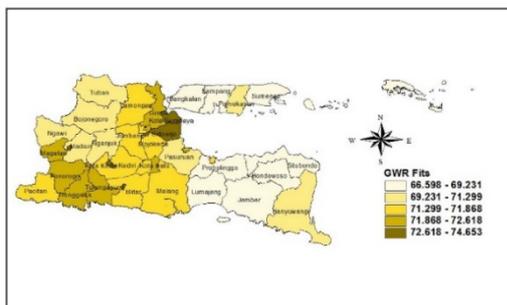| Indicator | Model | | |
|---|---|---|---|
| | OLR | GWR | GWPolR |
| $AIC_c$ | - | 104.7531 | 96.1493 |
| RSS | 43.4100 | 30.3514 | 21.7375 |
| $R^2$ | 71.8% | 80.31% | 85.90% |
| Model feasibility | Not feasible | Not feasible | Feasible |
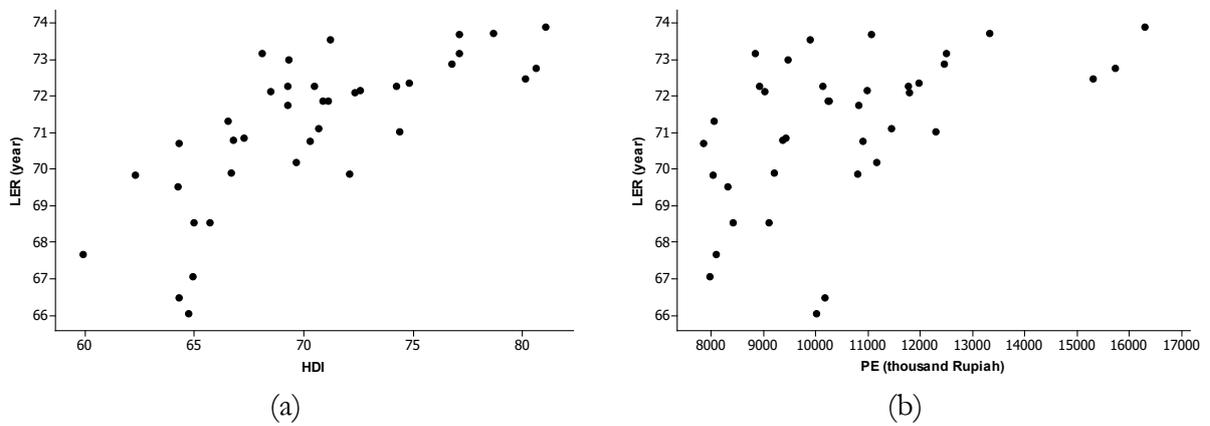


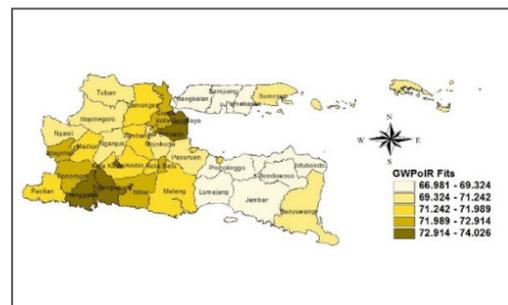**Figure 1.** Scatter plot of LER versus (a) HDI and (b) PE
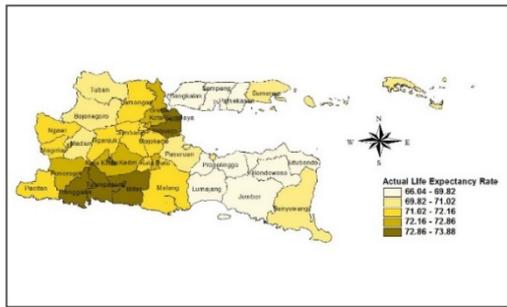
Geographical distribution of actual LER
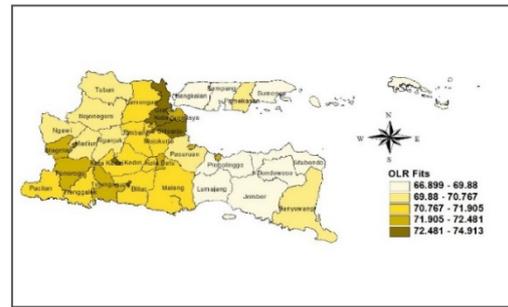


Geographical distribution of OLR Fits



Geographical distribution of GWR Fits



Geographical distribution of GWPolR Fits

**Figure 2.** Geographical distribution of actual LER and its estimates

**References**

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, *39*(3), 497–524.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298.

Brunsdon, C., McClatchey, J., & Unwin, D. J. (2001). Spatial variations in the average rainfall–altitude relationship in Great Britain: an approach using geographically weighted regression. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(4), 455–466.

Chamidah, N., Saifuddin, T., & Rifada, M. (2014). The Vulnerability Modeling of Dengue Hemorrhagic Fever Disease in Surabaya Based on Spatial Logistic Regression Approach. *Applied Mathematical Sciences*, *8*(28), 1369–1379.

Chiang, Y.-H., Peng, T.-C., & Chang, C.-O. (2015). The nonlinear effect of convenience stores on residential property prices: A case study of Taipei, Taiwan. *Habitat International*, *46*, 82–90.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Fotheringham, A. S., Charlton, M., & Brunsdon, C. (1997). Two techniques for exploring non-stationarity in geographical data. *Geographical Systems*, *4*(1), 59–82.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, *30*(11), 1905–1927.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (2001). Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical and Environmental Modelling*, *5*(1), 43–66.

Han, D., & Gorman, D. M. (2013). Exploring spatial associations between on-sale alcohol availability, neighborhood population characteristics, and violent crime in a geographically isolated city. *Journal of Addiction*, *2013*.

Harris, P., Fotheringham, A. S., Crespo, R., & Charlton, M. (2010). The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, *42*(6), 657–680.

Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, *32*(1), 9–32.

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(4), 660–681.

Wang, C., Zhang, J., & Yan, X. (2012). The use of geographically weighted regression for the relationship among extreme climate indices in China. *Mathematical Problems in Engineering*, *2012*.