Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment Rate in Kalimantan

Sifriyani¹*, Hillidatul Ilmi¹, and Zakiyah Mar'ah²
¹Statistics Study Program, Mathematics Department, Faculty of Mathematics and Natural Sciences, Mulawarman University, Indonesia
²Statistics Study Program, Faculty of Mathematics and Natural Science, Universitas Negeri Makassar, Indonesia

*Email: sifriyani@fmipa.unmul.ac.id

Abstract. This study was conducted specifically to GIS mapping based on Nonparametric-Geographically Weighted Spline Regression (NGWSR) Estimation Model for the factors that affect the open unemployment rate (OUR) in Kalimantan. Observational data in this study were categorized into 56 regions based on the Regency/City scale in Kalimantan. The variables used in this study consisted of the open unemployment rate, the labor force participation rate, population density, human development index, expected years of schooling, and regional minimum wage. This study utilized the spatial analysis of the NGWSR model with the geographic weighting of the Gaussian and Bisquare kernel functions. The NGWSR model was considered capable of providing a solution to the geographically weighted spatial regression for the unknown regression curve. Regarding to the result of this study, NGWSR with geographic weighting of the Bisquare kernel function was considered as the best model. The model criteria were based on the coefficient of determination and RMSE. The results of the significance test of model parameters for 56 Regencies/Cities data in Kalimantan had succeeded in mapping the area into 14 categories based on the significant variables of each region.

Keyword: GIS Mapping, Geographically Weighted Regression, Truncated Spline, Nonparametric Regression, Open Unemployment Rate in Kalimantan

1. Introduction
This research is about Nonparametric-Geographically Weighted Spline Regression (NGWSR) model. The NGWSR model is a development of the Geographically Weighted Regression (GWR) model [1],[2],[3]. The NGWSR is a truncated spline nonparametric regression that considers geographic or spatial factors [4]. Spline Truncated is a function that is built based on polynomial components and truncated components, namely polynomial pieces that have knot points, which can overcome patterns of changes in data behavior. The spline truncated approach is used as a solution to solve the problem of modeling spatial data analysis in which the relationship between the response variable and the predictor variable does not follow a specific pattern, and some patterns vary in certain sub-intervals [5]. The stages of completion of the NGWSR model consist of NGWSR model estimation with maximum likelihood optimization [4], followed by model suitability test [6], and simultaneous and partial significance test [7].
2. Material and method

2.1 Nonparametric-Geographically weighted spline regression

The Nonparametric-Geographically Weighted Spline Regression (NGWSR) model is frequently assumed to be normally distributed with mean 0 and variance $\sigma^2(u_i, v_i)$ at each location $(u_i, v_i)$. In mathematical terms, the form of the relationship between the response variable $y_i$ and the predictor variable $(x_{i1}, x_{i2}, \ldots, x_{it})$ at the i-th location for the NGWSR model is formulated in Equation (1).

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{d} \sum_{f=1}^{p} \beta_{kf}(u_i, v_i)x_{ik}^f + \sum_{k=1}^{d} \sum_{s=1}^{r} \delta_{k, m+s}(u_i, v_i)(x_{ik} - K_{ks})^m + \varepsilon_i$$

(1)

Where $y_i$ is referred to as the response variable at the i-th location. The parameter coefficient $\beta_{kf}(u_i, v_i)$ is the polynomial component of the NGWSR model with the f-th parameter of the k-th predictor variable in the i-th area. Moreover, $\delta_{k, m+s}(u_i, v_i)$ is defined as the parameter coefficient of the truncated component of the NGWSR model with a total order of m at the s-th knot point and with the k-th predictor variable. $K_{ks}$ is the s-th knot point of the k-th predictor variable component.

The NGWSR model was estimated by using the Maximum Likelihood Estimator (MLE) method. As stated by Sifriyani et al [4], if the regression model (1) is used with error $\varepsilon_i$, normally distributed with mean 0 and variance $\sigma^2(u_i, v_i)$ along with the weighted likelihood function as written in Equation (2).

$$\max_{\hat{\beta}, \hat{\delta}, \sigma^2} \left\{ L \left( \beta(u_i, v_i), \delta(u_i, v_i), \sigma^2(u_i, v_i) \right) \right\} = \max_{\hat{\beta}, \hat{\delta}, \sigma^2} \left\{ \left( \sigma^2(u_i, v_i) \right)^{-\frac{n}{2}} \right\}$$

$$\exp - \frac{1}{2\sigma^2(u_i, v_i)} \sum_{j=1}^{n} w_{l(j)} 
\left[ y_j - \left( \beta_0(u_i, v_i) + \sum_{p=1}^{i} \sum_{k=1}^{m} \beta_{pk}(u_i, v_i) x_{pi}^k + \sum_{p=1}^{r} \sum_{h=1}^{m} \delta_{p, m+h}(u_i, v_i)(x_{pi} - K_{ph})^m \right) \right]^2$$

(2)

then the MLE estimator for $\hat{\beta}(u_i, v_i)$ and $\hat{\delta}(u_i, v_i)$ is expressed by Equation (3)

$$\hat{\beta}(u_i, v_i) = A(K)\tilde{Y}$$
$$\hat{\delta}(u_i, v_i) = B(K)\tilde{Y}$$

(3)

where

$$A(K) = S \left( X^T W(u_i, v_i) X \right)^{-1} \left[ X^T - X^T W(u_i, v_i) P (P^T W(u_i, v_i) P)^{-1} P^T \right] W(u_i, v_i)$$
$$B(K) = R \left( P^T W(u_i, v_i) P \right)^{-1} \left[ P^T - P^T W(u_i, v_i) X (X^T W(u_i, v_i) X)^{-1} X^T \right] W(u_i, v_i)$$

Where $\hat{\beta}(u_i, v_i)$ is defined as estimator of parameter $\beta(u_i, v_i)$ with size vector $(1 + (d \times p)) \times 1$, $\hat{\delta}(u_i, v_i)$ estimator of parameter $\delta(u_i, v_i)$ with size vector $(d \times r) \times 1$, $A(K)$ is referred to as the Hat matrix for $\hat{\beta}(u_i, v_i)$ estimator with matrix size $1 + (d \times p) \times n$, and $B(K)$ is Hat matrix for $\hat{\delta}(u_i, v_i)$ estimator with matrix size $(d \times r) \times n$. Furthermore, $y$ is the response variable vector with size $n \times 1$. 

The NGWSR model may be applied to various fields in environmental management [8], [9], urban mapping, disease spread and population fields. One of the major social issues faced in the field of population is the increase in unemployment. Referring to that matter, this study was conducted to determine the factors that may provide an influence on the Open Unemployment Rate (OUR) in Kalimantan by using the NGWSR model. Furthermore, the estimation results of the NGWSR model were mapped to identify the level of influence of the predictor variables for each regency/city.
X is the predictor variable matrix with size $n \times (1 + (d \times p))$, P is the predictor variable matrix of truncated function with matrix size $n \times (d \times r)$, and W is weighted with matrix size $n \times n$.

### 2.2 Data and data sources

The data involved in this study consisted of secondary data, which were successfully derived from the National Bureau of Statistics (BPS) and the Office of Manpower in 2019. Furthermore, this study utilized observational data consisting of the Regency/City scale of 56 regions in Kalimantan. The response variable used in this study was the open unemployment rate. The predictor variable x is described in Table 1.

| Variables | Symbol | Description of Variables | Sources of Observational Data | Unit | Scale |
|-----------|--------|--------------------------|-------------------------------|------|-------|
| Respon    | $y$    | Open Unemployment Rate (OUR) | BPS of West Kalimantan [10], BPS of Central Kalimantan [11], BPS of South Kalimantan [12], BPS of East Kalimantan [13], BPS of North Kalimantan [14]. | Percent | 56 Regencies/Cities in Kalimantan |
| Predictor | $x_1$  | Labor Force Participation Rate (LFPR) | BPS of West Kalimantan [15], BPS of Central Kalimantan [16], BPS of South Kalimantan [17], BPS of East Kalimantan [18], BPS of North Kalimantan [19]. | Percent | 56 Regencies/Cities in Kalimantan |
|           | $x_2$  | Population Density        | BPS of West Kalimantan [20], BPS of Central Kalimantan [21], BPS of South Kalimantan [17], BPS of East Kalimantan [18], BPS of North Kalimantan [19]. | People | 56 Regencies/Cities in Kalimantan |
|           | $x_3$  | Human Development Index (HDI) | BPS of West Kalimantan [15], BPS of Central Kalimantan [22], BPS of South Kalimantan [17], BPS of East Kalimantan [23], BPS of North Kalimantan [19]. | Percent | 56 Regencies/Cities in Kalimantan |
|           | $x_4$  | Expected Years of Schooling (EYS) | BPS of West Kalimantan [15], BPS of Central Kalimantan [16], BPS of South Kalimantan [17], BPS of East Kalimantan [18], BPS of North Kalimantan [19]. | Year | 56 Regencies/Cities in Kalimantan |
|           | $x_5$  | Regional Minimum Wage      | Dinas Ketenagakerjaan of West Kalimantan [24], Dinas Ketenagakerjaan of Central Kalimantan [25], BPS of South Kalimantan [12], BPS of East Kalimantan [13], BPS of North Kalimantan [14]. | Rupiah | 56 Regencies/Cities in Kalimantan |

### 2.3 Data analysis procedure

Referring to the objectives of this study, the researchers were successfully conducted several stages of data analysis to map the GIS Based NGWSR Model for OUR Data in Kalimantan.

1. The Exploration of the distribution of response variables and predictor variables through the mapping of spatial distribution
2. Description statistics of the open unemployment rate data and predictor variables
3. Estimation of the NGWSR model
4. Mapping the distribution of OUR based on the estimation of the NGWSR model

### 3. Result and discussion

#### 3.1 Data exploration

Observational data are categorized based on variables and are described in Figures 1. Figure 1(a) shows that OUR in the highest category of 1.74 – 3.48 percent was found in 21 Regencies/Cities. The lowest OUR was found in Pulang Pisau Regency by 1.74 percent and the highest OUR was found in Bontang...
City by 9.19 percent. In Figure 1(b), LFPR in the highest category of 65.40 – 68.65 percent was found in 18 Regencies/Cities. The lowest LFPR was found in Pontianak City by 61.62 percent and the highest LFPR was found in Sekadau Regency by 77.19 percent. Moreover, Figure 1(c) indicates that the highest population in the category of 177,700 - 342,217 people was found in 20 Regencies/Cities. The lowest population was found in Mahakam Ulu Regency amounted to 26,400 people and the highest population was found in Samarinda City amounted to 872,800 people. Figure 1(d) shows that HDI in the highest category of 69.88-74.88 percent was found in 19 Regencies/Cities. The lowest HDI was found in North Koyang Regency by 62.66 percent and the highest HDI was found in Palangka Raya City by 80.77 percent. Figure 1(e) shows that EYS in the highest category of 12 - 13 Years was found in 20 Regencies/Cities. The lowest EYS was found in Melawai Regency by 11.15 years and the highest EYS was found in Pontianak City by 14.99 years. Figure 1(f) shows that the minimum wage in the highest category of IDR 2,787,920 – IDR 2,990,358 was found in 27 Regencies/Cities.
3.2. Description of OUR data and predictor variables

Descriptions of observational data in this study, which include response variables and predictor variables are presented in Table 2.

**Table 2.** Description of open unemployment rate data for 56 regencies/cities in Kalimantan and predictor variables

| N  | OUR  | LFPR | Population Density | HDI  | EYS  | Regional Minimum Wage |
|----|------|------|--------------------|------|------|----------------------|
|    | 56   | 56   | 56                 | 56   | 56   | 56                   |
| Mean | 4.35 | 69.04 | 288.967            | 70.41| 12.67| 2.753.530            |
| Minimum | 1.74 | 61.62 | 26.400             | 62.66| 11.15| 2.175.200            |
| Maximum | 9.19 | 77.19 | 872.800            | 80.77| 14.99| 3.462.192            |
| Median | 4.15 | 68.75 | 246.012            | 69.21| 12.51| 2.866.773            |

| Total | 243.86 | 3866.51 | 16.182.122 | 3943.23 | 709.71 | 154.197.669 |
| Range | 7.45 | 15.57 | 846.400 | 18.11 | 3.84 | 1.286.992 |
| Variance | 2.72 | 13.12 | 38.321.222.081 | 19.41 | 0.733 | 74.006.525.803 |

| Deviation Standard | 1.65 | 3.62 | 195.758 | 4.41 | 0.86 | 272.041 |

Statistical data description of OUR data of 56 Regencies/Cities in Kalimantan as presented in Table 2 shows that the total average of OUR data was amounted to 4.35 percent, with the lowest OUR found in Pulang Pisau Regency by 1.74 percent and the highest found in Bontang City.

3.3. Parameter estimation of nonparametric-geographically weighted spline regression

Table 3 shows a summary of the parameter estimator values for the NGWSR model that were successfully obtained by means of geographic weighting of the Bisquare Fixed Bandwidth Kernel Function.

**Table 3.** Summary of NGWSR model parameter estimator values

| Parameter Estimator | Minimum | Q1 | Median | Q3 | Maximum |
|---------------------|---------|----|--------|----|---------|
| $\hat{\beta}_0$    | -0.0323400 | -0.0003904 | 0.0120840 | 0.0186420 | 0.0372130 |
| $\hat{\beta}_1$    | -0.535456 | -0.18132 | -0.11206 | -0.06482 | 0.08336 |
| $\hat{\beta}_2$    | -9.180e-07 | -5.290e-08 | 1.635e-06 | 2.100e-06 | 3.080e-06 |
| $\hat{\beta}_3$    | -0.29425 | 0.09172 | 0.20515 | 0.24409 | 0.47536 |
| $\hat{\beta}_4$    | -1.87550 | -0.33793 | 0.03173 | 0.65153 | 1.12800 |
| $\hat{\beta}_5$    | -2.320e-06 | -1.340e-06 | -2.325e-07 | 7.385e-07 | 5.820e-06 |

Regarding to the estimation results of the NGWSR model parameter, the contributing factors to the open unemployment rate were successfully identified by the researchers, which included OUR, LFPR, Population Density, HDI, EYS and Regional Minimum Wage. The results of the estimator calculation of $\hat{\beta}$ indicated that the highest effect of LFPR was found in Sukamara Regency. The LFPR in this Regency showed the total working age population who were actively involved in the labor market, both individuals who were already employed and those who were still looking for work. Referring to this matter, the labor market is always expected to be improved, which subsequently able to reduce the OUR. The effect of the highest population was found in the Kotabaru Regency. This was due to a fairly high population density which was not in accordance with the existing area. Meanwhile, the highest
The HDI effect was found in the East Kutai region. The HDI in this region showed the achievement of human development based on the basic components of the quality of human life. Therefore, the government is highly required to allow the public to access all existing facilities development, so the OUR may be significantly reduced. Furthermore, the highest EYS effect was found in Sukamara Regency. EYS was perceived as an indicator that described the length of schooling (in years) by children at a certain age. Having regard to this matter, the government is expected to be consistent in providing adequate school facilities for school-age children for at least 12 years of education, thus reducing the OUR. Moreover, the effect of the minimum wage was found in Nunukan City. The amount of the minimum wage showed that the reward provided by the employer to the employee for a job or service that has been performed was very low. Consequently, this should be taken into consideration by the government to reduce the OUR.

3.4. GIS mapping based nonparametric-geographically weighted spline regression model estimation for OUR data in Kalimantan

The mapping based on the estimation results of the NGWSR model is shown in Figure 2. The estimated value of the model may be utilized to determine the level of influence of variables on OUR. Figure 2 indicates the distribution of the estimator model or the coefficient values of the LFPR, Population Density, HDI, EYS and Regional Minimum Wages variables in 2019.

Figure 2. The mapping based on each estimator value of 56 regencies/cities in Kalimantan in 2019 (a) LFPR, (b) population, (c) HDI, (d) EYS, (e) regional minimum wage in 2019

The estimated figures may be used to identify the level of influence provided by predictor variables on the OUR data of 56 Regencies/Cities in Kalimantan. Referring to Figure 2(a), it was found that the
LFPR of 2019 in all 56 Regencies/Cities in Kalimantan likely provided a high influence on OUR. This is indicated by a dark color that is evenly distributed throughout the area. While in Figure 2(b), the total population of 2019 in all 56 Regencies/Cities in Kalimantan was found to likely have a low influence on OUR. This can be seen through the light color that is evenly distributed throughout the area. In addition, Figure 2(c) shows that EYS in all 56 Regencies/Cities in Kalimantan likely provide a low impact on OUR. This is indicated by the light color that is evenly distributed throughout the area. In Figure 2(d) shows that the EYS of 2019 in all 56 Regencies/Cities in Kalimantan tended to provide a high influence on OUR. This is clearly shown by the dark color that is evenly distributed throughout the area. Figure 2(e) shows that the Regional Minimum Wages of 2019 in all 56 Regencies/Cities in Kalimantan tended to provide a low influence on OUR. This is indicated by the light color that is evenly distributed throughout the area.

Conclusion

The NGWSR model was considered capable of addressing unknown regression curves and spatially non-stationary data. The results of this study successfully showed that contributing factors to the OUR of 56 Regencies/Cities in Kalimantan consisted of LFPR, Population Density, HDI, EYS and Regional Minimum Wage. Having regard to the value of the parameter estimator of the NGWSR model, Sukamara Regency was found to have the highest influence on LFPR and EYS. The OUR in Sukamara Regency was assumed to increase if the LFPR provided a high percentage value, and the low value was provided by EYS. The increase in OUR in Kotabaru Regency was influenced by the high population. A high population in an area will be capable of providing an effect on the high percentage of OUR. Moreover, based on HDI as a basic component of human resources according to quality of life, it was found that East Kutai Regency had a low quality of life, thus influencing OUR to be high. Furthermore, the minimum wage in Nunukan City showed the highest parameter estimator value among other regional estimators. Therefore, OUR in that region was likely to increase. This value was eventually able to provide information to the reviewer to pay attention to the predictor variables for each region of 56 Regencies/Cities in Kalimantan, because it was capable of providing a fairly large influence on increasing OUR.

Acknowledgments

The study work is facilitated by the support of Mulawarman University.

References

[1] B. Lu, M. Charlton, and A. S. Fotheringham, 2011, Geographically Weighted Regression using a non-Euclidean distance metric with a study on London house price data, (Procedia Environ. Sci., vol. 7), pp. 92–97, doi: 10.1016/j.proenv.2011.07.017.
[2] E. H. Koh, E. Lee, and K. K. Lee, 2020, Application of geographically weighted regression models to predict spatial characteristics of nitrate contamination: Implications for an effective groundwater management strategy, (J. Environ. Manage., vol. 268) p. 110646, doi: 10.1016/j.jenvman.2020.110646.
[3] C. A. S. F. M. E. C. Brunsdon, Geographically Weighted Regression, (Encycl. Geogr., vol. 28, no. 4), 2014, doi: 10.4135/9781412939591.n478.
[4] Sifriyani, S. H. Kartiko, I. N. Budiantara, and Gunardi, 2018, Development of nonparametric geographically weighted regression using truncated spline approach, (Songklanakarin J. Sci. Technol., vol. 40, no. 4), pp. 909–920, 2018, doi: 10.14456/sjst-psu.2018.98.
[5] Sifriyani, Haryatmi, I. N. Budiantara, and Gunardi, 2017, Geographically weighted regression with spline approach, (Far East J. Math. Sci., vol. 101, no. 6), pp. 1183–1196, 2017, doi: 10.17654/MS101061183.
[6] Sifriyani, I. N. Budiantara, S. H. Kartiko, and Gunardi, 2018, A new method of hypothesis test for truncated spline nonparametric regression influenced by spatial heterogeneity and application, (Abstr. Appl. Anal., vol. 2018), doi: 10.1155/2018/9769150.
[7] S. Sifriyani, 2019, Simultaneous Hypothesis Testing of Multivariable Nonparametric Spline Regression in the GWR Model, (Int. J. Stat. Probab., vol. 8, no. 4), p. 32, doi: 10.5539/ijsp.v8n4p32.

[8] S. Sifriyani, R. Ruslan, and S. F. H, 2019, Mapping and Analysis Factors of Affecting Productivity Tropical Rain Forests in East Kalimantan, (Mod. Appl. Sci., vol. 13, no. 10), p. 112, doi: 10.5539/mas.v13n10p112.

[9] Sifriyani, Ruslan, and F. H. Susanty, 2019, Evaluation of forest productivity and governance on the preservation of tropical rain forests in kalimantan using the ngwr-ts nonparametric geospatial method, (EurAsian J. Biosci., vol. 13, no. 2,) pp. 2373–2379.

[10] BPS Kalimantan Barat. Keadaan Angkatan Kerja. Pontianak: Badan Pusat Statistik Provinsi Kalimantan Barat. 2019.

[11] BPS Kalimantan Tengah. Keadaan Angkatan Kerja. Palangkaraya: Badan Pusat Statistik Provinsi Kalimantan Tengah. 2019a.

[12] BPS Kalimantan Selatan. Keadaan Angkatan Kerja. Banjarmasin: Badan Pusat Statistik Provinsi Kalimantan Selatan. 2019.

[13] BPS Kalimantan Timur. Keadaan Angkatan Kerja. Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur. 2019a.

[14] BPS Kalimantan Utara. Keadaan Angkatan Kerja. Tarakan: Badan Pusat Statistik Provinsi Kalimantan Utara. 2019.

[15] BPS Kalimantan Barat. Provinsi Kalimantan Barat dalam Angka. Pontianak: Badan Pusat Statistik Provinsi Kalimantan Barat. 2020a.

[16] BPS Kalimantan Barat. Proyeksi Penduduk Kabupaten/Kota. Pontianak: Badan Pusat Statistik Provinsi Kalimantan Barat. 2020b.

[17] BPS Kalimantan Selatan. Provinsi Kalimantan Selatan dalam Angka. Banjarmasin: Badan Pusat Statistik Provinsi Kalimantan Selatan. 2020.

[18] BPS Kalimantan Tengah. Statistika Ketenagakerjaan. Palangkaraya: Badan Pusat Statistik Provinsi Kalimantan Tengah. 2019b.

[19] BPS Kalimantan Tengah. Provinsi Kalimantan Tengah dalam Angka. Palangkaraya: Badan Pusat Statistik Provinsi Kalimantan Tengah. 2020a.

[20] BPS Kalimantan Tengah. Proyeksi Penduduk Kabupaten/Kota. Palangkaraya: Badan Pusat Statistik Provinsi Kalimantan Tengah. 2020b.

[21] BPS Kalimantan Timur. Indeks Pembangunan Manusia. Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur. 2019b.

[22] BPS Kalimantan Timur. Provinsi Kalimantan Timur dalam Angka. Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur. 2020.

[23] BPS Kalimantan Utara. Provinsi Kalimantan Utara dalam Angka. Tarakan: Badan Pusat Statistik Provinsi Kalimantan Utara. 2020.

[24] Dinas Ketenagakerjaan. Penetapan Upah Minimum. Pontianak: Dinas Ketenagakerjaan Kalimantan Barat. 2020.

[25] Dinas Ketenagakerjaan. Upah Minimum Regional. Palangkaraya: Dinas Ketenagakerjaan Kalimantan Tengah. 2020.