

Research Article

Bandwidth Selection in Geographically Weighted Regression Models via Information Complexity Criteria

Tuba Koç 

Cankiri Karatekin University, Faculty of Science, Statistics Department, Cankiri, Turkey

Correspondence should be addressed to Tuba Koç; tubakoc@karatekin.edu.tr

Received 9 March 2022; Revised 23 May 2022; Accepted 16 June 2022; Published 28 June 2022

Academic Editor: Ding-Xuan Zhou

Copyright © 2022 Tuba Koç. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The geographically weighted regression (GWR) model is a local spatial regression technique used to determine and map spatial variations in the relationships between variables. In the GWR model, the bandwidth is very important as it can change the parameter estimates and affect the model performance. In this study, we applied the information complexity (ICOMP) type criteria in the selection of fixed bandwidth for the first time in the literature. The ICOMP-type criteria use a complexity measure that measures how parameters in the model relate to each other. A real dataset example and a simulation study have been conducted. Results of the simulation demonstrate that GWR models created with the bandwidth selection by ICOMP-type criteria show superior performance. In addition, when the bandwidth is selected according to the ICOMP-type criteria and the GWR model is created for the actual total fertility rate data, it is seen that the spatial distribution of the total fertility rate estimates is quite compatible with the distribution of the actual total fertility rate. According to the results, ICOMP-type criteria can be used effectively instead of the classical criteria in the literature in the selection of bandwidth in the GWR model.

1. Introduction

Spatial regression analysis has become one of the important fields in the field of statistics in recent years. In studies where space is important, it is clear that classical statistical methods are insufficient to explain statistical change and estimate statistical inferences. Therefore, spatial statistical methods started to use in this field. Spatial statistical methods include spatial models that contain spatial information and consider the effects of locations on observations. The geographically weighted regression (GWR) method is a local spatial regression technique used to model relationships on geography [1]. GWR can generate estimates for other points where the location is known by the regression method and the reference points where the location and features are known on the geography. Unlike the classical regression model, the coefficients are not fixed in the GWR model and each spatial point has its own coefficients [2]. In the GWR approach, the neighborhood weight of all reference points adjacent to the regression point is determined to estimate the value of the parameters of a regression point. This neighborhood weight

is usually determined by the kernel function using the Euclidean distance. Bandwidth is the distance metric or number of neighbors used for each local regression equation, and changing the bandwidth is the most important parameter to consider in the GWR approach because it can change the coefficient estimates [3]. As the bandwidth value increases, the weights decrease and the local variation of the parameters decreases. So the regression equations turn into a general equation rather than a local one. As the bandwidth value decreases, the weights increase and the local variation of the parameters increases. But in this case, the equation may not give correct results because few reference points will be taken into account [4]. Studies to improve the accuracy of GWR approaches generally focus on the calibration of bandwidths [5–7]. These approaches mainly focus on finding the appropriate bandwidth values for the datasets and are not concerned with the spatial datasets, altitude and temporal stationary. The bandwidth value can be a fixed value or an adaptive value for the all dataset, depending on the distribution density of the locations in the data [8]. Cross-validation (CV), generalized cross-validation (GCV), Akaike

information criterion (AIC), and Bayesian information criterion (BIC) are among the methods used to find the optimal bandwidth value in the GWR method [9, 10]. Guo [4] constructed a forest plot with a clustered spatial model of tree positions to investigate the effects of different kernel functions and different bandwidths on the model performance and coefficient estimates of the GWR method. Cho et al. [6] selected bandwidths using the GWR method, CV, and smallest spatial error LaGrange multiplier test statistic calibration methods. Yacim and Boshoff [11] defined five different GWR models by choosing different cores and bandwidths in the residential data sample and compared the model performances. Koc and Akin [12] selected the bandwidth with the CV criterion and applied GWR models with different kernel functions at fixed bandwidth. Yuan et al. [13] examined the effects of AIC and their different bandwidths on GWR results. Hu et al. [14] introduced a two-dimensional bandwidth matrix for parameter estimation in the GWR model. Punzo et al. [15] examined local differences in the effects of the main sociodemographic, economic, and institutional determinants of land consumption by using the bandwidth adjusted AIC in the GWR method. It is clear that bandwidth selection has a strong influence on the descriptive and predictive power of GWR models. Generally, CV and AIC are used for bandwidth selection in literature.

In this study, we propose to use the information complexity (ICOMP) criteria proposed by Bozdoğan [16, 17] in the selection of bandwidth in the GWR model. ICOMP use an information theoretic measure of general model complexity based on Van Emden's generalized covariance complexity and the Kullback–Leibler distance [18, 19]. The purpose of ICOMP-type criteria is to achieve the optimal balance between the complexity and fit of a model. ICOMP aims to establish this balance by considering a complexity measure that measures how the parameters in the model are related to each other. Therefore, although it is a measure based on the Akaike Information Criterion, unlike AIC, it penalizes the covariance complexity of the model instead of directly penalizing the number of independent parameters. The use of ICOMP-type criteria in the selection of bandwidth in the GWR model will increase the confidence in the selected bandwidth due to its theoretical basis and will bring a new perspective to the literature.

The study is organized as follows. In Section 2, we present the model and methods. First, we explain geographically weighted regression and we applied the ICOMP-type criteria for the regression model. In Section 3, we present the simulation results regarding bandwidth selection using ICOMP-type. In Section 4, a real dataset example of the total fertility rate is present. There are conclusions obtained from this study in Section 5.

2. Materials and Methods

In this section, we introduced the geographically weighted regression method and information complexity criteria. First, the geographically weighted regression model is a local spatial regression technique that produces the predicted values for other points whose positions and properties are

known [1]. Unlike the linear regression model, the coefficients in the GWR model are not constant. The coefficients of each spatial point are created [2]. The GWR model is given by

$$y_i = \alpha_0(u_i, v_i) + \sum_{k=1}^m \alpha_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (1)$$

In equation (1), (u_i, v_i) is the latitude and longitude coordinates of the location in space i th, y_i is the dependent variable, x_i ($k = 1, 2, \dots, m$) is the independent variable, α_k is the coefficient of the GWR regression model, and ε_i is the error of i th location which is assumed to be independent and identically distributed normal random variable with mean zero and constant variance σ^2 .

The weighted least square methods provide a basis of estimating the GWR parameters. Parameters estimation of the GWR model are obtained as follows:

$$\hat{\alpha}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y, \quad (2)$$

where X is the matrix of independent variables and consists of $m + 1$ column, $Y = (Y_1, Y_2, \dots, Y_n)^T$ is the dependent variable matrix, and $W(i)$ is a diagonal matrix of w_{ij} values and is as shown as follows [1].

$$W(i) = \begin{bmatrix} w_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{in} \end{bmatrix}, \quad (3)$$

where w_{ij} is the neighborhood ratio between the regression point and the reference point. w_{ij} is calculated using the global model, box-car, exponential, Gaussian, bi-square, and tricube [20]. The Gaussian function is commonly used with Gaussian kernel function [21, 22], and w_{ij} is calculated as in the following equation.

$$w_{ij} = \exp\left[-\frac{1}{2} \left(\frac{d_{ij}}{bw}\right)^2\right], \quad (4)$$

where bw is the bandwidth value, d_{ij} is the distance between the regression point i and the reference point j , and d_{ij} is usually the Euclidean distance and is calculated as shown in the following equation, where u and v are the point coordinates.

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}. \quad (5)$$

The value of the bw bandwidth parameter can be constant for a whole model in the GWR model or it can be variable according to the point density in the location. The optimal value of bandwidth can be determined by cross-validation (CV), generalized cross-validation (GCV), Akaike information criterion (AIC), and corrected Akaike information criterion (AICc) methods [23]. The cross-validation criterion is given as follows [24].

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(bw))^2, \quad (6)$$

where $\hat{y}_{\neq i}(bw)$ is the fitted value of y_i by omitting the i th point from the process [1].

The generalized cross-validation criterion for GWR is defined as [25]

$$GCV = n \sum_{i=1}^b \frac{(y_i - \hat{y}_i(bw))^2}{(n - \text{tr}(S))^2}, \quad (7)$$

where $\hat{y}_i(bw)$ is the fitted value of y_i using a bandwidth of bw , n is the sample size, and $\text{tr}(S)$ denotes the trace of the hat matrix [1].

The Akaike information criterion and corrected Akaike information criterion for GWR is defined as

$$\begin{aligned} AIC &= 2n \log(\hat{\sigma}) + n \log(2\pi) + n + \text{tr}(S), \\ AICc &= 2n \log(\hat{\sigma}) + n \log(2\pi) + n \left\{ \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right\}, \end{aligned} \quad (8)$$

where n is the sample size, $\hat{\sigma}$ is the estimated standard deviation of the error term, and $\text{tr}(S)$ denotes the trace of the hat matrix [26].

Second, the information complexity (ICOMP) criteria are a measure developed by Bozdoğan [27] on the basis of AIC. Unlike the AIC-based information criterion, ICOMP approximates the sum of the two Kullback and Leibler [28] distances that measure the model's lack of fit and model complexity in a criterion function using an entropic measure of the estimated covariance matrix of the model parameters. Thus, the concept of model complexity takes into account not only the number of free parameters in the model but also the interdependence of parameter estimates. Therefore, a general model selection criterion can be provided by ICOMP by understanding the relational structure between parameter estimates in the selected model [29]. ICOMP-type criteria provide the most appropriate balance between the complexity of a model and the goodness-of-fit [30]. ICOMP criteria can be defined in several formulations. The formulations of the information criteria are as follows [31].

$$\begin{aligned} AIC &= -2 \log L(\hat{M}) + 2k, \\ CICOMP &= -2 \log L(\hat{M}) + k[\log(n) + 1] + 2C(\hat{\Sigma}_{\text{model}}), \\ ICOMP_{\text{PEU}} &= -2 \log L(\hat{M}) + k + 2C(\hat{\Sigma}_{\text{model}}), \\ ICOMP_{\text{PEULN}} &= -2 \log L(\hat{M}) + k + \log(n)C(\hat{\Sigma}_{\text{model}}), \end{aligned} \quad (9)$$

where n is the sample size, $L(\hat{M})$ is the maximized likelihood function, k is the number of variables, C is a real-valued complexity measure, and $\hat{\Sigma}_{\text{model}} = \widehat{\text{Cov}}(\hat{M})$ is the estimated covariance matrix of the parameter vector of the models [29].

3. Simulation Study

A simulation study was conducted to understand the performance of the information complexity criteria on bandwidth. Simulation design is in the same way as follows:

$$\begin{aligned} Y &= X_1 + 1.5X_2 + 3X_3 + \varepsilon, \\ \text{lon} &\sim N(533200, 1000), \\ \text{lat} &\sim N(159400, 1000), \end{aligned} \quad (10)$$

where $X_1 \sim N(0, 1)$, $X_2 \sim N(2, 4)$, $X_3 \sim N(5, 8)$, and $\varepsilon \sim N(0, 1)$ have been generated. Lon and lat are the spatial coordinates of the locations. Sample size is $n = 300$, and kernel function is taken as Gaussian. In Table 1, the values of the optimal bandwidth selection by different methods are given.

The ICOMP-type information criteria selected the bandwidth value as 2811.487. It was found that ICOMP-PEULN performs better with the lowest information criterion in the selection of bandwidth. While the GWR model was created with fixed bandwidth selection, the GWR model established with ICOMP-type was GWR-ICOMP-type, the GWR model established with CV bandwidth selection was taken as GWR-CV, the GWR model established with GCV bandwidth selection was taken as GWR-GCV, the GWR model established with AIC bandwidth selection was taken as the GWR-AIC model, and the GWR model established with AICc bandwidth selection was taken as GWR-AICc. Performance evaluations of the models are as given in Table 2.

In Table 2, the GWR-ICOMP-type model was performed the best with the highest $\text{Adj.}R^2$ value of 0.9901 and the lowest information criterion $AIC = 169.0912$.

4. Real Data Application

In this study, total fertility rate data were used for 81 provinces of Turkey in 2020. Total fertility rate refers to the average number of children a woman can have in the 15–49 age group. The data obtained are available at [32]. A set of 6 continuous variables are used in this study and described as dependent variable (Y): total fertility rate, and independent variables: gross domestic product (GDP-per city) (x_1), mean age of mother by provinces, (2009–2020) (x_2), number of illiterate (x_3), number of women with higher education (x_4), unemployment rate (% of GDP) (x_5). Furthermore, the coordinates (longitude, latitude) for the 81 cities in the Turkey (see [33]) are used to fit the GWR. The spatial distribution of the total fertility rate in Turkey for 2020 is shown in Figure 1.

The provinces with the highest total fertility rate in Turkey are Şanlıurfa with 3.71 which are shown in red on the map. This province was followed by Şırnak with 3.22 and Ağrı and Siirt with 2.88. The province with the lowest total fertility rate was Karabük with 1.29, which is shown in dark blue on the map. This province was followed by Zonguldak and Kütahya with 1.31. First, factors affecting total fertility rates in Turkey are modeled with the multiple regression method. Furthermore, the GWR model is used to determine whether there is an effect of locations.

Table 3 provides the coefficients of multiple regression models. According to the multiple regression models, gross domestic product, mean age of mother by provinces, and

TABLE 1: Optimal bandwidth for different criteria.

| Criteria | Score | Optimal bandwidth |
|------------|-----------------|-------------------|
| CV | 96.6466 | 15390.96 |
| | 96.63455 | 13963.07 |
| | 96.62347 | 14520.52 |
| | 96.62348 | 14547.6 |
| | 96.62344 | 14513.78 |
| GCV | 96.83989 | 15421.74 |
| | 96.7201 | 14535.04 |
| | 96.24801 | 13907.22 |
| | 95.56062 | 14387.58 |
| | 95.36734 | 14325.1 |
| AIC | 282.6431 | 17774.06 |
| | 282.5941 | 15435.91 |
| | 282.5786 | 16037.34 |
| | 282.5785 | 16215.42 |
| | 282.5783 | 16129.28 |
| | 270.54 | 15524.6 |
| AICc | 269.842 | 15314.5 |
| | 269.42 | 14716.3 |
| | 268.946 | 16912.7 |
| | 265.407 | 15149.1 |
| CICOMP | 776.6168 | 18805.38 |
| | 502.4187 | 11652.69 |
| | 350.805 | 7232.089 |
| | 256.1577 | 4500.007 |
| | 142.1605 | 2811.487 |
| ICOMPPEU | 1389.638 | 18805.38 |
| | 771.4441 | 11652.69 |
| | 450.8292 | 7232.089 |
| | 285.8024 | 4500.007 |
| | 138.8512 | 2811.487 |
| ICOMPPEULN | 514.2803 | 18805.38 |
| | 372.1259 | 11652.69 |
| | 285.3836 | 7232.089 |
| | 217.7516 | 4500.007 |
| | 116.404 | 2811.487 |

Bold values indicate values with the lowest information criterion in the selection of bandwidth.

TABLE 2: Model performance evaluation.

| Model | Adj. R^2 | AIC | AICc |
|----------------|---------------|-----------------|-----------------|
| GWR-CV | 0.9365 | 270.1586 | 282.6757 |
| GWR-GCV | 0.9374 | 269.8887 | 282.3933 |
| GWR-AIC | 0.9272 | 271.3406 | 282.5783 |
| GWR-AICc | 0.9353 | 268.9736 | 280.1133 |
| GWR-ICOMP-type | 0.9901 | 169.0912 | 256.1141 |

Bold values indicate values that give the best model performance with the highest Adj and the lowest information criteria.

number of women with higher education and unemployment rate had a significant effect on total fertility rates. Testing the goodness-of-fit of the GWR model is important for experimental analysis. This test is called the global test of nonstationary [1].

In Table 4, the nonstationary global test result fits the GWR model to the total fertility rate data. Therefore, we can apply the GWR method to the total fertility rate data. The selection of optimal bandwidths in the GWR model for the total fertility rate data is given in Table 5.

The $ICOMP_{PEULN}$ criterion, which has the lowest information criteria among the ICOMP-type criteria, has chosen the optimal bandwidth. GWR models were found significant in the multiple regression models and selected the fixed bandwidth values.

These models are the GWR-CV model with a bandwidth value of 1.6477 for the CV score, the GWR-GCV model with a bandwidth value of 1.6526 for the GCV score, the GWR-AIC model with bandwidth value of 16.0042, the GWR-AICc model with bandwidth value of 14.9964, and the

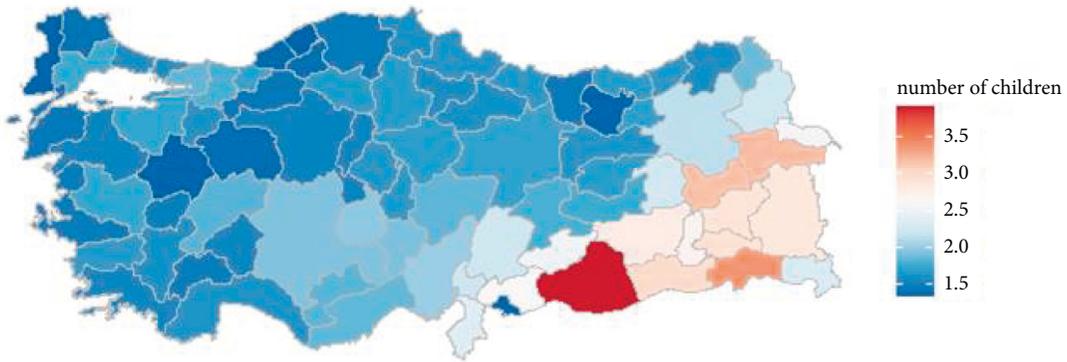


FIGURE 1: Spatial distribution of total fertility rate in Turkey for 2020.

TABLE 3: Coefficients for the multiple regression models.

| Coefficient | Estimate | Std. error | <i>t</i> value | Pr (> <i>t</i>) |
|-------------|-------------|-------------|----------------|--------------------|
| Intercept | 7.65 E + 00 | 1.95 E + 00 | 3.924 | 0.0001 |
| x1 | -1.14E-04 | 2.44E-05 | -4.678 | 0.0000 |
| x2 | -1.90E-01 | 6.90E-02 | -2.753 | 0.0074 |
| x3 | 1.77E-06 | 1.35E-06 | 1.308 | 0.1950 |
| x4 | 8.42E-07 | 3.77E-07 | 2.236 | 0.0283 |
| x5 | 3.47E-02 | 9.68E-03 | 3.586 | 0.0005 |

Bold values are *p* values ≤ 0.05 and are statistically significant.

TABLE 4: Analysis of variance for goodness-of-fit of the GWR model.

| Source | Df | Sum sq. | Mean sq. | F | <i>P</i> value |
|-----------------|--------|---------|----------|--------|----------------|
| OLS residual | 5 | 11.2918 | | 6.8976 | 0.00000 |
| GWR improvement | 30.058 | 9.2435 | 0.3075 | | |
| GWR residuals | 45.924 | 2.0483 | 0.0445 | | |

TABLE 5: Optimal bandwidths for the total fertility rate data.

| Criteria | Score | Optimal bandwidth |
|----------|-------------|-------------------|
| CV | 7.44 | 4.3735 |
| | 5.84 | 2.7100 |
| | 5.10 | 1.6819 |
| | 7.04 | 1.0465 |
| | 5.09 | 1.6477 |
| GCV | 7.45 | 4.3835 |
| | 5.83 | 2.7072 |
| | 5.10 | 1.6785 |
| | 7.06 | 1.0504 |
| | 5.08 | 1.6526 |
| AIC | 1.89 | 16.2104 |
| | 2.13 | 16.3004 |
| | 1.98 | 16.0055 |
| | 2.00 | 16.0056 |
| | 1.99 | 16.0042 |
| AICc | 1.69 | 14.5188 |
| | 1.81 | 13.8984 |
| | 1.85 | 14.9626 |
| | 1.98 | 15.8659 |
| | 1.86 | 14.9964 |

TABLE 5: Continued.

| Criteria | Score | Optimal bandwidth |
|------------------------|----------------|-------------------|
| CICOMP | 118.14 | 2.7100 |
| | 5.66 | 1.6819 |
| | -99.25 | 1.0465 |
| | -237.05 | 0.6538 |
| | -508.62 | 0.4111 |
| ICOMP _{PEU} | 40.98 | 2.7100 |
| | -38.38 | 1.6819 |
| | -128.35 | 1.0465 |
| | -260.45 | 0.6538 |
| | -530.64 | 0.4111 |
| ICOMP _{PEULN} | 223.30 | 2.7100 |
| | 31.53 | 1.6819 |
| | -109.15 | 1.0465 |
| | -260.58 | 0.6538 |
| | -535.46 | 0.4111 |

TABLE 6: Model performance evaluation.

| Model | Adj. R^2 | AIC | AICc |
|----------------|---------------|------------------|------------------|
| GWR-CV | 0.8759 | -40.97521 | 19.3678 |
| GWR-GCV | 0.8776 | -41.0572 | 19.3291 |
| GWR-AIC | 0.8064 | -25.11549 | 16.0042 |
| GWR-AICc | 0.8244 | -26.2959 | 15.2857 |
| GWR-ICOMP-type | 0.9993 | -456.2319 | -28271.21 |

Bold values are the values that give the best model performance with the highest Adj and the lowest information criteria.

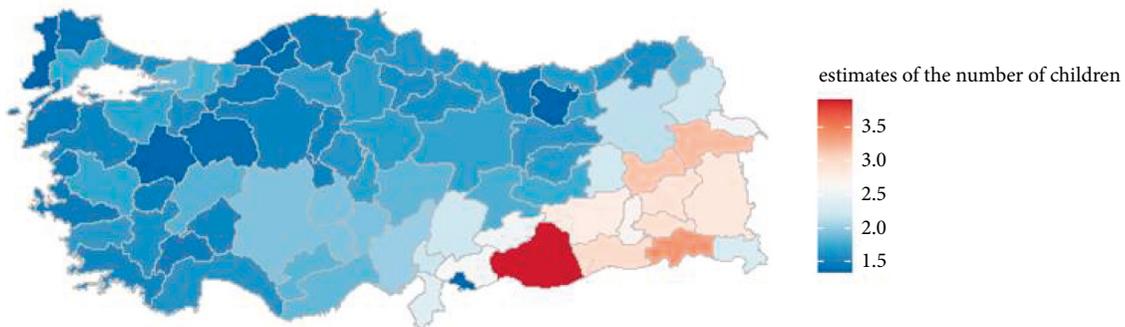


FIGURE 2: Spatial distribution of total fertility rate estimates in Turkey for 2020.

GWR-ICOMP-type with a bandwidth value of 0.4111. Model performances are as given in Table 6.

In Table 6, it is seen that the GWR-ICOMP-type model performed the best result with the highest Adj. R^2 and lowest model information criteria values. Spatial distributions of total fertility rate estimates are shown in Figure 2.

In Figure 2, after selecting the bandwidth with ICOMP-type, the spatial distribution of the total fertility rate estimates in the GWR model is given. In Figure 2, it is seen that the distribution of the total fertility rate estimates is quite compatible with the distribution of the actual total fertility rate shown in Figure 1.

5. Conclusions

The selection of bandwidth in the GWR model is very important to increase the efficiency and accuracy of the model, and the selection of bandwidth can be considered as a model selection problem. When bandwidth is large, more data points will be included in the regression. Thus, the variance will be small, while the deviation will be large. If a small bandwidth is used, the regression will be confined to a local area and parameter estimates will depend on observations close to the regression point. Therefore, the variance of parameter estimates will increase, but the bias will be

small and more anomalies can be discovered. In this study, we applied ICOMP-type criteria in the selection of bandwidth for the first time in literature. We found in the simulation design that the bandwidths selected with the ICOMP-type criteria increase the model performance and prediction accuracy of the GWR model. When GWR models with different bandwidths are established for actual total fertility rate data, it is concluded that the best performance is the GWR-ICOMP-type model with the highest R^2 and lowest information criteria. In addition, when the ICOMP- $PEULN$ criterion chose the most suitable model with the smallest information criterion in fixed bandwidth. The spatial distribution of the total fertility rate estimates in the GWR-ICOMP-type model seems to be quite compatible with the actual total fertility rate distribution. As a result, the use of ICOMP-type criteria in the selection of bandwidth can improve the GWR model in terms of prediction accuracy and model performance.

Data Availability

The data used to support this study are available from <https://data.tuik.gov.tr/>.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] A. S. Fortheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley and Song, 2003.
- [2] B. Lu, W. Yang, Y. Ge, and P. Harris, "Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths," *Computers, Environment and Urban Systems*, vol. 71, pp. 41–57, 2018.
- [3] G. Kauermann and J. Opsomer, "Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models," *Journal of Computational & Graphical Statistics*, vol. 13, no. 1, pp. 66–89, 2004.
- [4] L. Guo, Z. Ma, and L. Zhang, "Comparison of bandwidth selection in application of geographically weighted regression: a case study," *Canadian Journal of Forest Research*, vol. 38, no. 9, pp. 2526–2534, 2008.
- [5] X. Zhang, M. L. King, and R. J. Hyndman, "A Bayesian approach to bandwidth selection for multivariate kernel density estimation," *Computational Statistics & Data Analysis*, vol. 50, no. 11, pp. 3009–3031, 2006.
- [6] S. H. Cho, D. M. Lambert, and Z. Chen, "Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data," *Applied Economics Letters*, vol. 17, no. 8, pp. 767–772, 2010.
- [7] B. Lu, C. Brunsdon, M. Charlton, and P. Harris, "Geographically weighted regression with parameter-specific distance metrics," *International Journal of Geographical Information Science*, vol. 31, no. 5, pp. 982–998, 2017.
- [8] P. E. Bidanset and J. R. Lombard, "The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal," *Journal of Property Tax Assessment and Administration*, vol. 10, pp. 5–14, 2014.
- [9] X. Zhang, M. L. King, and R. J. Hyndman, "A Bayesian approach to bandwidth selection for multivariate kernel density estimation," *Computational Statistics & Data Analysis*, vol. 50, no. 11, pp. 3009–3031, 2006.
- [10] P. E. Bidanset, J. R. Lombard, P. Davis, M. McCord, and W. J. McCluskey, "Further evaluating the impact of kernel and bandwidth specifications of geographically weighted regression on the equity and uniformity of mass appraisal models," *Advances in Automated Valuation Modeling*, vol. 865, pp. 191–199, 2017, *Advances in Automated Valuation Modeling*.
- [11] 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, 2019.
- [12] T. Koc and P. Akin, "Comparison of kernel functions in geographically weighted regression model: suicide data as an application," *Acta Infologica*, vol. 5, pp. 7–8, 2021.
- [13] Y. Yuan, M. Cave, H. Xu, and C. Zhang, "Exploration of spatially varying relationships between Pb and Al in urban soils of London at the regional scale using geographically weighted regression (GWR)," *Journal of Hazardous Materials*, vol. 393, 2020.
- [14] X. Hu, Y. Lu, H. Zhang, H. Jiang, and Q. Shi, "Selection of the bandwidth matrix in spatial varying coefficient models to detect anisotropic regression relationships," *Mathematics*, vol. 9, no. 18, pp. 23–43, 2021.
- [15] G. Punzo, R. Castellano, and E. Bruno, "Using geographically weighted regressions to explore spatial heterogeneity of land use influencing factors in Campania Southern Italy," *Land Use Policy*, vol. 112, Article ID 105853, 2022.
- [16] H. Bozdogan, "Akaike's information criterion and recent developments in information complexity," *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 62–91, 2000.
- [17] H. Bozdogan, *Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2003.
- [18] M. H. van Emden, "An analysis of complexity," *Mathematics Centrum Tracts*, vol. 35, 1971.
- [19] S. Kullback, *The Kullback-Leibler Distance*, Springer, Berlin/Heidelberg, Germany, 1987.
- [20] I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris, "GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models," *Journal of Statistical Software*, vol. 63, no. 17, pp. 1–50, 2013, arXiv preprint arXiv: 1306.0413. 2013.
- [21] T. Hu, J. Fan, Q. Wu, and D. X. Zhou, "Regularization schemes for minimum error entropy principle," *Analysis and Applications*, vol. 13, no. 04, pp. 437–455, 2015.
- [22] X. Guo, L. Li, and Q. Li, "Modeling interactive components by coordinate kernel polynomial models," *Mathematical Foundations of Computing*, vol. 3, no. 4, pp. 263–277, 2020.
- [23] M. Tasyurek and M. Celik, "RNN-GWR: a geographically weighted regression approach for frequently updated data," *Neurocomputing*, vol. 399, pp. 258–270, 2020.
- [24] A. W. Bowman, "An alternative method of cross-validation for the smoothing of density estimates," *Biometrika*, vol. 71, no. 2, pp. 353–360, 1984.
- [25] P. Wahba and G. Wahba, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, 1978.

- [26] D. C. Hoaglin and R. E. Welsch, "The hat matrix in regression and ANOVA," *The American Statistician*, vol. 32, no. 1, pp. 17–22, 1978.
- [27] H. Bozdogan, "ICOMP: a new model selection criterion," in *Classification and Related Methods of Data Analysis*, H. H. Bock, Ed., pp. 599–608, Elsevier Science Publishers, Amsterdam, Netherlands, 1988.
- [28] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [29] E. Kartal Koc and H. Bozdogan, "Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function," *Machine Learning*, vol. 101, no. 1-3, pp. 35–58, 2015.
- [30] E. Pamukçu, H. Bozdogan, and S. Çalık, "A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification," *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1–14, 2015.
- [31] H. Koç, E. Dündar, S. Gümüştekin, T. Koç, and M. A. Cengiz, "Particle swarm optimization-based variable selection in Poisson regression analysis via information complexity-type criteria," *Communications in Statistics - Theory and Methods*, vol. 47, pp. 5298–5306, 2018.
- [32] <https://data.tuik.gov.tr/>.
- [33] <https://www.mapsofworld.com>.