IJEGEO

*International Journal of Environment and Geoinformatics 11(2):061-066 (2024)*

**Research Article**

# Prediction of Precipitation using Multiscale Geographically Weighted Regression

**Murat Taşyürek[1], Mete Çelik[2], Ali Ümran Kömüşcü[3], Filiz Dadaşer-Çelik[4*]**

[1] Department of Computer Engineering, Kayseri University, Kayseri, Türkiye
[1] Department of Computer Engineering, Erciyes University, Kayseri, Türkiye
[3] Department of Geography, Ankara Hacı Bayram Veli University, Ankara, Türkiye
[4] Department of Environmental Engineering, Erciyes University, Kayseri, Türkiye

* Corresponding author: F.Dadaşer-Çelik
E-mail: fdadaser@erciyes.edu.tr.edu.tr

Received 01.12.2023
Accepted 11.06.204

**How to cite:** Taşyürek et al., (2024). Prediction of Precipitation using Multiscale Geographically Weighted Regression, *International Journal of Environment and Geoinformatics (IJEGEO)*, 11(2): 061-066. doi. 10.30897/ijegeo.1399172

## Abstract

Prediction of precipitation at locations that lack meteorological measurements is a challenging task in hydrological applications. In this study, we aimed to demonstrate the potential use of multiscale geographically weighted regression (MGWR) method used to predict precipitation based on relevant meteorological parameters. Geographically weighted regression (GWR) is a regression technique proposed to explore spatial non-stationary relationships. Compared to the linear regression technique, GWR considers the dynamics of local behaviour and therefore provides an improved representation of spatial variations in relationships. Multiscale geographically weighted regression (MGWR) is a modified version of GWR that examines multiscale processes by providing a scalable and flexible framework. In this study, the MGWR method was used to predict precipitation, which is an essential problem not only in meteorology and climatology but also in many other disciplines, such as geography and ecology. A meteorological dataset including elevation, precipitation, air temperature, air pressure, relative humidity, and cloud cover data from 184 stations in Türkiye was used, and the performance of the MGWR was assessed in comparison with that of global regression and classical GWR based on root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), and correlation coefficient (R) calculated between measured and simulated precipitation. The RMSE values calculated for the global regression, GWR, and MGWR methods were 4.64 mm, 3.53 mm, 2.9 mm, respectively. NSE values and R values were -0.63, 0.03, 0.35 and 0.04, 0.42, 0.63, respectively. These results demonstrated that the MGWR model outperformed other approaches in precipitation prediction.

**Keywords:** Multiscale Geographically Weighted Regression, MGWR, Data Mining, Precipitation

## Introduction

Meteorological data have a wide range of applications across various domains, such as meteorology, hydrology, water resource planning, forestry, ecology, agriculture, and climate change analysis. The magnitude and variety of data on meteorological parameters, such as precipitation, air temperature, and evaporation, are diverse and growing every day. Precipitation is the most important meteorological parameter affecting the water cycle and is an important aspect of the Earth's climate system. Precipitation ensures that water moves between the atmosphere, land, and the oceans. Its distribution and availability also influence the local and regional climate by causing extreme hydrological conditions, such as floods and droughts. Precipitation also plays a vital role in the maintenance of Earth's water supply. Thus, accurate and reliable rainfall measurements are essential for water resource management and agricultural practices. Both meteorologists and decision makers always demand a higher spatial and temporal accuracy of precipitation. There may be hundreds of precipitation stations in a region, and precipitation data are collected at particular points (stations). Although the number of stations is high, the data received from these stations can represent the points where they are located, and the level of representation for the points without station data may be low/insufficient. Today, applications in many different areas, for example, in water resource planning, require numerical solutions for spatially scattered parameters. Therefore, spatial analysis using point data creates large uncertainties for many hydrological applications.

Approximating precipitation over locations where direct field observations are not available is one of the most important and challenging tasks in meteorology and hydrology. The spatial characteristics of precipitation have been examined using various hydrological research techniques (Mays, 2001). Classical techniques include statistical analysis of precipitation data (e.g., arithmetic mean), where the spatial characteristics of precipitation data are ignored. In recent years, advances in geographical information systems have increased the use of different spatial interpolation techniques (such as kriging and inverse distance weighing). Precipitation data have spatial and temporal characteristics; therefore, the inclusion of these characteristics in data analysis can improve the accuracy of precipitation prediction (Ashiq et al., 2010; Brunsdon et al., 2001; Celik et al., 2014; Diodato, 2005). Meteorological events occur on different spatial and temporal scales. Unfortunately, most meteorological measurements are obtained at or near settlements because of access to energy sources and operational flexibility. However, hydrological processes occur at the basin level, where hydrological inputs are needed at desired temporal and spatial scales. Therefore, to take into account different scales, in this study, we examine the applicability of multiscale geographically weighted regression (MGWR) (Fotheringham et al., 2017) for the prediction of precipitation in ungauged locations, where there are no meteorological measurements.

GWR was used to model spatially varying relationships (Fotheringham et al., 2002; Lu et al., 2017). The GWR method has a wide range of applications in various fields. It has been used to predict the soil total nitrogen of a region (Wang et al., 2013), predict cancer risks (Hsueh et al., 2012), analyze dengue fever cases (Zhang et al., 2011), and model hotel room prices (Wei and Qi, 2012), etc.

There are different variations of the GWR model in the literature. One of them is the mixed geographically weighted regression, which takes some of the explanatory variable coefficients as either constant or spatially varying, and is proposed to analyze spatial non-stationarity (Brunsdon et al., 2001; Leong and Yue, 2017). A conditional GWR was proposed to select the bandwidth for each variable (Dong et al., 2018). A geographically weighted ordinal regression (GWOR) model was proposed to handle ordinal categorical variables (da Silva and de Oliveira Lima, 2017). Geographically weighted beta regression (GWBR) was proposed to model spatially varying rates and proportions (A. Fotheringham et al., 2015). Geographically and temporally weighted regression (GTWR) was proposed to handle the temporal dimensions of the data (Ma et al., 2018). The FastGTWR method was proposed to accelerate the GTWR model (Taşyürek and Celik, 2021). RNN-GWR algorithm was proposed to handle frequently updated data (Tasyurek and Celik, 2020). Grid-based GWR (Harris et al., 2010), distributed GWR (Hung Tien et al., 2016), and parallel GWR of FastGWR (Li et al., 2018) were proposed to accelerate GWR models. 4D-GWR was proposed to consider the altitude and temporal and spatial dimensions of the dataset (Tasyurek and Celik, 2022). In this study, we propose the use of the MGWR (Fotheringham et al., 2017) method to analyze meteorological data. The MGWR determines the bandwidth of each scale to improve the prediction accuracy.

## Method

### Geographically Weighted Regression (GWR)

In this section, the details of GWR and MGWR methods are described. The GWR, which is based on standard multiple-parameter regression, aims to analyze locally spatially varying relationships (Fotheringham et al., 2002). In GWR, the coefficients are not constant (Eq.1).

$$y_i = \sum_{j=0}^{m} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i \qquad \text{(Eq.1)}$$

In Eq. (1), ($u_i$, $v_i$) represents the location of the regression point $i$, $y_i$ represents the dependent variable, $x_{ij}$ ($j$ = 1, ..., $m$) represents the independent variable, $\beta_j$ represents the coefficient of regression, and $\varepsilon_i$ represents the error.

In GWR, there is only one bandwidth and its optimal value is calibrated using $AIC_c$ (Eq.2). In trials, the bandwidth, which minimizes $AIC_c$ is selected as the optimal bandwidth.

$$AIC_c = 2nln(\hat{\sigma}) + nln(2\pi) + n\frac{n+tr(S)}{n-2-tr(s)} \qquad \text{(Eq.2)}$$

In the equation, $n$ is the number of observations, $\hat{\sigma}$ is the estimated standard deviation of the error term, and $tr(S)$ refers to the trace of the hat matrix. In MGWR, there is more than one bandwidth to be optimized and therefore, a different strategy is needed to calibrate the bandwidths.

### Multiscale Geographically Weighted Regression (MGWR)

Multiscale geographically weighted regression (MGWR), an extension of the classical GWR model, was proposed by Fotheringham et al. (A.S. Fotheringham et al., 2017). Classical GWR assumes that the spatial scale is fixed for all processes, and a single optimal bandwidth is found for the processes. In contrast, MGWR assumes that different spatial scales can be used for different processes, and a bandwidth vector is used whose elements represent the spatial scales of different processes (Eq. 3) (A.S. Fotheringham et al., 2002).

$$y_i = \sum_{j=0}^{m} \beta_{bwj}(u_i, v_i)x_{ij} + \varepsilon_i \qquad \text{(Eq.3)}$$

In Eq. (2), the regression coefficient $\beta_{bwi}$ includes $bwj$, which represents the bandwidth used for determining the $j^{th}$ conditional relationship.

To calibrate the MGWR model, a back-fitting algorithm is used. It maximizes log-likelihood and is generally used in generalized additive models (GAM). In the algorithm, the term $\beta_{bwi}\, x_j$ in MGWR is represented as the $j^{th}$ additive term of $f_j$ of GAM, as shown in (Eq.4).

$$y = \sum_{j=0}^{m} f_j + \varepsilon \qquad \text{(Eq.4)}$$

In the first iteration of the back-fitting algorithm, it produces the optimal bandwidths (*bw0, bw1, ..., bwm*) by regressing on every $x$ variable using GWR one by one. That is in one of the iterations, the GWR is run for $m$ times to generate $m$ optimal bandwidths for each $x$ variable. The iteration continues until a stopping criterion is satisfied.

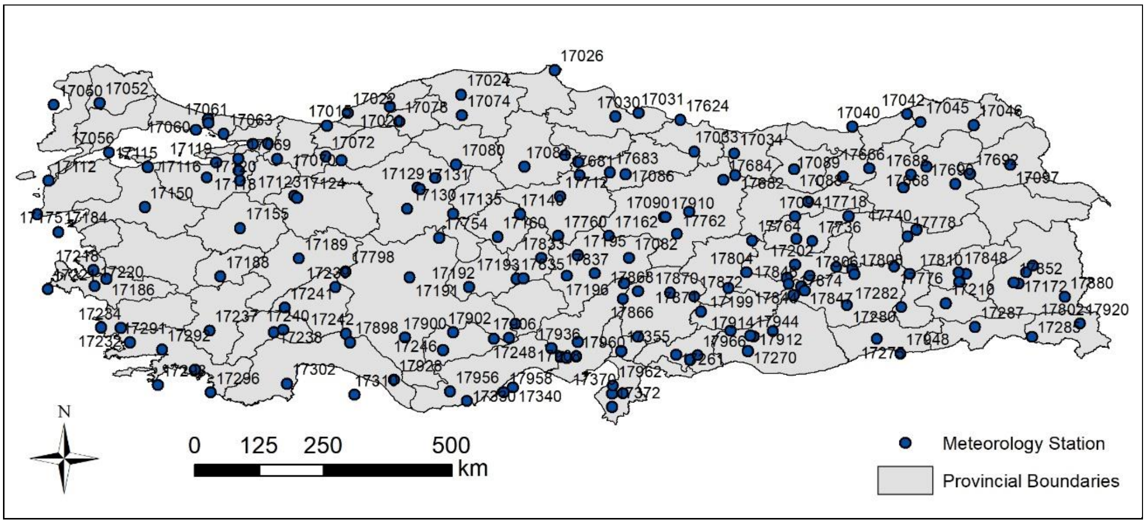


Fig. 1. Stations used

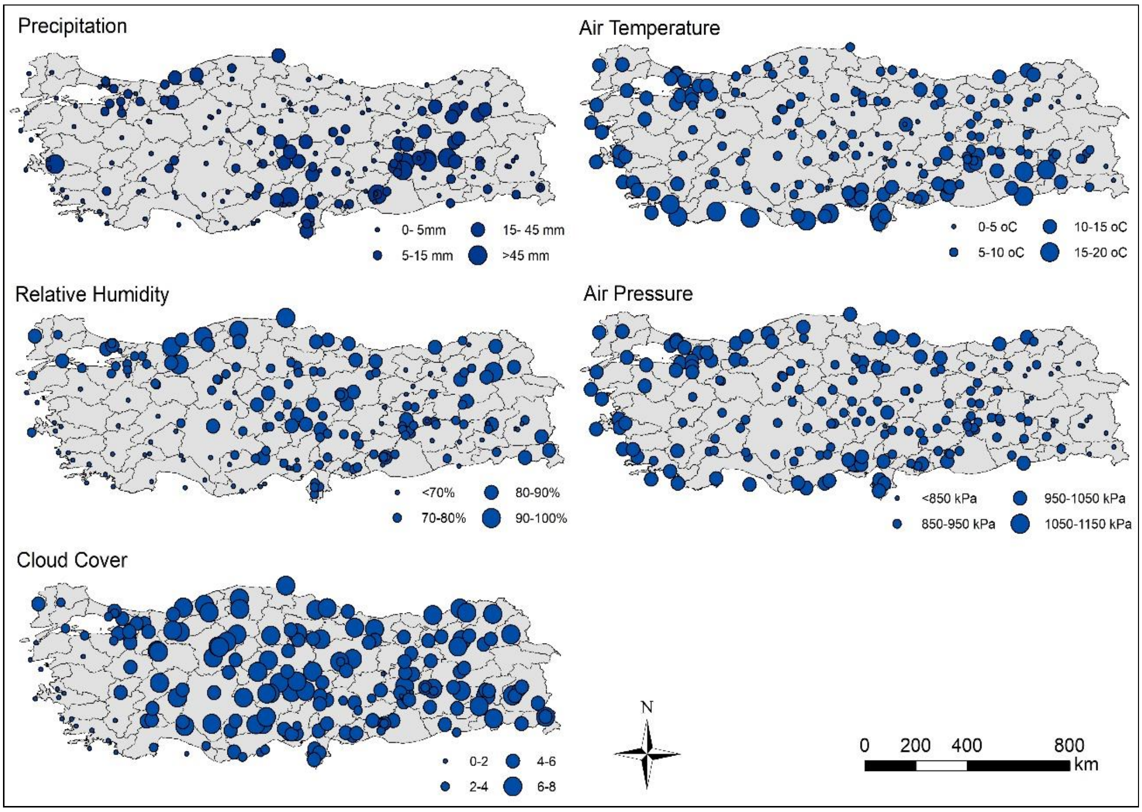


Fig. 2. Training Data Used

**Experimental Evaluation**

The performance of the MGWR model was compared with those of global regression and classical GWR on a meteorological dataset. The comparison was done based on root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), and correlation coefficient (R) calculated between measured and simulated precipitation. In the remainder of this section, the dataset used is introduced, model settings are discussed, and experimental evaluations are presented.

**Dataset**

The data used in this study was obtained from the Turkish State Meteorological Service (Fig. 1, Fig 2). The dataset included data from 184 stations on the daily timescale. Altitude, air pressure, relative humidity, air temperature, and cloud coverage were used as independent variables, and the value of precipitation was estimated using these variables. Data were normalized before the experimental evaluation. The methods were trained by using data from data from 30.03.2010 and tested using data from 01.04.2010.

The statistical characteristics of the meteorological data are provided in Table 1. Precipitation values ranged from 0 to 143 mm while air temperature values were between 0 and 17 ºC. The average air pressure was 924 kPa, the average relative humidity was 70% and the average cloud cover was 5.

Table 1. Statistical Characteristics of the Data Used

| Parameter | Minimum | Maximum | Average | Standart Deviation |
|---|---|---|---|---|
| Precipitation (mm) | 0 | 147 | 8 | 15 |
| Air Pressure (kPa) | 771 | 1020 | 924 | 65 |
| Relative Humidity (%) | 40 | 99 | 70 | 10 |
| Air Temperature (ºC) | 0 | 17 | 9 | 4 |
| Cloud Cover | 0 | 7 | 5 | 2 |

**Model settings**

This study aimed to predict the precipitation of a location using the altitude and meteorological parameters of average air pressure, average relative humidity, average temperature, and cloud coverage. The precipitation estimation model can be expressed as follows in mode (Eq.5):

$Total\_precipitation_i = \beta_{0i} + \beta_{1i}Altitude_i + \beta_{2i}Air\_Pressure_i + \beta_{3i}Rel\_Humidity + \beta_{4i}Temperature_i + \beta_{5i}cloud\_coverage_i$. (Eq. 5)

In the equation, $i$ is the regression point and $\beta$s are the coefficients.

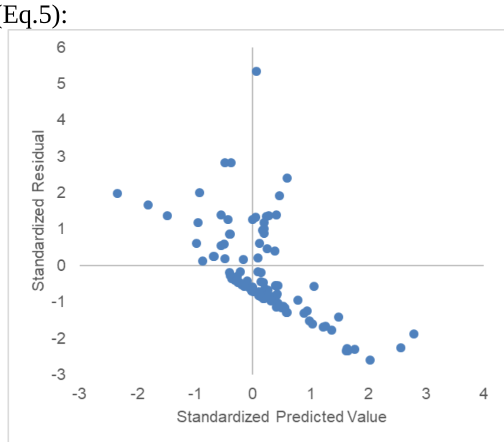

Fig. 3. Global Regression Residuals

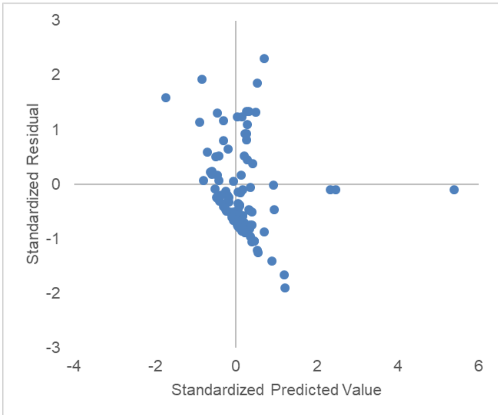

Fig. 4. GWR Residuals

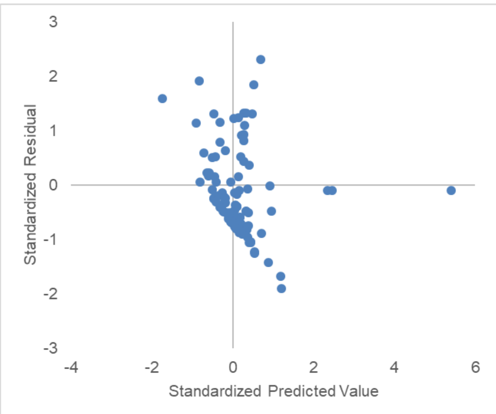

Fig. 5. MGWR Residuals

Table 2. Comparison of methods based on different evaluation criteria

| Evaluation Criteria/Method | NSE | RMSE | r |
|---|---|---|---|
| Global regression | -0.63 | 4.64 | 0.04 |
| GWR | 0.03 | 3.54 | 0.42 |
| MGWR | 0.35 | 2.90 | 0.63 |

In order to find the bandwidth and the coefficients of the variables, meteorological records of 184 stations collected on 30.03.2010 were used. A one-day record of 101 stations with precipitation parameter values from 01.04.2010 was used as the test data. In the GWR model, the coefficients vary for each point. The coefficients of the other points within the bandwidth were directly proportional to their distances. While the points close to each other affect each other more, distant points affect each other less. The optimal bandwidth for the GWR was found to be 168 km. In the MGWR method, the bandwidth is determined separately for each variable. MGWR algorithm finds the optimal bandwidth for bw0 as 183 km and the optimal bandwidth for the variables of altitude, air pressure, air temperature, and cloud coverage as 60, 183, 183, 183, and 183 km, respectively.

**Experiments**

Experiments were performed to determine the residuals of the models by measuring the difference between the precipitation amount measured at the stations and the precipitation amount obtained by the methods. The RMSE, NSE, and R values of the methods were calculated for global regression, GWR, and MGWR methods. The RMSE values calculated for the global regression, GWR, and MGWR methods were 4.64 mm, 3.53 mm, 2.9 mm, respectively. NSE values and R values were -0.63, 0.03, 0.35 and 0.04, 0.42, 0.63, respectively. The estimated precipitation and residual values of the methods are presented in Fig. 3, 4 and 5, respectively.

As shown in Fig. 3, 4, and 5, the results estimated by the MGWR method were closer to the measured values. GWR and MGWR methods outperform the global regression model because global regression assumes that the effects of altitude, air pressure, relative humidity, air temperature, and cloud coverage variables on precipitation do not change as the location changes. On the other hand, the MGWR method outperforms both the global regression and GWR methods. In GWR, a single bandwidth was used for all variables. In contrast, in the MGWR model, different bandwidths are used for each variable. In other words, the bandwidth was calculated for all variables. Although the bandwidth calculation cost of the MGWR method is higher than that of the other methods, the estimated precipitation value is better than that of the other methods.

**Conclusion**

This study aims to predict precipitation using other relevant meteorological parameters involved in precipitation formation. The methodology proposed herein is particularly important for locations that lack precipitation measurements at sufficient spatial resolution. Multiscale geographically weighted regression (MGWR) is applied to predict precipitation. In contrast to the classical GWR, the spatial scale of the MGWR is not fixed; therefore, no single bandwidth exists. MGWR uses different spatial scales and bandwidths for different processes. In this study, the prediction performance of the MGWR was compared with that of the global regression and classical GWR. The experimental evaluation, which was conducted on actual meteorological data, proved that the MGWR outperformed the other approaches. In future work, we plan to reduce the computational complexity of MGWR by using new techniques and parallel or distributed strategies and to use different techniques to determine optimal bandwidth values efficiently.

**Acknowledgements**

This study is supported by the Erciyes University Research Fund (FBA-2022-12224).

**References**

Ashiq, M. W., Zhao, C., Ni, J., Akhtar, M. (2010). GIS-based high-resolution spatial interpolation of precipitation in mountain–plain areas of Upper Pakistan for regional climate change impact studies. Theoretical and Applied Climatology, 99(3), 239-253.

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, 21(4), 455-466.

Celik, M., Dadaser-Celik, F., Dokuz, A. S. (2014). Discovery of hydrometeorological patterns. Turkish Journal of Electrical Engineering and Computer Sciences, 22(4), 3.

da Silva, A. R., de Oliveira Lima, A. (2017). Geographically Weighted Beta Regression. Spatial Statistics, 21, 279-303.

Diodato, N. (2005). The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. International Journal of Climatology, 25(3), 351-363.

Dong, G., Nakaya, T., Brunsdon, C. (2018). Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. Computers, Environment and Urban Systems, 70, 35-42.

Fotheringham, A., Crespo, R., Yao, J. (2015). Geographical and Temporal Weighted Regression (GTWR). Geographical Analysis, 47.

Fotheringham, A. S., Brunsdon, C., M. Charlton. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley.

Fotheringham, A. S., Yang, W., Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247-1265.

Harris, R., Singleton, A., Grose, D., Brunsdon, C., Longley, P. (2010). Grid-enabling Geographically Weighted Regression: A Case Study of Participation in Higher Education in England. Transactions in GIS, 14(1), 43-61.

Hsueh, Y.-H., Lee, J., Beltz, L. (2012). Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008. Applied Geography, 34, 587-594.

Hung Tien, T., Hiep Tuan, N., Viet-Trung, T. (2016, 6-8 Oct. 2016). Large-scale geographically weighted regression on Spark. Paper presented at the 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE).

Leong, Y.-Y., Yue, J. C. (2017). A modification to geographically weighted regression. International Journal of Health Geographics, 16(1), 11.

Li, Z., Fotheringham, A., Li, W., Oshan, T. (2018). Fast Geographically Weighted Regression (FastGWR): A

Scalable Algorithm to Investigate Spatial Process Heterogeneity in Millions of Observations. International Journal of Geographical Information Science.

Lu, B., Brunsdon, C., Charlton, M., Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. International Journal of Geographical Information Science, 31(5), 982-998.

Ma, X., Zhang, J., Ding, C., Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. Computers, Environment and Urban Systems, 70, 113-124.

Mays, L. (2001). Water Resources Engineering. New York: John Wiley & Sons, New York.

Tasyurek, M., Celik, M. (2020). RNN-GWR: A geographically weighted regression approach for frequently updated data. Neurocomputing, 399, 258-270.

Tasyurek, M., Celik, M. (2022). 4D-GWR: geographically, altitudinal, and temporally weighted regression. Neural Computing and Applications, 34(17), 14777-14791.

Taşyürek, M., Celik, M. (2021). FastGTWR: A fast geographically and temporally weighted regression approach. Journal of the Faculty of Engineering and Architecture of Gazi University, 36, 715-726.

Wang, K., Zhang, C., Li, W. (2013). Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. Applied Geography, 42, 73-85.

Wei, C.-H., Qi, F. (2012). On the estimation and testing of mixed geographically weighted regression models. Economic Modelling, 29(6), 2615-2620.

Zhang, H., Zhang, J., Lu, S., Cheng, S., Zhang, J. (2011). Modeling hotel room price with geographically weighted regression. International Journal of Hospitality Management, 30(4), 1036-1043.