



1st Conference on Spatial Statistics: Mapping Global Change

Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wrocław (SW Poland)

Mariusz Szymanowski^{a*}, Maciej Kryza^a

^a*Institute of Geography and Regional Development, University of Wrocław, pl. Uniwersytecki 1, 50-137 Wrocław, Poland*

Abstract

Geographically weighted regression (GWR) and Multiple Linear Regression (MLR) models have been applied to derive the spatial structure of urban heat island (UHI) in Wrocław, SW Poland and compared. It was found that GWR is better suited for spatial modeling of UHI than MLR, as it takes into account non-stationarity of the spatial process. Both local and global models were extended by the interpolation of regression residuals, and used for spatial interpolation of the UHI structure. The combined: GWR + interpolated regression residuals (GWRK) approach is recommended for spatial modeling of UHI.

© 2010 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Selection and/or peer-review under responsibility of Spatial Statistics 2011

Keywords: urban heat island; geographically weighted regression; multiple regression; residual kriging, spatial interpolation, GIS

1. Introduction

Urban heat island (UHI) is probably the most significant phenomenon of urban climate, with further strong impact on various aspects of urban environment. The most important features describing UHI are its magnitude and spatio-temporal structure. The general spatial structure of UHI is characterized by the occurrence of three distinct zones, named cliff, plateau and peak [1]. This general structure of UHI can be strongly modified depending on land-use types and urban structures. During calm and clear-sky

* Corresponding author. Tel.: +48-71-375-28-30; fax: +48-71-343-51-84.
E-mail address: mariusz.szymanowski@uni.wroc.pl.

meteorological conditions, which are favorable for UHI development, and especially during nighttime, UHI takes a multicellular, irregular shape [2].

Most of the recent studies on spatial characteristic of UHI are based on multidimensional interpolation algorithms, with the multiple linear regression (MLR) being the most often applied [3]. Nevertheless MLR could lead to distorted results when the spatial process is non-stationary. The main goal of this paper is the application and evaluation of GWR for determination of the spatial structure of UHI from Wrocław (SW Poland). A statistical analysis is performed to verify if there is methodological justification for a more complex, if compared to MLR, approach with local models. Finally, local and global regression models, both raw and extended by interpolation of the regression residuals are used for spatial interpolation of UHI. Interpolation results are evaluated with cross-validation (CV). The analysis is performed here basing on the case from 3 Jan 2002 when strong heat island (intensity > 8 K) was observed and distorted by regional wind resulting in the non-stationarity of the spatial process.

2. Study area

The city of Wrocław (293 km²; population ~640 000) is located in SW Poland (51°N, 17°E) at ~120 m a.s.l. The altitude in the city area varies from 105 to 148 m a.s.l., thus the temperature field is practically unaffected by elevation. Wrocław is situated along the Odra River. Approximately 31.4% of the city area is built-up, mainly with housing estates, industrial and warehouse buildings. The rest of the area consists of urban green space (36.6%), agricultural areas (28.9%) and water (3.1%).

3. Methods

3.1. Meteorological data

Air temperature measurements of Wrocław UHI were gathered during nighttime with relatively weak winds (<2 ms⁻¹) and cloudless sky. Air temperature data were measured by automatic mobile meteorological stations mounted on two cars. The sensors were placed in the aspirated radiation shields at 2 m a.g.l. Total of 206 measurement points (Fig. 1) were selected along the routes systematically to represent different land-use categories [3].

3.2. The set of potential auxiliary variables

The set of potential predictors was previously prepared basing mostly on land-use map of Wrocław and the buildings database available only for the selected areas of the city [3]. The auxiliary variables for interpolation were: roughness length, percentage of artificial and semi-natural surfaces, thermal admittance and artificial heat emission. For the purpose of this study the set of potential predictors was expanded by including LIDAR scan and satellite imagery (Landsat ETM+) derivatives: porosity, sky view factor, daily sums of solar irradiation, albedo, vegetation indices (NDVI, NDMI, SAVI), land surface temperature and emissivity [4].

3.3. Multiple linear regression model

The MLR method can be used to spatialize discrete point data on the assumption that auxiliary, independent variables are known and continuous in space. It has been successfully used for climatological purposes, as well as for UHI spatialization [3]. Selection of the predictors was performed stepwise, taking into account their statistical significance and the lack of collinearity with other independent variables, by

analyzing the variance inflation factor (VIF). Also, the direction of dependence (sign of the β coefficient) between air temperature and independent variable was checked to ensure that the final equation can be explained in terms of known physical processes that influence UHI formation.

3.4. Geographically weighted regression model

The MLR method can be applied to spatial data under assumption of spatial non-stationarity and location-independency. This is usually hard to meet, especially in the field of meteorology and climatology, where various processes should be considered as spatially unstable. One solution is to substitute the global regression model with the local model (GWR) [5]. It is a non-parametric model of spatial drift that relies on a sequence of locally linear regressions to produce estimates for every point in space by using a subset of data information from nearby observations. The local estimates are made using weighted regression, and weights assigned to observations are the function of the distance from point i . Larger weights are assigned to observations closer to the point i . Therefore, the weighting of an observation is not constant, but is a function of geographical location. For the continuous processes the kernel functions are suggested for constructing the weights.

For retaining the comparability of both global and local regression models, the same explanatory variables, as specified for MLR, were used in GWR model. Calibration of the local regression model in a deterministic manner included selection of the kernel type, size of bandwidth and verification that the local model can be physically explained over the whole study area. Due to irregular distribution of the sampling points (Fig. 1), the adaptive kernel (Gaussian shape) was used. The optimal bandwidth size was determined with iterative procedure, and all bandwidth sizes were tested for expected physical relation between UHI and the predictor. The following diagnostic measures were calculated: AICc – corrected Akaike Information Criterion [6], estimated standard deviation for the residuals (σ), global, and local minimum and maximum determination coefficients (R^2) as measures of goodness of fit.

After specification of the GWR model the set of parameter estimates was tested if it exhibits significant spatial variation over the study area. The reason is that if localized parameter estimates do not meet statistically significant differences, the GWR model can be considered as equivalent to the MLR model, although the local parameter estimates show spatial variation. In this paper, two tests were employed to address the issue: Monte Carlo [7] and F3 [8].

Both models, MLR and GWR, were evaluated and compared using a set of statistics to check a goodness-of-fit of the models to the observations. Additionally, spatial autocorrelation of regression residuals (Moran's I statistics) was analyzed to detect possible problems with proper specification of the model in the non-stationary conditions of the spatial processes. To test whether the GWR model offers an improvement over the MLR model, an analysis of variance (ANOVA) was used [5].

3.5. Residual kriging

In the regression methods described above, there is a part of variance that is not explained by the model and the regression residuals are assumed to be randomly distributed. The residual kriging (RK) algorithm sums up the trend component (deterministic, explained by the regression model) and residuals (stochastic part) interpolated with kriging (usually ordinary kriging, OK) technique.

3.6. Evaluation of interpolation results

Interpolation results derived with spatialization techniques applied in this study (MLR, GWR, RK and GWRK) were evaluated and compared using CV approach (“leave-one-out” method). The procedure is

repeated consecutively for all measuring sites and the interpolation errors are calculated as the difference between the observed and the modeled values. The CV errors were used to calculate diagnostic measures, including: mean bias (BIAS), the root mean square error (RMSE) and the mean absolute error (MAE).

4. Results

4.1. Specification of regression models

Statistical summary of the MLR model for 3 Jan 2001 is given in Table 1. The global model was able to explain 85% of the air temperature variance.

Table 1. Multiple linear regression analysis and spatial autocorrelation of regression residuals for 3 Jan 2002

Predictor	p-value	R ² adj.	AICc	F	Std Error	Moran's I E(I) = -0.0049	p-value
<i>Artificial heat</i>	0.0000						
<i>Albedo</i>	0.0395	0.85	479.1	373.8	0.77	0.3142	0.0004
<i>Land-surface temperature</i>	0.0000						

The process of GWR model specification was started by calibration of optimal kernel size. It can be stated that, with the decrease of the kernel size, the better fit of modeled data to observations is achieved, quantified by decreasing values of AICc and σ , and increasing global R². The analysis of changes in local parameter estimates (β) showed a similar tendency as in the case of R²: the smaller kernel the greater variance of β , but also the model is less biased. For the analyzed UHI case we found one independent variable (albedo) for which the use of small adaptive bandwidths (< 60 points) causes the change of sign of the local β estimate. This means that in some parts of the study area, the model does not properly describe the physical processes affecting the air temperature field. For this reason, the optimum adaptive kernel size cannot be defined simply by selecting the best R², AICc or σ statistics. In this case the kernel size should be as small as possible and the smallest possible for which the β values do not change sign over the study area and can be physically explained (Table 2).

Table 2. Geographically weighted regression analysis and spatial autocorrelation of regression residuals for 3 Jan 2002

Optimal adaptive kernel (no of objects)	Global R ² adj.	Min local R ² adj.	Max local R ² adj.	AICc	Moran's I E(I) = -0.0049	p-value
60	0.88	0.82	0.91	437.7	-0.0237	0.3540

The analysis of Moran's I index of GWR residuals showed significantly better specification of GWR model in comparison to MLR. A spatial autocorrelation of regression residuals was detected for the MLR model, suggesting the misspecification of the model due to the non-stationarity of the spatial process (Table 1). After the implementation of GWR, statistically significant tendency for clustering of similar residuals was no longer observed (Table 2).

The straightforward comparison of MLR and GWR models in each UHI case was performed by the analysis of AICc values, global R² and ANOVA approach. The analysis of AICc revealed a better fit of GWR model to observations than that for MLR. This was also supported by the analysis of global R² that

showed the increase of air temperature variation explained by the model up to 88% (locally even 91%) (Table 2). ANOVA results showed that the GWR estimates reduced the residual sum of squares over the MLR estimates. The test was statistically significant and showed that GWR model performed better than MLR (Table 3).

Table 3. Comparison between MLR and GWR model – ANOVA test (SS - sum of squares, DF – degrees of freedom) for 3 Jan 2002

Source	SS	DF	F
MLR Residuals	118.6	4.00	
GWR Improvement	68.5	25.2	
GWR Residuals	50.2	174.8	9.47

In the last step of GWR model calibration, non-stationarity tests for regression parameters were performed. The results of two tests are generally similar, but the existence of spatial instability was detected for artificial heat emission by Monte Carlo test and for albedo by F3 test. Even though it can be stated that GWR model for 3 Jan 2002 is specified properly.

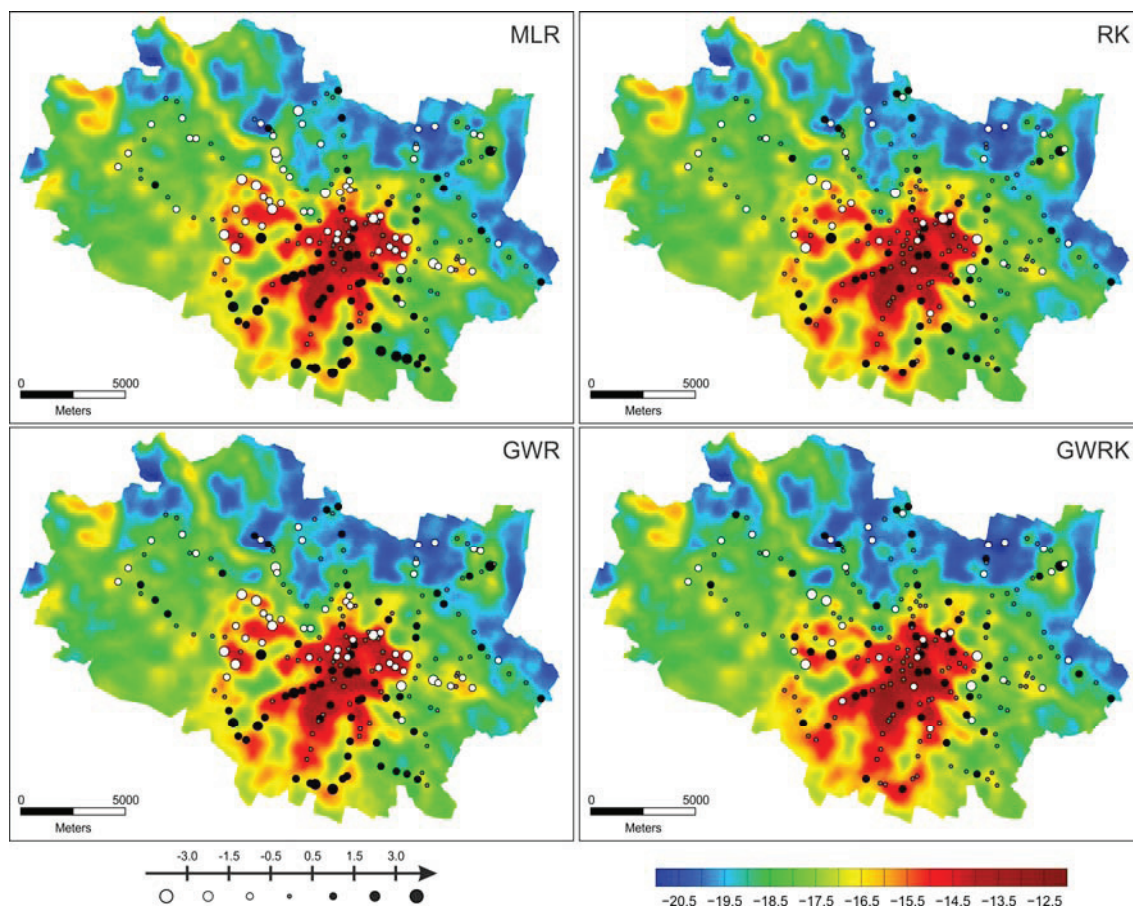


Fig. 1. Air temperature [°C] and standardized cross-validation (observed - modelled) errors for 3 Jan 2002

4.2. Spatial interpolation

The results of spatial interpolation by four methods are presented on Fig. 1. MLR is not able to detect exterior factor, i.e. the regional northern wind, which can be recognized by the analysis of spatial distribution of CV errors and results in the underestimations in southern parts and overestimations in northern part of the city. The GWR model works better (CV errors are smaller than for MLR) and improves the quality of the interpolation, but still does not fully recognize the role of wind. It should be stressed that the incorporation of stochastic part in the process of spatialization (RK, GWRK) improves the model and the zones of over- or underestimation are eliminated (Fig. 1).

5. Conclusions

Comparing the spatialization results achieved by the MLR and GWR techniques, even if the maps looks similar, the latter has a strong advantage in better recognition of non-stationarity characteristics of spatial process. The goodness-of-fit of the GWR model is the function of the kernel size, and increase with decreasing kernel size. There are two main reasons limiting decrease of the kernel size: too many independent variables for too few observations leads to the misspecification of the model and the physical interpretation of the local model is often lost for a given predictor if too small kernel size is selected. The GWR and MLR results can be significantly improved by adding the stochastic part to the interpolation (RK and GWRK) procedures. The results of those procedures are similar while comparing the CV errors statistical characteristics and spatial distribution. The main reason that is decisive in recognizing GWRK as the most proper method is its statistical correctness due to unexplained (by independent variables) and non-stationary phenomena.

Acknowledgement

The study was supported by the Polish Ministry of Science and Higher Education (science financial resources), project no N N306 155038.

References

- [1] Oke TR. The distinction between canopy and boundary layer urban heat island. *Atmosphere* 1976;**14**:268-277.
- [2] Kłysik K, Fortuniak K. Temporal and spatial characteristics of the urban heat island of Łódź, Poland. *Atmos Environ* 1999;**33**:3885-3895.
- [3] Szymanowski M, Kryza M. GIS-based techniques for urban heat island spatialization, *Clim Res* 2009;**38**:171-187.
- [4] Szymanowski M, Kryza M. Application of remotely sensed data for spatial approximation of urban heat island in the city of Wrocław, Poland. IEEE Xplore Digital Library; 2011 (in print).
- [5] Fotheringham AS, Brunsdon C, Charlton ME. Geographically weighted regression. Chichester: John Wiley & Sons; 2002.
- [6] Hurvich CM, Simonoff JS, Tsai C-L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc* 1998;**60**:271-293.
- [7] Charlton ME, Fotheringham AS, Brunsdon C. GWR 3. Software for Geographically Weighted Regression. Spatial Analysis Research Group, Department of Geography, University of Newcastle, Tyne, England; 2003.
- [8] Leung Y, Mei CL, Zhang WX. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ Plan* 2000;**A32**:9-32.