MAUSAM, **68**, 1 (January 2017), 9-22

551.571.3 (55)

Spatial variability of water vapour in south and southwest of Iran

YOUNES KHOSRAVI, HASAN LASHKARI* and HOSEIN ASAKEREH**

Faculty of Sciences, Department of Environmental Sciences, University of Zanjan, 45371-38791, Zanjan, Iran

**Faculty of Earth Sciences, Department of Geography, Shahid Beheshti University, Tehran, Iran*

***Department of Physical Geography, University of Zanjan, 45371-38791, Zanjan, Iran*

(*Received 27 January 2016, Accepted 3 May 2016*)

e mail : khosravi@znu.ac.ir

सार – इस शोध पत्र में दो समाश्रयण मॉडल, साधारण अतिलघु वर्ग और व्यापक रूप से प्रचलित तकनीक वाले भौगोलिक रूप से भारित समाश्रयण का उपयोग जलवाष्प और इससे संबंधित भौगोलिक विशेषताओं जैसे: अक्षांश, देशांतर, उन्नयन, ढलान तथा अवस्थिति का समाश्रयण संबंधों के मॉडलिंग में किया गया है। तदनुसार वर्ष 1981-2010 के समय अंतराल में दक्षिण और दक्षिण-पश्चिमी इरान के जल वाष्प आंकड़े पिक्सेल्स में संग्रहित किए गए हैं। सामान्य ओ एल एस समाश्रयण के अनुसार जलवाष्प एवं अक्षांश, ऊँचाई एवं अवस्थिति के बीच के संबंध विपरीत और देशांतर एवं ढलान के साथ सकारात्मक रहे हैं। भैगोलिक विशेषताओं तथा जल वाष्प के संबंध का जी डब्ल्यू आर मॉडल के द्वारा विश्लेषण करने से पता चलता है कि व्याख्यात्मक चरों (बैरिएबल्स) में सबसे बड़ा गुणांक क्रमश: देशांतर, अक्षांश, ढाल (स्लोप) और उन्नयन में रहा है। मॉडल के कार्य निष्पादन के बारे में कहा जा सकता है कि जलवाष्प के आकलन में GWR में OLS से अधिक सुधार दिखता है और इससे यथार्थ एवं उपयोगी परिणाम प्राप्त हुए हैं। इसकी वजह से GWR का R^2 समायोजित R^2 तथा AIC_c के मान क्रमश: 0.967, 0.968 तथा 9329.38 आए हैं जबकि OLS के लिए ये मान (फैक्टर्स) 0.8478, 0.8475 एवं 14559.04 है।

ABSTRACT. In this study two regression models, ordinary least square and geographically weighted regression as widely applied techniques, were used in modeling the regression relationships between water vapour and related geographic features, *i.e.*, longitude, latitude, elevation, slope and aspect. Accordingly, the water vapour data in south and southwest of Iran were collected in pixels in the time interval 1981-2010. According to the general OLS regression, the relationship between WV and latitude, elevation and aspect were reverse and with longitude and slope were positive. Analyzing the relationship between geographic features and WV by GWR model determined that greatest coefficients of explanatory variables were in longitude, latitude, slope, aspect and elevation, respectively. Regarding to the model performance, GWR showed an improvement over OLS in estimating the WV and provided more realistic and useful results. So that the R^2, Adjusted R^2 and AICc for GWR were 0.967, 0.968 and 9329.38, respectively while these factors for OLS were 0.8478, 0.8475 and 14559.04.

Key words – Water vapour, Geographically weighted regression (GWR), Ordinary least square (OLS), Spatial autocorrelation, Multicollinearity.

1. Introduction

Water vapour plays a crucial role in climate system as an important feedback variable associated with the earth's energy balance and hydrologic cycle (Naud *et al*., 2012). This climatic parameter has an important role in explaining the climate change or changes in climatic parameters, because of (*i*) It is the main source of rainfall in all weather systems, (*ii*) It supplies the latent heat in this process and controls the heat in the troposphere (Trenberth and Stepaniak, 2003; Serrano *et al*., 1999; Wentz *et al*., 2007), (*iii*) It is the booster of the storm's speed (Allen and Sodden, 2008) and (*iv*) plays a major role in the dynamics of atmospheric circulation (Ross and Elliott, 1996). So, determination and interpretation of the likely reasons of WV changes and its variability are vitally important for human as well as other living-beings (Tonkaz *et al*., 2007).

There are two important approaches to analyze the variability of a climatic parameter such as water vapour, spatial and temporal. Mainly in the temporal variation of climatic parameters, the trend analysis has been noted (Begert *et al*., 2005; Brunetti *et al*., 2000; Kampata *et al*., 2008; Yue and Hashino, 2003). Another approach is analyzing, interpreting and detecting of spatial variations by determining an optimal model (Delhomme, 1979; Burgess and Webster *et al*., 1980). In this context, the statistical models have received considerable attention in environmental sciences (Cressie, 1993; Anselin and Getis, 2010).

One of the most common statistical models in environmental sciences is multivariate regression that explains the relationship between the variables, count as a tool for recreating, estimating and forecasting (Dodd, 2006). There is a variety of regression modelling but one of them that is more common between environmentalists is Ordinary Least Square (OLS). This model is a form of linear regression that yield a single estimate of the relationships between the dependent variable and a set of explanation variables based on the entire study area. Because of this characteristics this model has known as a global regression model. The basic assumption in OLS is independence of observations that often violated due to temporal or/and spatial autocorrelations in data, which leads to a biased estimation of the standard errors of model parameters and, consequently, misleading significance tests (Anselin and Griffith, 1988; Fox *et al*., 2001). This problem is called multicollinearity. In fact this problem happens when one independent variable is nearly combination of other independent variables (Lin, 2008). In dealing with multicollinearity, there are two solutions : (*i*) dropping redundant variables from model directly (Bowerman and O'Connell, 1993) and (*ii*) using another regression model that able to solve the multicollinearity. It is difficult to decide which redundant variables cause the multicollinearity in the model, then it said that using of the first solution is hard.

On the other hand, most of climatic parameters such as water vapour have spatial non-stationary or spatial heterogeneity. Thus, based on this characteristic, by using of a global regression is not enough for describing the whole area and determining relationships between parameters. Therefore, when spatial autocorrelation and multicollinearity is present, an alternative model is recommended for OLS. One of the best alternative model for more accurate estimating is geographically weighted regression (GWR).

Geographically Weighted Regression (GWR) is a local regression to deal with spatial autocorrelation and heterogeneity for predicting environmental parameters. This model was specifically designed to deal with the spatial non-stationarity of regression coefficients between the target variable and explanatory variables by measuring those coefficients locally using local data (Brunsdon *et al*., 1998). This is not possible with traditional regression (OLS) because the parametric stability hypothesis is assumed, which is equivalent to considering that calculated coefficients do not have significant differences in space (Cardozo *et al*., 2012).

One of the most important of usages of OLS and GWR in climatology for the identification of the interaction and behavior of a climatic parameter with climatic factors and other climatic parameters that will lead to achievement of overall view of their spatial distribution. Recent studies about the relationship between climatic parameters and geographic features such as longitude, latitude, elevation, slope and aspect have focused mainly on the linear regression model in determination of variability of temperature (Hudson and Wackernagel, 1994; Bolstad *et al*., 1998) and precipitation (Basist *et al*., 1994; Konrad, 1996; Singh and Kumar, 1997; Johansson and Chen, 2003; Um *et al*., 2011).

But recently, the GWR method has been increasingly employed to model spatial distributions and relationships in environmental sciences such as geography (Kamarianakis *et al*., 2008; Tian *et al*., 2012; Zhao *et al*., 2010; Li *et al*., 2010), agriculture (Mishra *et al*., 2010; Zhang *et al*., 2011; Wang *et al*., 2012; Wang *et al*., 2013) and water management (Brunsdon *et al*., 1998; Huang *et al*., 2015; Pratt and Chang, 2012). However, few studies have addressed the relationships between climatic parameters and climatic factors by GWR (Brunsdon *et al*., 2001; Foody, 2003; Diodato, 2005; Zhao *et al*., 2010; Fotheringham *et al*., 2003).

Therefore, the objective of this study is to evaluate the spatial variations of WV in association with geographic features such as longitude, latitude, elevation, slope and aspect by using two forecasting models, traditional regression model (OLS) and local spatial regression (GWR) in South and Southwest of Iran. The results of this study can help researchers and decision makers in Iran to achieve more information about spatial variability of WV in South and Southwest of Iran.

2. Materials and method

2.1. *Study area*

The study area, with about 360,200 km^2 area, is located in the south and southwest of Iran and approximately between 25° 00' N and 34° 25' N latitudes and between 45° 38' E and 59° 17' E longitudes (Fig. 1). Southern and southwestern parts of the study area are located beside of two massive sources of moisture, Persian Gulf and Oman Sea. The main mountain chain in the study area is Zagros that extends from the northwest to the southern part of study area (Dinpashoh *et al*., 2011). The Zagros mountain range is responsible for the major portion of rain - producing air masses that enter the region

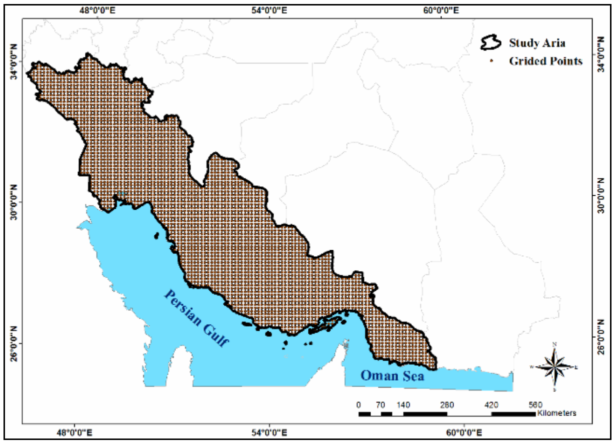

Fig. 1. The daily WV gridded data in pixels with dimension of 10 × 10 km

from the western and northwestern sides, with relatively high amounts of rainfall (Sadeghi *et al*. 2002).

2.2. *Data collection*

In this study, WV data in pixels (dimension of 10×10 km) in the time interval 1981-2010 (Fig. 1) were collected by the Iranian Meteorological data website (http://www.weather.ir). As regards the aim of this study, analyzing the spatial variability of WV in regional and local scale, the most important geographic features such as elevation, longitude, latitude, slope and aspect (as independent variables) were chosen. Topography maps of the study area collected by the Geological Survey of Iran (http://www.gsi.ir). By mosaicking, georeferencing and editing these maps in Arc GIS 10.2 software, the Digital Elevation Model (DEM) by 10 km cell size was derived and based on it, the geographic features are prepared.

2.3. *Data processing*

One of the global regression models that has been used widely in climatology and meteorology is OLS. The output of the analysis in this model is a single regression equation describing the relationship between the dependent and explanatory variables across the whole study area (Łukawska-Matuszewska and Urbański, 2014):

$$Y_i - \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} + \varepsilon_1, \quad i = 1, 2, \ldots, n$$

where, Y_i is a dependent variable; $X_i - X_{i.p-1}$ are explonatory variables; $\beta_0 - \beta_{p-1}$ are the regression coefficients; and ε_1 is the random error term (residuals).

The basic assumption in OLS is the independence of observations; but this condition in climatology usually hard to occur, where many processes can be considered as spatially unstable (Szymanowski and Kryza, 2012). In this regard, a local regression appropriates for non-stationary cases (Fotheringham *et al*., 2003; Foody, 2003; Su *et al*., 2012). GWR is a practical technique to examine the spatial variation and non-stationarity for continuous surface of parameter values at regional scale (Brunsdon *et al*., 1996; Fotheringham *et al*., 1996). While the OLS calculates the coefficients for the whole study area, the GWR based on the variability of the parameters in the space calculates regression coefficients at each individual location (Fotheringham *et al*., 2003). The GWR model for each regression point *i* described as:

$$Y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i) X_{ik} + \varepsilon_i$$

where, y_i is the i_{th} observation of the dependent variable, a_0 is the regression constant, a_k is the coefficient of the k_{th} independent variable, X_{ik} is the i_{th} observation of the k_{th} independent variable and ε_i is the residual of the i_{th} observation, (u_i, v_i) indicates the spatial location of each observation. Thus, GWR provides for all parameters to vary in terms of location. Some advantages of GWR model in relation to other regression model such as OLS include: (*i*) GWR allows moving from a global perspective to a local analysis of the, (*ii*) GWR allows the relationships to vary over space (Lee and Schuett, 2014), (*iii*) Estimated errors in GWR are lesser than the OLS and thereby the local coefficient of determination (R^2) increase (Hadayeghi *et al*., 2010).

In this study, the geographic features and WV are as independent parameters and dependent parameter, respectively. In the first step, based on the OLS model and the interpolated WV map, relationships between the WV and geographic features were evaluated. The spatial variability of WV based on the geographic features were examined by GWR model. In this step, the distributions of the model's spatial fits (local R^2 values) and the local coefficients were mapped.

Finally based on the results of two models, GWR and OLS, the optimum model was determined. In order to evaluate the fit of the OLS and GWR, the following tests were performed : R-Squared (R^2), adjusted R-Squared (R^2), corrected Akaike Information Criteria (AICc), Jarque-Bera p-value (JB), Koenker's studentized Breusch-Pagan p-value (BP), Variance Inflation Factor (VIF), and Global Moran's I *p*-value (MI) using ArcGIS (ESRI, 2012). The normality, homoscedasticity and the absence of autocorrelation in the residuals obtained from the linear regression models can be tested with the Jarque-Bera

TABLE 1

Coefficients matrix of correlation between the water vapour and geographic features

		WV	Longitude	Latitude	Elevation	Slope	Aspect
WV	Coefficient correlation	1	.626**	-.781**	-.671**	-.187**	.026
	Significance		.000	.000	.000	.000	.138
Longitude	Coefficient correlation		1	-.89**	-.022	-.054**	-.018
	Significance			.000	.203	.002	.313
Latitude	Coefficient correlation			1	.307**	.122**	.009
	Significance				.000	.000	.623
Elevation	Coefficient correlation				1	.407**	-.06**
	Significance					.000	.001
Slope	Coefficient correlation					1	-.029
	Significance						.098
Aspect	Coefficient correlation						1

Correlation is significant at the 0.01 level**

statistic (J-B) (Jarque and Bera, 1980). When this test is statistically significant ($p < 0.05$), model predictions are biased (the residuals are not normally distributed). In a linear regression model, the Koenker Bruesch-Pagan Statistic (BP) is used to test for heteroskedasticity,that estimates whether the variance of regression residuals is dependent on the values of independent variables (Breusch and Pagan, 1979).

One of the methods for determining the presence of multicollinearity is the Variance Inflation Factor (VIF). The VIF tells us how much the variance of a coefficient associated with the explanatory variable increases because of the linear dependence between the explanatory variables (Lukawska-Matuszewska, 2014). Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables (ESRI, 2012). The Global Moran's Index is a stronger indicator for spatial non-stationarity among predictors and the response variables (Ivajnsic *et al*., 2014). To assess the model performance, widely used diagnostic tools are the Akaike's information criterion (AIC) (Akaike, 1998), R^2 and adjust R^2. AIC is used to compare the performance of models with different sets of independent variables or to compare the global (OLS) and local (GWR) models (Burnham and Anderson, 2002). The adjust R^2 reflects model complexity and is considered a more accurate measure of the model performance. So the best model is the one with the highest adjusted R-square (Adj. R^2) and the lowest AICc (Staub *et al*., 2014). The magnitude of residuals, *i.e*., the differences between the observed and predicted values of dependent variable, is another measure of a model fit; the smaller the residuals, the better fit of the model (Fahrmeir *et al*., 2013; Kuhn and Johnson, 2013). The global Moran's I index and its correlogram were used to determine whether residuals of GWR and OLS models are spatially auto correlated. The spatial autocorrelation in the independent variables in an OLS model is showing the multicollinearity. Several important transformations is considered for solving the multicollinearity such as *Zi*, *Wi*, *Mi*, *Ni*, *Oi*, *Vi*, *Yi* and *Ti*. In the present study, all multicollinearity transformations were tested and one of them that can solve this problem will be chosen to enter the GWR model.

3. Results and discussion

3.1. *Ordinary least square*

One of the prerequisites of spatial variation is estimating and evaluating the correlation between the dependent variable and explanatory variables. Therefore in the first step, bivariate correlation was calculated between the dependent and explanatory variables for 3338 pixels (Table 1 and Fig. 2).

Table 1 shows an overall overview of how various geographic features influence the behavior of the WV. According to this table the highest and lowest positive correlation coefficients are observed between WV and

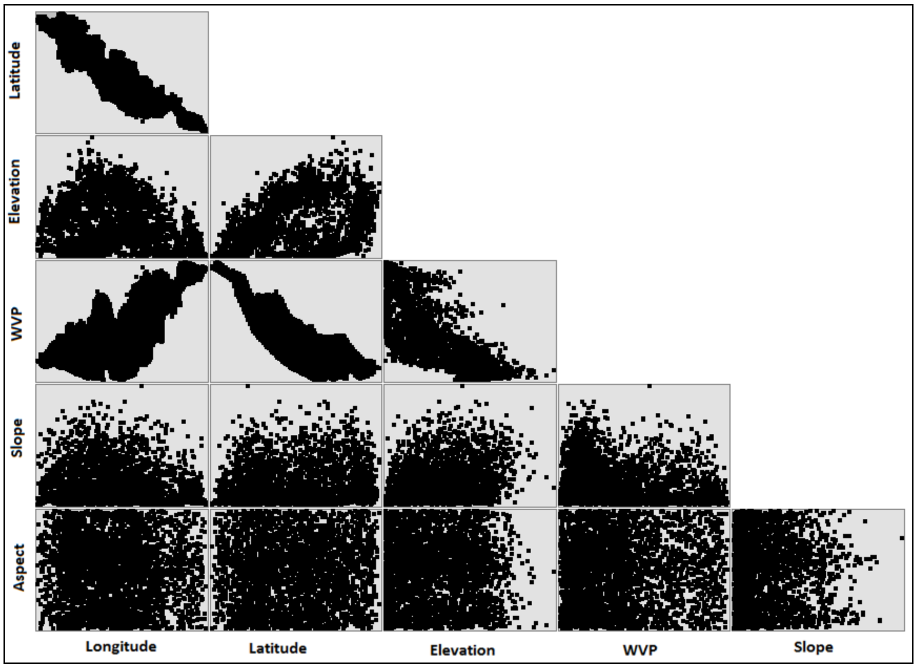

Fig.2 . Scatter plot matrix of bivariate relationships between WV and explanatory variables

TABLE 2

Diagnosis of the OLS analysis

Variable	Coefficient	Std Error	*t*-statistic	Probability	Robust_SE	Robust_t	VIF
Intercept	127.87	14.71	8.689	0.000*	14.681	8.709	-
Longitude	0.000008	0	22.961	0.000*	0	23.0641	7.5015
Latitude	-0.000005	0	-11.301	0.000*	0	-11.323	8.2593
Elevation	-0.004242	0.000063	-67.599	0.000*	0.00006	-63.949	2.0084
Slope	0.0743	0.0044	16.699	0.000*	0.00486	15.279	1.2283
Aspect	-0.00036	0.00036	-0.299	0.764	0.00036	-0.297	1.0034

OLS diagnostic information

Number of observations	3338	Joint F-Statistic Value:	3712.346
Number of variables	5	Joint F-Statistic Probability (p-value):	0.000000*
R^2	0.8478	Wald statistic	14247.358
Adjusted R^2	0.8475	Wald Statistic Probability (p -value):	0.000000*
AIC	14559.044		

*Statistically significant at the 0.05 level

longitude (0.626) and aspect (0.026), respectively. On the other hand, the WV is in an inverse relationship with elevation, latitude and slope.

Thus, it can be said all of the independent variables have an ability to fit into a regression model for estimating the WV in the study area.

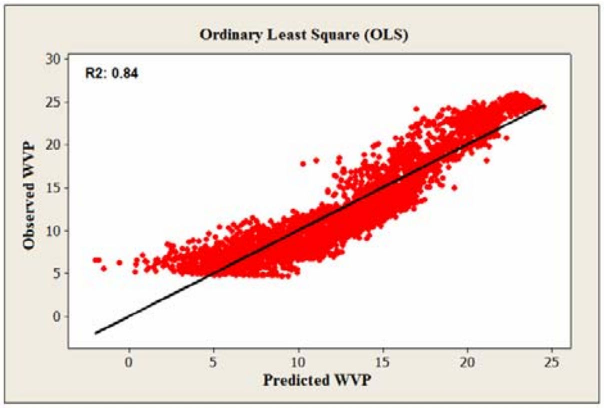

Fig. 3 Scatter plot of observed and OLS-estimated WV

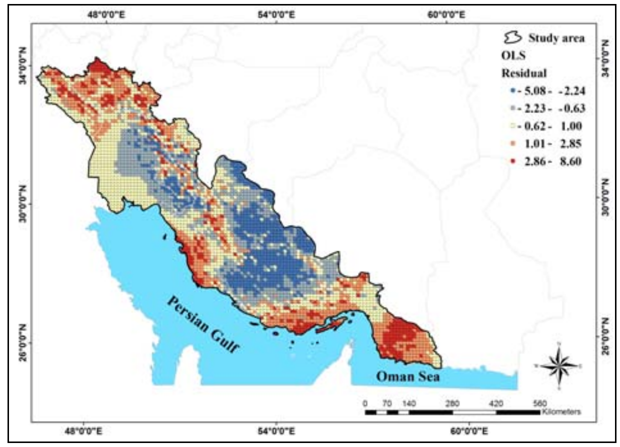

Fig. 4. Spatial distribution of the WV residuals of OLS model in south and southwest of Iran

The following Ordinary Least Squares regression model describes the relationship between the WV and the 5 explanatory variables, *i.e.*, longitude (*X*), latitude (*Y*), elevation, slope and aspect:

$$WV = 128 + 0.000008\ X - 0.000005\ Y - 0.00424\ \text{Elevation} + 0.0743\ \text{Slope} - 0.000109\ \text{Aspect}$$

The global OLS model suggests a linear relationship and explains 84.8% of the variability of WV using the five geographic parameters. This model indicates that each of the explanatory variables has a different influence on the dependent variable.

According to the regression equation, slope with the highest coefficient is as the most influential geographic feature for estimating the WV in the study area. Based on this equation, it can be said if a degree of slope rises, WV increases by 0.0743 hPa. Elevation has a negative influence on the WV and as prediction by the model, for increasing the elevation in each meter, 0.0042 hPa of WV decreases. The relationship between longitude and WV is slightly positive which means that with an increasing change of a longitude, approximately 0.000008 hPa of WV rises. In other words, over 100 km to the East, about 0.8 hPa will be added to the amount of WV. Thus, in 1250 km (the longest distance between the eastern and western points of the study area) it is expected that WV decreases about 10 hPa. The relationship between WV and aspect is reverse. It was estimated that by moving from south to north, about 0.000005 hPa of WV decrease for every meter. Therefore, it is expected that with 900 km (longest distance between the southern and northern points of the study area), about 0.0042 hPa of WV decrease. According to the relationship between WV and aspect, it is clear that by clockwise moving from north 0° to 359°, 0.000109 hPa of WV will be reduced. However, the predicted values of WV for per amount of geographic features is low, but it can be said that they are significant at any desired level.

Summary of OLS results is shown in Table 2. According to Table 2, slope with the highest coefficient (0.074) is as the most influential geographic feature in estimating the WV. Probability column showed that all independent variables coefficients are statistically significant (0.000000*) with the exception of aspect. The global model fit gives R^2 and adjusted R^2 values of 0.8478 and 0.8475, respectively. The F-statistic (3712.34) and Wald statistic (14247.358) values and their associated *p* - value (0.000000*) indicated that overall OLS model is statistically significant.

In order to diagnostic the error in the estimation values, scatter plot of observed and estimated values of WV was drawn (Fig. 3). This figure indicated that incoordination between observed and estimated values of WV was in low and high values of WV. Likewise, coordination or the higher accuracy of the OLS is observed in the middle values of WV, 7 to 17 hPa. Thus, it can be said that the error in the OLS model in low and high values of WV is higher than the other values.

The spatial distribution of the residuals of OLS model (obtained by subtracting the observed and estimated values of WV) and also spatial distribution of estimated values of WV are presented in Figs. 4 and 5, respectively. According to Fig. 4, that shows the deviations of observed and estimated values of WV, the highest values of residuals clustered in Persian Gulf and Oman Sea shoreline and highest parts of northwest of the study area. It means that in these regions, the WV tends to underestimate the output values.

TABLE 3

Diagnosis of the accuracy OLS analysis

Variables	Value	Prob	
Koenker (BP) statistic	287.918	0.000000*	Prob (>chi-squared), (5) degrees of freedom
Jarque e Bera statistic (J-B)	47.557	0.000000*	Prob (>chi-squared), (2) degrees of freedom
Global Moran's Index	0.491	0.000000*	-

* P < 0.5

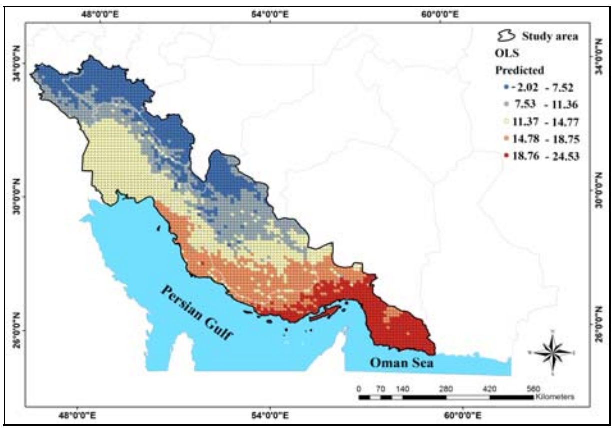

Fig. 5. Spatial distribution of the OLS-estimated WV in south and southwest of Iran

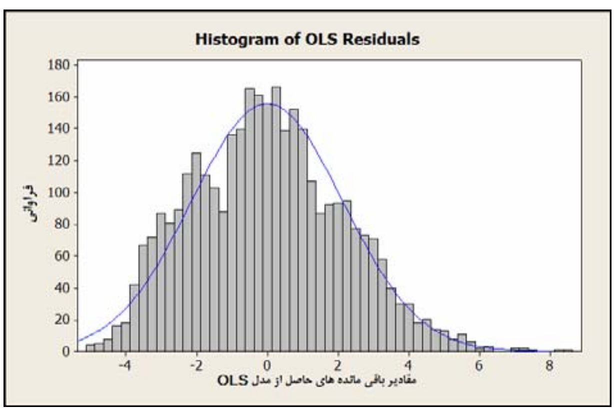

Fig. 6. Histogram of OLS residuals

On the other hand, the lowest residuals took place in the central part of Zagros Mountains. It means that the OLS model significantly overestimated the WV in this part of the study area.

According to Fig. 5, WV values reduced by moving away from massive sources of moisture, Persian Gulf and

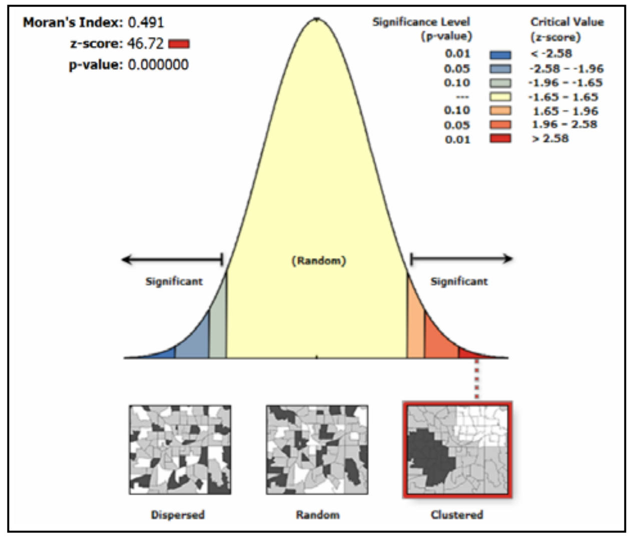

Fig. 7. Global Moran's Index of OLS residuals

Oman Sea and moving towards higher latitudes and heights.

In order to diagnosing the accuracy of OLS model and knowing normality, homoscedasticity and the absence of autocorrelation in the residuals, the Jarque-Bera (JB), Breusch-Pagan (BP) and Global Moran's Index tests were performed. In this context the Moran's Index is a stronger indicator for detecting the spatial autocorrelation. It means that, if spatial relationships do exist, results of the OLS model wouldn't be completely reliable. The results of these tests are shown in Table 3.

According to the results displayed in Table 3, the statistically significant J-B and BP ($p < 0.05$) indicate that the residuals are not normally distributed and the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). The result is better appreciated in Fig. 6. The existence of spatial autocorrelation in OLS model residuals was confirmed with the Global Moran's Index test (MI = 0.491; z score = 46.72; p = 0) (Fig. 7).

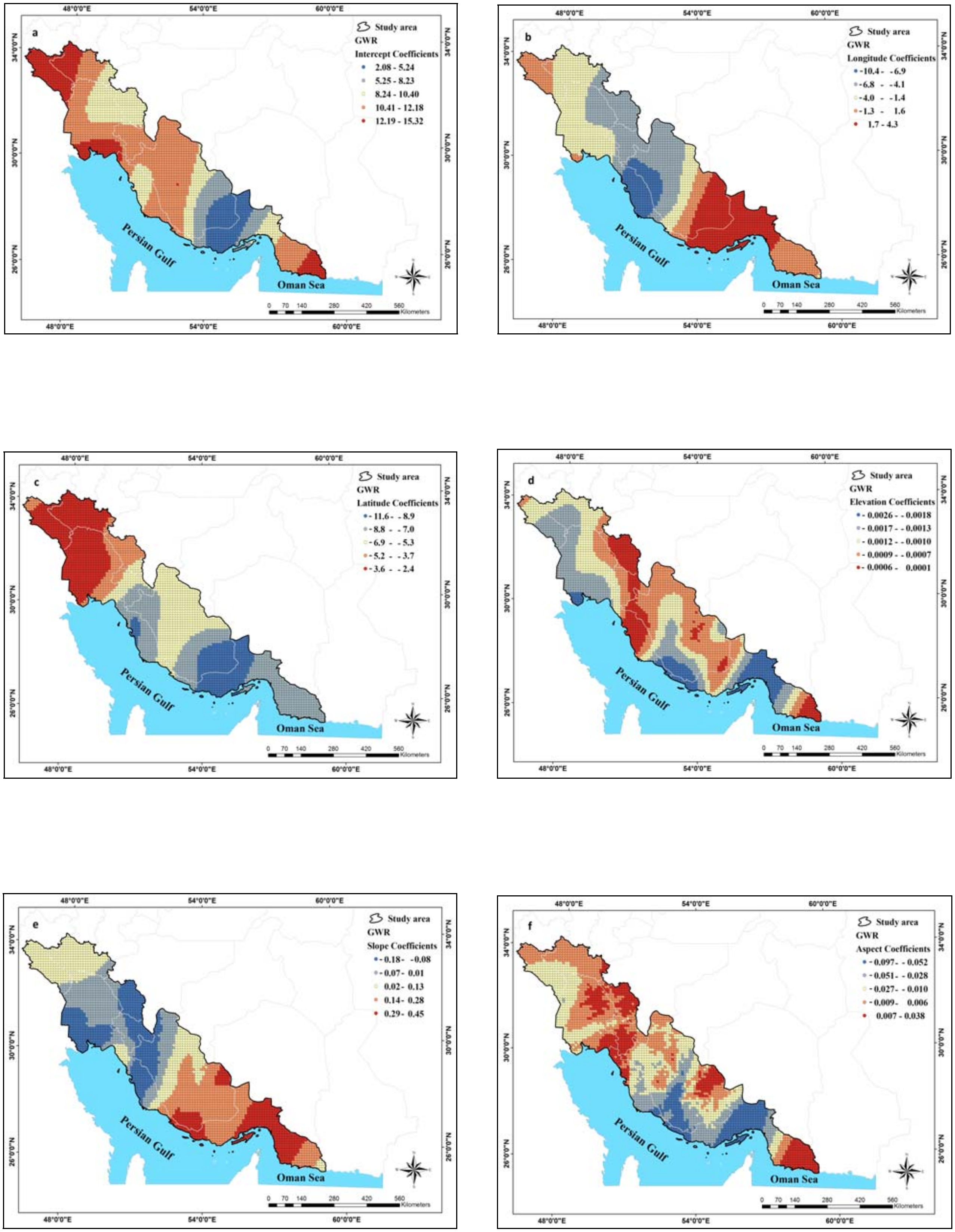

Fig. 8 (a-f). Spatial distribution of local coefficients of geographic features

TABLE 4

Estimated parameters and diagnostic statistics in OLS and GWR models

OLS model		GWR model			
Variable		Min	Max	Mean	Std Desv
Intercept	127.87	2.29	14.44	9.91	2.66
Longitude	0.000008	-9.98	4.17	-2.23	3.67
Latitude	-0.000005	-11.4	-2.54	-6.19	2.39
Elevation	-0.004242	-0.002	-0.0002	-0.001	0.0004
Slope	0.0743	-0.019	0.05	0.009	0.018
Aspect	-0.00036	-0.0009	0.0004	-0.0001	0.0002
Number of observations	3338			3338	
Number of variables	5			5	
R^2	0.8478			0.967	
Adjusted R^2	0.8475			0.968	
AIC	14559.044			9329.38	

*Statistically significant at the 0.05 level

According to these values, an alternative model is recommended for OLS. One of the best alternative models for more accurate estimating is geographically weighted regression (GWR).

3.2. *Geographically weighted regression model (GWR)*

The Strong alignment between independent variables, causes the occurrence of large variances for regression coefficients and then the estimates are unrealistic (Hooman, 2001). Then it is emphasized using another model that can provide a local model for analysis the variables by fitting a regression equation. One of the best models for that is GWR.

To analysis the presence or absence multicollinearity between independent variables, the VIF test was used. This test showed that all explanatory variables in OLS model were less than the critical value of 7.5 except of longitude and latitude (7.501 and 8.259, respectively) (Table 2). Thus it can be said 2 variables, longitude and latitude, cause the multicollinearity. Given the importance of these two variables in the WV estimation and impossibility of their regression, it should be transform their multicollinearity and then enter into another model. In the present study, all multicollinearity transformations (stated in introduction) were tested on longitude and latitude values and only one of them, *Z*-score, could solve the multicollinearity problem.

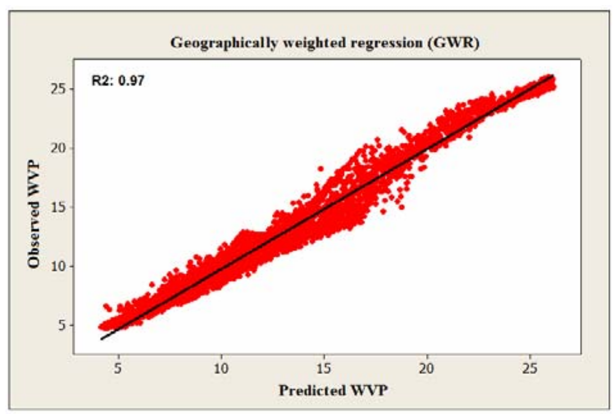

Fig. 9. Scatter plot of observed and GWR-estimated WV

According to these analyses, the important role of local non-stationary explanatory variables was confirmed. As we know, with the GWR, it was possible to survey the spatial variability of the local coefficients of explanatory variables. Fig. 8(a) shows the spatial distribution of intercept coefficients. According to this figure, the highest intercept coefficients were seen in the west and northwest of the study area and also in eastern parts of Oman Sea. On the other hand, the lowest values of intercept coefficients are located in southern parts of middle Zagros, south of the study area, and along the 59° E longitude. According to the Fig. 8(b), the calculated

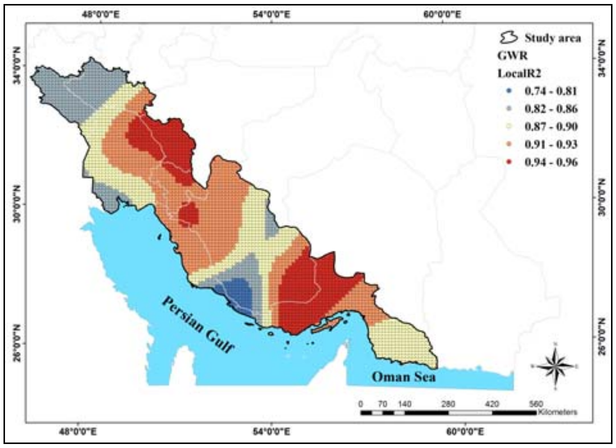

Fig. 10. Spatial distribution of local R^2

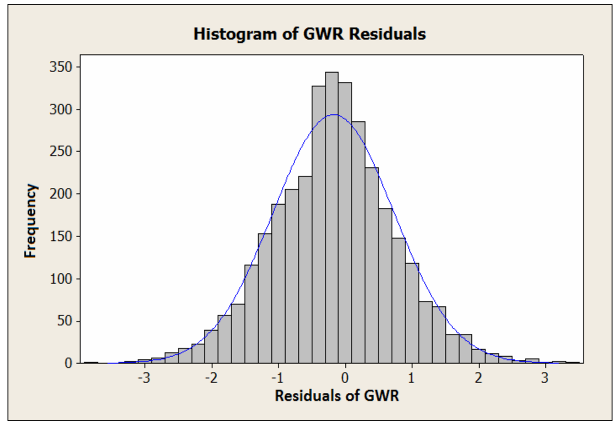

Fig. 12. Histogram of GWR residuals

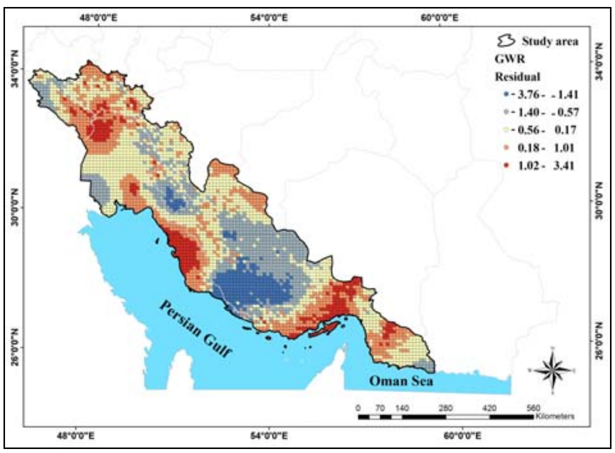

Fig. 11. Spatial distribution of residuals of GWR model

maximum and minimum coefficients by GWR of longitude were in the eastern and western parts of the study area, respectively. So, it can be said there is a positive relationship between longitudes and WV. Fig. 8(c) shows the latitude coefficients that are opposite to longitude. It means that latitude has an inverse correlation to WV. The lowest and highest values of coefficients of latitude were in southern parts of the study area and high latitude, respectively. The spatial distribution of elevation coefficients are shown in Fig. 8(d). According to this figure, negative coefficients of elevation were in flat surfaces such as Khuzestan plain, some parts of Persian Gulf and lowlands between Oman Sea and Persian Gulf. On the other hand, the highest coefficients of elevation were in heights such as Zagros Mountains. Positive coefficients of elevation indicated a direct relationship between elevation and WV. Therefore it can be realized the important role of elevation and wind directions that transport water vapour decreasing the WV. The role of elevation in reducing the WV decrease by moving towards the heights. The highest coefficients of slope were in eastern and central parts of the study area [Fig. 8(e)]. This situation is just the opposite of elevation. It means that where the coefficient of elevation is high, slope has low coefficient and therefore has less impact on WV. Fig. 8(f) shows the coefficients of aspect. The low coefficients of aspect can be found in the southern parts of Persian Gulf. In other regions, these coefficients indicated a greater impact of aspect on WV. Analyzing these maps determined that largest coefficients of explanatory variables can be seen in longitude, latitude, slope, aspect and elevation, respectively.

As it is said, difference between the traditional regression (OLS) and GWR is the global perspective of the OLS. It means that the relationships are considered for whole study area. However, the GWR calculates regression coefficients at each individual location.

R^2 and adjusted R^2 obtained using GWR (0.967 and 0.968, respectively) implied a considerable improvement with respect to the OLS model (0.8478 and 0.8475, respectively).

Fig. 9 shows the scatter plot of observed and GWR-estimated WV. The AIC value in GWR (9329.38) was substantially lower than in OLS model. Therefore, based on the AIC and R^2 Values, the GWR are chosen as the preferred model (Table 4).

Fig. 10 shows the spatial distribution of local R^2. According to this figure, the highest frequency of R^2 (94 to 96%) was more in heights. On the other hand, low values of this factor (74 to 81%) can be seen in some coastal areas and central part of Zagros Mountains.

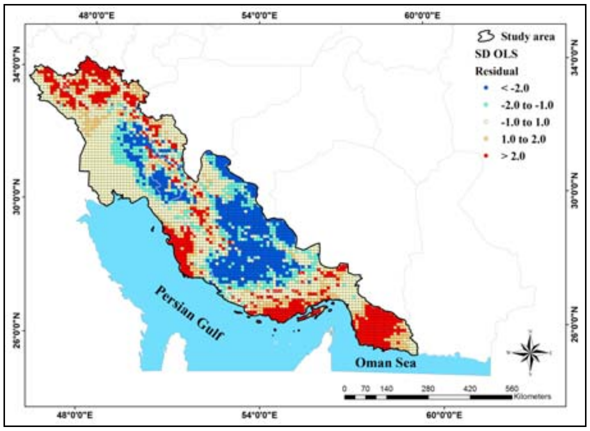

Fig. 13. Spatial distribution of standardized residuals of OLS

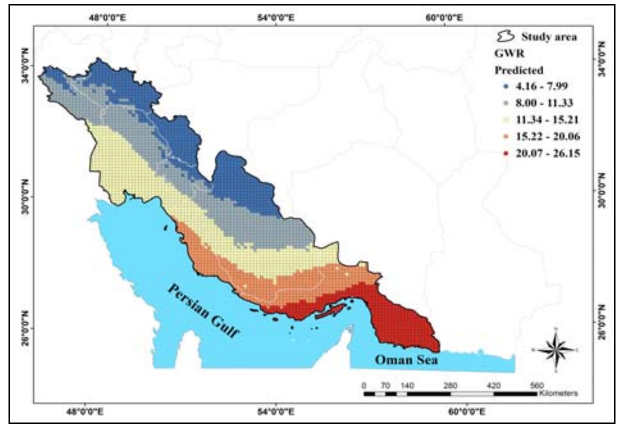

Fig. 15. Spatial distribution of GWR-estimated WV in south and southwest of Iran

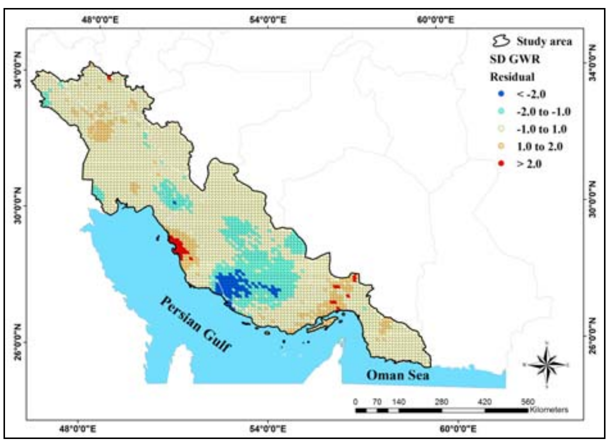

Fig. 14. Spatial distribution of standardized residuals of GWR

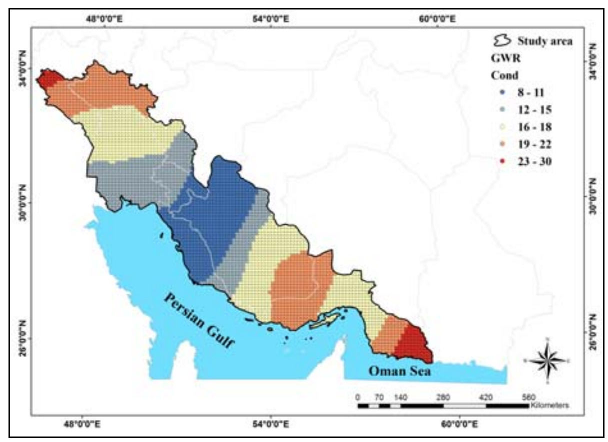

Fig. 16. Spatial distribution of condition numbers

The residuals of the GWR model are shown in Fig. 11. According to this figure, the residuals of the GWR model were smaller than those from the OLS model. Thus it can be said that the GWR model had better results.

In the GWR model, the values of residuals varied from -3.76 to 3.41 while it varied from -2.69 to 8.6 in OLS model. According to residuals of GWR and OLS it clear that the GWR model had a better fit than the traditional one (OLS). The residuals in the GWR model were almost normally distributed (Fig. 12) than the OLS (Fig. 5). The normality of GWR and OLS residuals are also shown in Figs. 13 and 14, the spatial distribution of standardized residuals of GWR and OLS, respectively. The standardized residuals values clear the under- (standard deviations of residuals > 2) and over-predictions (standard deviations of residuals < -2) of the dependent variable. According to these figures, the standard deviations of residuals characterize the lower under-predictions and over-predictions from GWR model, represent more normal distribution of GWR residuals than the OLS model.

Fig. 15 shows the predicted WV values by GWR model. According to this figure, the amount of WV reduced by moving away from the moisture sources of Persian Gulf and Oman Sea, moving towards the heights and the highest latitudes.

The condition number is the square root of the largest eigenvalue divided by the smallest eigenvalue (Lin & Wen, 2011) and is a reliable method for evaluating the local collinearity (Siordia *et al*., 2012). Multicollinearity

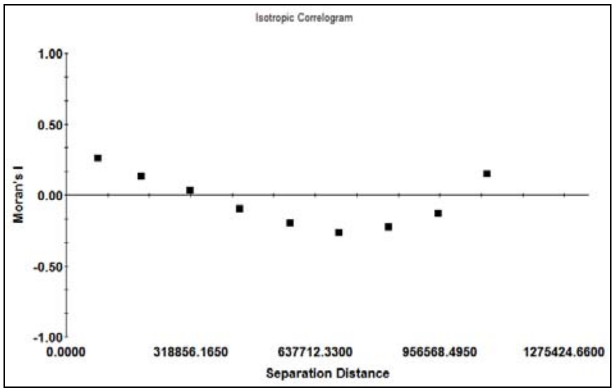

Fig. 17. Moran's I correlogram of OLS residuals

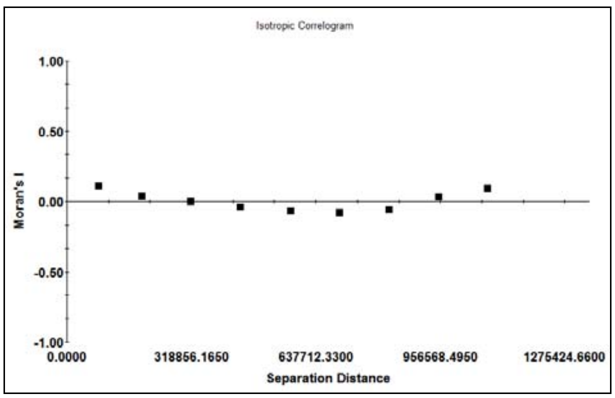

Fig. 18. Moran's I correlogram of GWR residuals

would be a great concern when condition numbers are greater than 30. The GWR model showed that the condition number is less than 30, which means that there are no serious local multicollinearity problems (Fig. 16).

The existence of autocorrelation in residuals of a model is showing the inadequacy of the model to estimates the values of the dependent variable. In this study, the global Moran's I index and its correlogram (autocorrelation plot) were used to determine whether residuals of GWR and OLS models are spatially autocorrelated. The global Moran's I demonstrated that residuals of OLS were spatially clustered. This index showed a positive spatial autocorrelation (MI = 0.491, Z = 46.72, P = 0). While the GWR residuals show less clustering than OLS (MI = 0.248, Z = 22.91, P = 0).

The correlogram of both OLS and GWR residuals are demonstrated in Figs. 17 and 18, respectively. The correlogram of OLS residuals demonstrated that the spatial autocorrelation of residuals observed up to 318 km that which indicated the stronger relationship between residuals and there dependency (Fig. 17). On the other hand, the spatial autocorrelation of the GWR residuals was seen partially only in initial intervals. These values reached to zero critical point too early and indicating no spatial autocorrelation between residuals. In this figure, the absence of spatial autocorrelation can be seen in more than 95% of study area. Therefore, it certainly can be said that estimation of the GWR model is more accurate than the OLS model.

4. Conclusion

The objective of this study is to provide a better understanding in spatial variability of water vapour in south and southwest of Iran, as well as the relationship with main geographic features. Accordingly, two important regression models, traditional regression model (OLS) and local spatial regression (GWR) were used to analyse the relationship between water vapour (dependent variable) and geographic features (response variables) in 3338 pixels. The global OLS model explained 84.8% of the variability of WV which used five geographic parameters. This model indicated that each explanatory variable has a different influence on the dependent variable. So that, slope had the highest coefficient as the most influential geographic feature for estimating the WV. In order to diagnosing the accuracy of OLS model, the Jarque-Bera (JB), Breusch-Pagan (BP) and Global Moran's I were performed. The results showed that the residuals are not normally distributed, the modeled relationships are not consistent and existence of spatial autocorrelation was confirmed. Analysis independent variables pointed out the presence of multicollinearity in two variables, longitude and latitude. After transforming, these variables fit into GWR model. Surveying the spatial variability of the local coefficients of explanatory variables in GWR model showed that the greatest coefficients of explanatory variables can be seen in longitude, latitude, slope, aspect and elevation, respectively.

The R^2 and adjusted R^2 obtained using GWR (0.967 and 0.968, respectively) implied a considerable improvement with respect to the OLS model (0.8478 and 0.8475, respectively). The AIC value in GWR (9329.38) was substantially lower than OLS model. Therefore AIC and R^2 Values of GWR indicated that the GWR was the preferred model. Analyzing the condition number for evaluating the local collinearity in GWR model showed not serious local multicollinearity problems.

References

Akaike, H., 1998, "Information theory and an extension of the maximum likelihood principle", In Petrov, B. and Csaki, F. (eds) 2nd Symposium on information theory, *Budapest*, Akademiai Kiado, 267-281.

Allen, R. P. and Sodden, B. J., 2008, "Atmospheric warming and the amplification of precipitation extremes", *Science*, **321**, 1481-1484.

Anselin, L. and Getis, A., 2010, "Spatial statistical analysis and geographic information systems", *In Perspectives on Spatial Data Analysis*, Springer Berlin Heidelberg, 35-47.

Anselin, L. and Griffith, D. A., 1988, "Do spatial effecfs really matter in regression analysis?", *Papers in Regional Science*, **65**, 1, 11-34.

Basist, A., Bell, G. D. and Meentemeyer, V., 1994, "Statistical relationships between topography and precipitation patterns", *Journal of climate*, **7**, 9, 1305-1315.

Begert, M., Schlegel, T. and Kirchhofer, W., 2005, "Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000", *International Journal of Climatology*, **25**, 1, 65-80.

Bolstad, P. V., Swift, L., Collins, F. and Régnière, J., 1998, "Measured and predicted air temperatures at basin to regional scales in the southern Appalachian Mountains", *Agricultural and Forest Meteorology*, **91**, 3, 161-176.

Bowerman, B. L. and O'Connell, R. T., 1993, "Forecasting and time series : An applied approach", 3rd, *Belmont: Duxbury Press*, 108-109.

Breusch, T. S. and Pagan, A. R., 1979, "A simple test for heteroscedasticity and random coefficient variation", *Econometrica : Journal of the Econometric Society*, 1287-1294.

Brunetti, M., Buffoni, L., Maugeri, M. and Nanni, T., 2000, "Trends of minimum and maximum daily temperatures in Italy from 1865 to 1996", *Theoretical and Applied Climatology*, **66**, 1-2, 49-60.

Brunsdon, C., Fotheringham, S. and Charlton, M., 1998, "Geographically weighted regression-modeling spatial non-stationarity", *Statistician*, **47**, 431-443.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E., 1996, "Geographically weighted regression : A method for exploring spatial nonstationarity", *Geographical analysis*, **28**, 4, 281-298.

Brunsdon, C., McClatchey, J. and 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.

Burgess, T. M. and Webster, R., 1980, "Optimal interpolation and isar1thmic mapping of soil properties", *Journal of soil science*, **31**, 2, 315-331.

Burnham, K. and Anderson, P., 2002, "Model Selection and Multimodel Inference : A Practical Information-theoretic Approach, second ed. *Springer-Verlag*", New York, p488.

Cardozo, O. D., García-Palomares, J. C. and Gutiérrez, J., 2012, "Application of geographically weighted regression to the direct forecasting of transit ridership at station-level", *Applied Geography*, **34**, 548-558.

Cressie, N., 1993, "Statistics for Spatial Data", Wiley Series in Probability and Statistics publication, 121-122.

Dinpashoh, Y., Jhajharia, D., Fakheri-Fard, A., Singh, V. P. and Kahya, E., 2011, "Trends in reference crop evapotranspiration over Iran", *Journal of Hydrology*, **399**, 3, 422-433.

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.

Delhomme, J. P., 1979, "Spatial variability and uncertainty in groundwater flow parameters : A geostatistical approach", *Water Resources Research*, **15**, 2, 269-280.

Dodd, S., Bassi, A., Bodger, K. and Williamson, P., 2006, "A comparison of multivariable regression models to analyse cost data", *Journal of Evaluation in Clinical Practice*, **12**, 1, 76-86.

ESRI, 2012, ArcGIS desktop. Retrieved from http://resources.arcgis.com/en/help/ main/10.1/ index.html#// 005p00000031000000.

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B., 2013, "Regression : Models, methods and applications", *Springer Science & Business Media,* 189-190.

Foody, G. M., 2003, "Geographical weighting as a further refinement to regression modeling : An example focused on the NDVI-rainfall relationship", *Remote sensing of Environment*, **88**, 3, 283-293.

Fotheringham, A. S., Charlton, M. and Brunsdon, C., 1996, "The geography of parameter space : An investigation of spatial non-stationarity", *International Journal of Geographical Information Systems*, **10**, 5, 605-627.

Fotheringham, A. S., Brunsdon, C. and Charlton, M., 2003, "Geographically weighted regression : The analysis of spatially varying relationships", John Wiley & Sons, p231.

Fox, J. C., Ades, P. K. and Bi, H., 2001, "Stochastic structure and individual-tree growth models", *Forest Ecology and Management*, **154**, 1, 261-276.

Hadayeghi, A., Shalaby, A. S. and Persaud, B. N., 2010, "Development of planning level transportation safety tools using geographically weighted Poisson regression", *Accident Analysis & Prevention*, **42**, 2, 676-688.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A., 2005, "Very high resolution interpolated climate surfaces for global land areas", *International journal of climatology*, **25**, 15, 1965-1978.

Hooman, H.A., 2001, "Multivariate data analysis in behavioral research", *Parsa Pub*, 28-29.

Hudson, G. and Wackernagel, H., 1994, "Mapping temperature using kriging with external drift : Theory and an example from Scotland", *International journal of Climatology*, **14**, 1, 77-91.

Huang, J., Huang, Y., Pontius, R. G. and Zhang, Z., 2015, "Geographically weighted regression to measure spatial variations in correlations between water pollution versus land use in a coastal watershed", *Ocean & Coastal Management*, **103**, 14-24.

Ivajnšič, D., Kaligarič, M. and Žiberna, I., 2014, "Geographically weighted regression of the urban heat island of a small city", *Applied Geography*, **53**, 341-353.

Jarque, C. M. and Bera, A. K., 1980, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics letters*, **6**, 3, 255-259.

Johansson, B. and Chen, D., 2003, "The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modeling", *International Journal of Climatology*, **23**, 12, 1523-1535.

Kamarianakis, Y., Feidas, H., Kokolatos, G., Chrysoulakis, N. and Karatzias, V., 2008, "Evaluating remotely sensed rainfall estimates using nonlinear mixed models and geographically weighted regression", *Environmental Modelling & Software*, **23**, 12, 1438-1447.

Kampata, J. M., Parida, B. P. and Moalafhi, D. B., 2008, "Trend analysis of rainfall in the headstreams of the Zambezi River Basin in Zambia", *Physics and Chemistry of the Earth, Parts A/B/C*, **33**, 8, 621-625.

Konrad, II C. E., 1996, "Relationships between precipitation event types and topography in the southern Blue Ridge Mountains of the southeastern USA", *Int. J. Climatology*, **16**, 1, 49-62.

Kuhn, M. and Johnson, K., 2013, "Applied predictive modeling (61-92)", *Springer publication*, 61-92.

Lee, K. H. and Schuett, M. A., 2014, "Exploring spatial variations in the relationships between residents' recreation demand and associated factors : A case study in Texas", *Applied Geography*, **53**, 213-222.

Li, S., Zhao, Z., Miaomiao, X. and Wang, Y., 2010, "Investigating spatial non-stationary and scale-dependent relationships between urban surface temperature and environmental factors using geographically weighted regression" *Environmental Modelling & Software*, **25**, 12, 1789-1800.

Lin, F. J., 2008, "Solving multicollinearity in the process of fitting regression model using the nested estimate procedure", *Quality & Quantity*, **42**, 3, 417-426.

Lin, C. H. and Wen, T. H., 2011, "Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue", *International journal of environmental research and public health*, **8**, 7, 2798-2815.

Łukawska-Matuszewska, K. and Urbański, J. A., 2014, "Prediction of near-bottom water salinity in the Baltic Sea using Ordinary Least Squares and Geographically Weighted Regression models", Estuarine, *Coastal and Shelf Science*, **149**, 255-263.

Mishra, U., Lal, R., Liu, D. and Van Meirvenne, M., 2010, "Predicting the spatial variation of the soil organic carbon pool at a regional scale", *Soil Science Society of America Journal*, **74**, 3, 906-914.

Naud, C. M., Miller, J. R. and Landry, C., 2012, "Using satellites to investigate the sensitivity of longwave downward radiation to water vapour at high elevations", *J. Geophys. Res*., **117**, 464-480.

Pratt, B. and Chang, H., 2012, "Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales", *J. Hazard. Mater*., **209-210**, 48-58.

Ross, R. J. and Elliott, W. P., 1996, "Tropospheric water vapour climatology and trends over North America : 1973-93", *Journal of Climate*, **9**, 12, 3561-3574.

Sadeghi, A. R., Kamgar-Haghighi, A. A., Sepaskhah, A. R., Khalili, D. and Zand-Parsa, S., 2002, "Regional classification for dryland agriculture in southern Iran", *Journal of Arid Environments*, **50**, 2, 333-341.

Serrano, A., Mateos, V. L. and Garcia, J. A., 1999, "Trend analysis of monthly precipitation over the Iberian Peninsula for the period 1921-1995", Physics and Chemistry of the Earth, Part B: Hydrology, *Oceans and Atmosphere*, **24**, 1, 85-90.

Siordia, C., Saenz, J. and Tom, S. E., 2012, "An introduction to macro-level spatial nonstationarity : A geographically weighted regression analysis of diabetes and poverty", *Human Geographies*, **6**, 5-13.

Singh, P. and Kumar, N., 1997, "Effect of orography on precipitation in the western Himalayan region", *Journal of Hydrology*, **199**, 1, 183-206.

Staub, C. G., Stevens, F. R. and Waylen, P. R., 2014, "The geography of rainfall in Mauritius : Modelling the relationship between annual and monthly rainfall and landscape characteristics on a small volcanic island", *Applied Geography*, **54**, 222-234.

Su, Y. F., Foody, G. M. and Cheng, K. S., 2012, "Spatial non-stationarity in the relationships between land cover and surface temperature in an urban heat island and its impacts on thermally sensitive populations", *Landscape and Urban Planning*, **107**, 2, 172-180.

Szymanowski, M. and Kryza, M., 2012, "Local regression models for spatial interpolation of urban heat island-an example from Wrocław, SW Poland", *Theoretical and Applied Climatology*, **108**, 1-2, 53-71.

Tian, F., Qiu, G. Y., Yang, Y., H., Xiong, Y. J. and Wang, P., 2012, "Studies on the relationships between land surface temperature and environmental factors in an inland river catchment based on geographically weighted regression and MODIS Data", Selected Topics in Applied Earth Observations and Remote Sensing, *IEEE Journal of*, **5**, 3, 687-698.

Tonkaz, T., Çetin, M. and Tülücü, K., 2007, "The impact of water resources development projects on water vapour pressure trends in a semi-arid region", Turkey, *Climatic change*, **82**, 1-2, 195-209.

Trenberth, K. E. and Stepaniak, D. P., 2003, "Seamless poleward atmospheric energy transports and implications for the Hadley circulation", *Journal of Climate*, **16**, 22, 3706-3722.

Um, M. J., Yun, H., Jeong, C. S. and Heo, J. H., 2011, "Factor analysis and multiple regression between topography and precipitation on Jeju Island", Korea, *Journal of Hydrology*, **410**, 3, 189-203.

Wang, K., Zhang, C. and Li, W., 2012, "Comparison of geographically weighted regression and regression kriging for estimating the spatial distribution of soil organic matter", *GIScience & Remote Sensing*, **49**, 6, 915-932.

Wang, K., Zhang, C. and 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.

Wentz, F. J., Ricciardulli, L., Hilburn, K. and Mears, C., 2007, "How much more rain will global warming bring?", *Science*, **317**, 5835, 233-235.

Yue, S. and Hashino, M., 2003, "Long term trends of annual and monthly precipitation in Japan", *Journal of the American Water Resources Association*, **39**, 5-587-596.

Zhang, C., Tang, Y., Xu, X. and Kiely, G., 2011, "Towards spatial geochemical modeling : Use of geographically weighted regression for mapping soil organic carbon contents in Ireland", *Applied Geochemistry*, **26**, 7, 1239-1248.

Zhao, N., Yang, Y. and Zhou, X., 2010, "Application of geographically weighted regression in estimating the effect of climate and site conditions on vegetation distribution in Haihe Catchment", China, *Plant ecology*, **209**, 2, 349-359.