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

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 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 km² 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.
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:

$$ Y_i = a_0(u_i, v_i) + \sum_{k} a_k (u_i, v_i) X_{ik} + \epsilon_i $$

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 $\epsilon_i$ 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

<table>
<thead>
<tr>
<th></th>
<th>WV</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Elevation</th>
<th>Slope</th>
<th>Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>WV</td>
<td>1</td>
<td>(0.626^*)</td>
<td>(-0.781^*)</td>
<td>(-0.671^*)</td>
<td>(-0.187^*)</td>
<td>(0.026)</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.138)</td>
</tr>
<tr>
<td>Longitude</td>
<td>Coefficient correlation</td>
<td>1</td>
<td>(-0.89^*)</td>
<td>(-0.022)</td>
<td>(-0.054^*)</td>
<td>(-0.018)</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>(0.000)</td>
<td>(0.203)</td>
<td>(0.002)</td>
<td>(0.313)</td>
<td></td>
</tr>
<tr>
<td>Latitude</td>
<td>Coefficient correlation</td>
<td>1</td>
<td>(0.307^*)</td>
<td>(0.122^*)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.623)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>Coefficient correlation</td>
<td>1</td>
<td>(0.407^*)</td>
<td>(-0.06^*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>Coefficient correlation</td>
<td>1</td>
<td></td>
<td>(-0.029)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Significance</td>
<td></td>
<td></td>
<td>(0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspect</td>
<td>Coefficient correlation</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation is significant at the 0.01 level**

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 I 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 \(Z_i\), \(W_i\), \(M_i\), \(N_i\), \(O_i\), \(V_i\) and \(T_i\). 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>t-statistic</th>
<th>Probability</th>
<th>Robust SE</th>
<th>Robust t</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>127.87</td>
<td>14.71</td>
<td>8.689</td>
<td>0.000*</td>
<td>14.681</td>
<td>8.709</td>
<td>-</td>
</tr>
<tr>
<td>Longitude</td>
<td>0.000008</td>
<td>0</td>
<td>22.961</td>
<td>0.000*</td>
<td>0</td>
<td>23.0641</td>
<td>7.5015</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.000005</td>
<td>0</td>
<td>-11.301</td>
<td>0.000*</td>
<td>0</td>
<td>-11.323</td>
<td>8.2593</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.004242</td>
<td>0.000063</td>
<td>-67.599</td>
<td>0.000*</td>
<td>0.00006</td>
<td>-63.949</td>
<td>2.0084</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0743</td>
<td>0.0044</td>
<td>16.699</td>
<td>0.000*</td>
<td>0.00486</td>
<td>15.279</td>
<td>1.2283</td>
</tr>
<tr>
<td>Aspect</td>
<td>-0.00036</td>
<td>0.00036</td>
<td>-0.299</td>
<td>0.764</td>
<td>0.00036</td>
<td>-0.297</td>
<td>1.0034</td>
</tr>
</tbody>
</table>

**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.
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:

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

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 0.0002 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.

Using 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.
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).

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 Oman Sea and moving towards higher latitudes and heights.
Fig. 8 (a-f). Spatial distribution of local coefficients of geographic features
### TABLE 4
Estimated parameters and diagnostic statistics in OLS and GWR models

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS model</th>
<th>GWR model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Intercept</td>
<td>127.87</td>
<td>2.29</td>
</tr>
<tr>
<td>Longitude</td>
<td>0.000008</td>
<td>-9.98</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.000005</td>
<td>-11.4</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.004242</td>
<td>-0.002</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0743</td>
<td>-0.019</td>
</tr>
<tr>
<td>Aspect</td>
<td>-0.00036</td>
<td>-0.0009</td>
</tr>
<tr>
<td>Number of observations</td>
<td>3338</td>
<td>3338</td>
</tr>
<tr>
<td>Number of variables</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.8478</td>
<td>0.967</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.8475</td>
<td>0.968</td>
</tr>
<tr>
<td>AIC</td>
<td>14559.044</td>
<td>9329.38</td>
</tr>
</tbody>
</table>

*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.

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...
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.

R² and adjusted R² 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 R² Values, the GWR are chosen as the preferred model (Table 4).

Fig. 10 shows the spatial distribution of local R². According to this figure, the highest frequency of R² (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.
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
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 estimate 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 R² and adjusted R² 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 R² 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


