EXPLORING THE MEAN ANNUAL PRECIPITATION AND TEMPERATURE VALUES OVER TURKEY BY USING ENVIRONMENTAL VARIABLES

Pınar Aslantaş Bostan [a], Zuhal Akyürek [b]

[a] METU, Geodetic and Geographic Inf. Technologies Natural and Applied Sciences-06531 Ankara, Turkey-aslantas@metu.edu.tr
[b] METU, Civil Eng. Dept, 06531 Ankara, Turkey- zakyurek@metu.edu.tr

KEYWORDS: Precipitation; Temperature; Spatial Analysis; Exploration; Co-kriging; Geographically Weighted Regression.

ABSTRACT:

Global warming and climate change that threat seriously population living as well as agriculture, environment, economy, and industry are very important topics for all over the world. Therefore efficient usage of available water resources must be considered carefully to be prepared for all contingencies. Exploring spatial distribution and variation of precipitation and temperature that occurs in the length of time can give an idea about water resources in future. According to Love (1999), trends in water thought likely to have the greatest influence on the future situation including population growth, economic expansion and, in the longer term, climate change.

In this study**,** mean annual precipitation and temperature values measured at 225 meteorological observations over Turkey are used for visualization, exploration and modeling processes to reveal spatial distribution of mean annual precipitation and temperature values. Data components were obtained from the Turkish State Meteorological Service for 34 years period (1970-2003). The basic objectives of the study are: to infer the nature of spatial variation of precipitation and temperature over Turkey based on meteorological observations and to model the pattern of variability of these data components by using secondary variables extracted from DEM. Visualization that gives an initial impression about the data is implemented by using proportional symbols. Exploration that provides good descriptions of the data is performed by using spatial moving averages and co-variogram for first and second order effects. Modeling part is implemented with Co-kriging (COK) and Geographically Weighted Regression (GWR) techniques with using secondary variables such as elevation, aspect, distance to coastline, distance to river, roughness, drop (elevation differences between station and grid), and plan-profile curvature. Correlations among the listed variables were analyzed and highly correlated ones were removed from the analysis. These two approaches are evaluated and discussed in finding the optimum spatial distribution of mean annual precipitation and temperature over Turkey.

1. INTRODUCTION

Geographical variables can not be measured at all part of space, therefore researchers who work with those variables generally use interpolation techniques in some part of their studies. Thus observations are taken at points and spatial interpolation is used to obtain a full spatial coverage. There are many examples such as; soil physical properties, air quality, groundwater pressure, plant species abundance (Heuvelink, 2006).

A common theme in many similar studies is that techniques that make use of the relation between precipitation and secondary data, such as elevation data, often provide more accurate estimates than approaches that are based only one parameter like precipitation measurements (Lloyd, 2005).

This study is concerned with mapping annual average precipitation and temperature for Turkey from sparse point data using Co-kriging (COK) and Geographically Weighted Regression (GWR) methods. By using the spatial relationships between meteorological observations and variables derived from DEM, optimum spatial distributions of mean annual precipitation and temperature are aimed to be defined.

2. STUDY AREA AND DATA

The study area covers all Turkey. The data used in this study is obtained from the Turkish State of Meteorological Service. Data consist of mean monthly precipitation and temperature values measured at big climate stations between 1970-2003 years. Analyses are performed on annual average values. Data from 225 meteorological stations as illustrated in Figure 1, were selected to use in the analyses because of their consistent number of data years and length in the observation period.

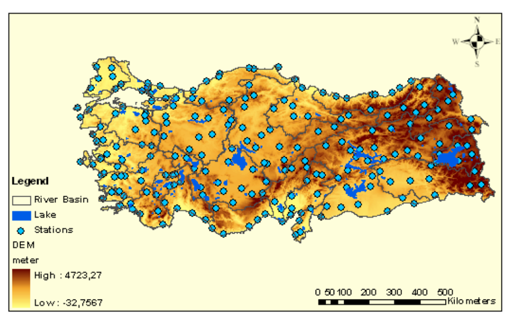

Figure 1. Distribution of meteorological stations on digital elevation model (DEM) of Turkey.

2.1. Variables Obtained From Digital Elevation Model

The use of digital elevation data to guide the interpolation of monthly temperatures is becoming accepted practice, but the spatial variability of air temperature is significantly affected by topographic relief together with several other geographical factors such as latitude and distance to coast line (Rigol et al., 2000). Therefore in this study as well as topographical parameters, some geographical variables were selected as the ancillary variables in finding the distribution of precipitation and temperature. Totally 9 topographical and geographical variables as listed in Table 1, are used as additional input data for spatial interpolation analyses. All the variables except station elevation are obtained from SRTM3 and river network digitized from 1/250000 scaled map.

Table 1. Ancillary data derived from DEM and river network.

Variable	**Description**
Elevation	Height of meteorological stations
Aspect	Function of aspect derived from DEM
Curvature	Degree of curvature derived from DEM
Roughness	DEM cell height minus 5 km grid mean height
Drop	DEM cell height minus 5 km grid minimum height
West	Distance to nearest west coast, calculated by using the Euclidean distance computation method
South	Distance to nearest south coast, calculated by using the Euclidean distance computation method
North	Distance to nearest north coast, calculated by using the Euclidean distance computation method
River	Distance to nearest river, calculated by using the Euclidean distance computation method

3. METHODOLOGY

Flowchart of the methodology applied in this study as illustrated in Figure 2, is composed of visualisation, exploration for first and second order effects and modeling part.

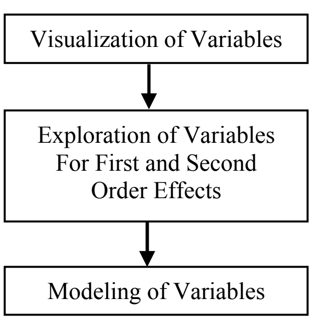

Figure 2. Methodology of study.

3.1. Visualization of Variables

Visualization should be the first step in any spatial data analysis to get an initial impression about the data for ones who deal with spatial nature of data. In this study visualization part is carried out by using proportional symbols. However it is hard to come to any conclusions purely on the basis of a visual analysis. Exploration and modeling of variables should be implemented to make comprehensive analyses.

3.2. Exploration of Variables

Data exploration is aimed at developing hypotheses and makes extensive use of graphical views of the data such as maps or scatter plots. Modeling of spatial phenomena has to incorporate the possibility of spatial dependence in order to provide a true representation of the existing effects. Such spatial effects can be either large scale trends or local effects. The first is also called as a **first order effect** and it describes overall variation in the mean value of a parameter such as rainfall. In this study spatial moving averages method is used as exploration for first order effects. The second which is named as **second order effect** is produced by spatial dependence and represents the tendency of neighboring values to follow each other in terms of their deviation from the mean. The presence of second order effects would result in positive covariance between observations a small distance apart and lower covariance or correlation if they are further apart (Pfeiffer, 1996).

3.3. Modeling of Variables

i. Inverse Distance Weighting (IDW) estimate is a weighted average of the data available in a specific neighborhood. As the exponent becomes larger the weight assigned to observations at large distances from the estimation location becomes smaller. In other words, as the value of the exponent is increased, the estimate at a given location becomes more similar to the closest observations. The exponent is usually set to 2 (so, inverse square distances are used in estimation), as is the case in this study (Lloyd, 2005).

ii. Geographically weighted regression (GWR) is a local statistical technique to analyze spatial variations in relationships which is based on Tobler’s (1970) “First law of geography”: everything is related with everything else, but closer things are more related.

The simple linear model usually fitted by ordinary least squares (OLS) methods is given in Equation (1).

$$P= Co+C1(H)+C2(A)+e \qquad (1)$$

- P=rainfall (mm)
- Co=rainfall at sea level (mm) and flat area
- $C1$= dimensionless rate of increase in rainfall with altitude, or height coefficient (mm/m)
- H=station altitude (m)
- $C2$= change of rainfall with aspect
- A= aspect of that station
- e= error term

In GWR by retaining the same linear model, we can allow parameters, the intercept constant, the height and aspect coefficient to change, or ‘drift’, over space. That is, if (x, y) is a coordinate pair, the simple linear model of Equation (1) can be expanded to Equation (2).

$$P= Co(x,y)+C1\,(x,y)(H)+C2\,(x,y)(A)+e \qquad (2)$$

This revised model as seen in Equation (2), allows the coefficients to vary as continuous functions over space, so that each may be thought
of as a three-dimensional surface over the geographical study area rather than as a single, fixed, real number (Brunsdon et al., 2001).

iii. Co-kriging (CO) method is an extension of ordinary kriging that takes into account the spatial cross-correlation from two or more variables. The usual situation is one where the primary or target variable, $Z_u(x)$, has been measured at many fewer places, x, than the secondary one, $Z_v(x)$, with which it is co-regionalized.

The influence of the secondary information on estimating *Z* depends on (i) the correlation between primary and ancillary variables, (ii) the spatial continuity of the attributes, and (iii) the sampling density and spatial configuration of primary and ancillary variables (Simbahan et al., 2005).

4. ANALYSIS AND RESULTS

4.1. Visualization of Variables

By using proportional symbols, precipitation and temperature values are analyzed visually as seen in Figure 3 and 4.

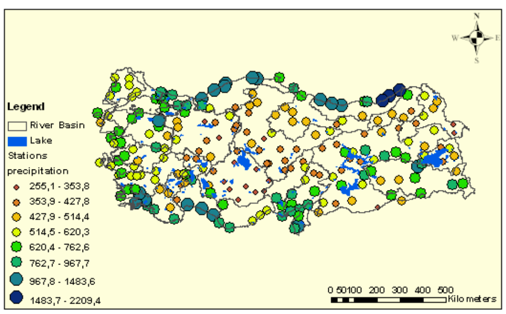

Figure 3. Mean annual precipitation distribution.

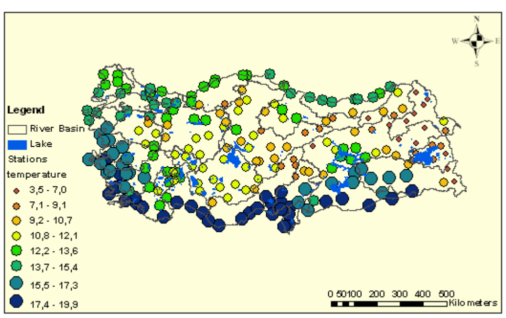

Figure 4. Mean annual temperature distribution.

As it is illustrated in Fig. (3), south, west and north coasts and some parts of south-eastern Anatolia have higher precipitation. The continental interiors of Turkey have lower precipitation than other regions. Mean temperature distribution is observed high at south, south-eastern, and west Anatolia as presented in Fig. (4).

4.2. Exploration of Variables

i. First Order Effects

The spatial moving average method with rectangular kernel, size equal to 180 km*180 km is used as exploration for first order effects (Figure 5 and 6).

From Figure (5), it is observed that density of stations by considering precipitation item is higher at west, north-west, north-east, and south regions of Turkey. From Figure (6), it is understood that west and south regions of Turkey have higher point density based on temperature values.

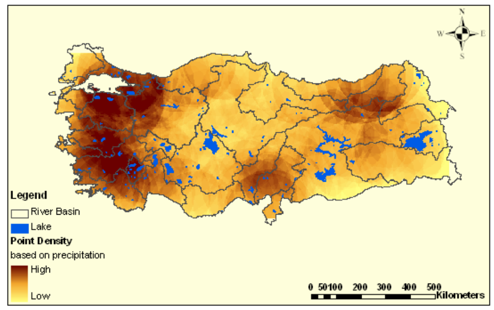

Figure 5. Point density of stations based on precipitation values.

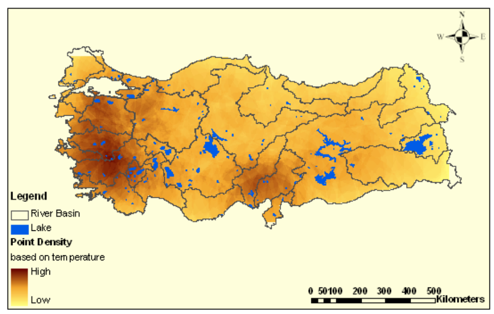

Figure 6. Point density of stations based on temperature values.

ii. Second Order Effects

Co-variograms of precipitation and temperature are analyzed. For precipitation, correlation among values obtained low even in small distances. When range is 180 km distance, correlation is zero among values. As stations become farther apart, they have more dissimilar precipitation values. It means that, spatial dependency among precipitation values is low.

Differently from precipitation values, correlations among temperature values of stations are higher. Correlations become zero beyond 278 km. Positive covariance is higher in temperature values than precipitation in specific distances.

4.3. Modeling of Variables

i. GWR Application

The output obtained from GWR can be voluminous. Predicted precipitation, temperature values, local r-square values and root-mean-square errors (RMSE) for each meteorological station are used for discussion. Cross validation is used to compare the prediction performances.

It is a desired condition that predictions should be unbiased (centered on the measurement values) after interpolation. If the prediction errors are unbiased, the mean prediction error should be near zero. Mean errors of precipitation and temperature are obtained close to zero and prediction mean error of precipitation is obtained lower than prediction mean error of temperature (Table 2).

There is also very important topic that should be considered carefully when performing interpolation; predictions obtained after interpolation should be as close to the measurement values as possible. If the RMSE are small, better predictions are obtained. RMSE of temperature is lower than RMSE of precipitation as illustrated in Table (2). Prediction of temperature is better when using derived secondary variables based on RMSE values.

Table 2. Cross validation results of GWR prediction

Precipitation		Temperature	
Mean error	RMSE	Mean error	RMSE
0,0036	0,8717	0,0047	0,7925

The predicted precipitation and temperature values obtained from GWR are presented in Figures 7 and 8.

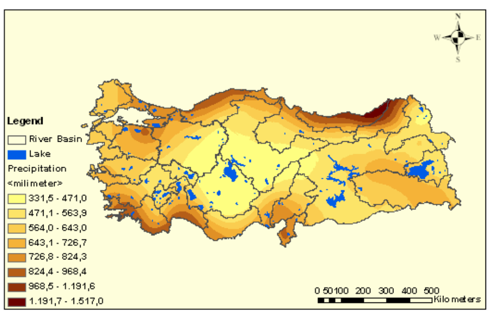

Figure 7. Precipitation prediction with GWR.

North, south, and west coasts and south-eastern of Turkey have more precipitation. Average of annual precipitation in long term is comparatively lower than other regions in Central Anatolia (331-471 mm) (Figure 7).

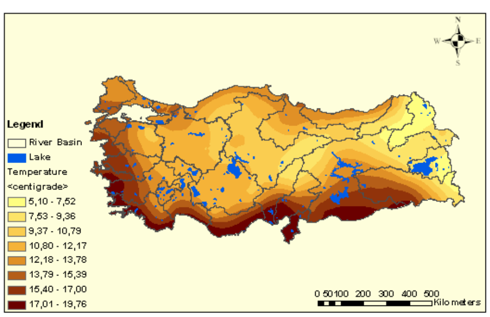

Figure 8. Temperature prediction with GWR.

In respect of Fig. (8), south, south-eastern and west coasts of Turkey have higher mean annual temperature values than other regions. In the north-east part of Anatolia this value is low (5-8 centigrade).

Local r-square values calculated for each meteorological station are interpolated with kriging operation. Mean errors and RMSE values are very low for both precipitation and temperature as seen in Table (3). So generated maps that show local r-square distribution are very confident (Figure 9 and 10).

Table 3. Cross validation results of GWR local r-square.

Precipitation		Temperature	
Mean error	RMSE	Mean error	RMSE
0,0043	0,4754	-0,0005	0,3833

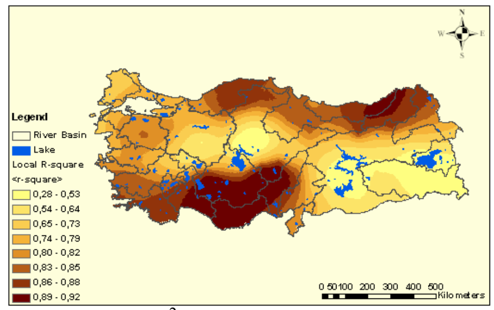

Figure 9. Local r^2 map of precipitation.

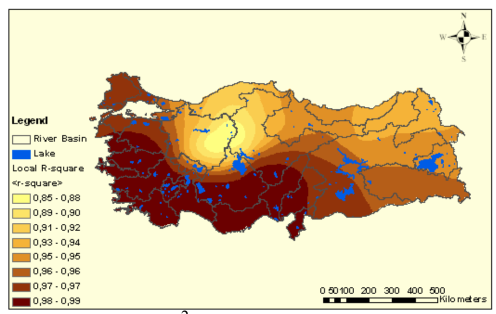

Figure 10. Local r^2 map of temperature.

High r-square values (0,89-0,92) are observed at the south, north-east and north-west regions of Turkey (Figure 9). This indicates that the model best fits to these regions when predicting precipitation. Also it can be reported that, the effects of secondary variables on spatial distribution of precipitation are not so significant for south-east and central Anatolia.

In contrast to precipitation r-square map, r-square values of temperature are considerably high for all Turkey. At south parts, r-square values exceed to 0,99. Variables that are extracted from DEM are very suitable when extracting temperature spatial distribution for whole Turkey.

In addition, measured and predicted values obtained from GWR are compared for two meteorological variables. Having high r^2 values, 0,826 for precipitation and 0,965 for temperature, indicates truthful predictions are obtained with GWR.

RMSE distributions of GWR predictions for precipitation and temperature are mapped to highlight the over and underestimated regions. The underestimated regions for precipitation are appeared at the north, north-west, and some parts of south-eastern Anatolia (Figure 11). RMSE values of temperature prediction are negatively high at the eastern parts of Turkey (Figure 12). The overestimated regions are dominated at south-eastern Anatolia.

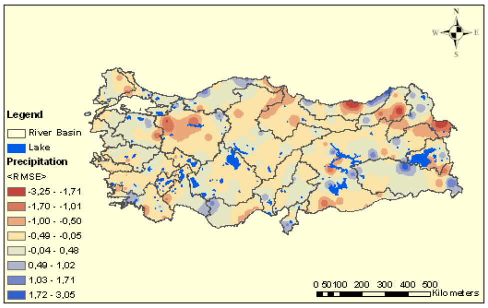

Figure 11. RMSE distribution of GWR precipitation prediction.

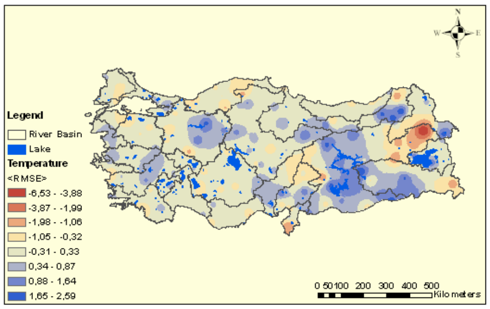

Figure 12. RMSE distribution of GWR temperature prediction.

The inadequate number and non-uniform distribution of meteorological stations over Turkey can be one of the cause of over and underestimations.

ii. Co-Kriging Application

In co-kriging application, only three secondary variables can be used due to restrictions of used software. Because of this limitation the secondary variables are grouped into nine different combinations, where each of them consists of three variables. DEM and aspect are used in all combinations since they are considered to be the most identifier variables to predict precipitation and/or temperature.

All of the variable combinations give low mean errors when predicting precipitation (Table 4). Mean error of “d” variable combination (-0,0133) is closest to zero. Mean error of “g” variable combination has maximum mean error value (-0,021). According to temperature cross validation results; mean errors are lower than precipitation mean errors. Mean errors of “f” and “h” variable combinations are closest to zero (0,0049).

Like mean errors, temperature RMSE values are lower than precipitation RMSE values. According to precipitation cross validation results (Table 4), RMSE values are very close to each other and all of them are below 1. Minimum RMSE values are obtained by “g” (0,9321); “e” (0,9324); and “i” variable combinations (0,9325). The same variable combinations have minimum RMSE values in temperature prediction (0,7081, 0,7093, 0,7203).

Table 4. Cross validation results of co-kriging analysis.

	Precipitation		**Temperature**	
	Mean error	RMSE	Mean error	RMSE
a)DEM-Asp-Curv	-0,0154	0,9332	0,0091	0,7592
b)DEM-Asp-Drop	-0,0145	0,9331	0,0111	0,7563
c)DEM-Asp-Rough	-0,0155	0,9332	0,0089	0,7595
d)DEM-Asp-River	-0,0133	0,9345	0,0093	0,7682
e)DEM-Asp-South	-0,0191	0,9324	-0,0068	0,7203
f)DEM-Asp-North	-0,0158	0,9352	0,0049	0,7454
g)DEM-Asp-West	-0,0210	0,9321	-0,0055	0,7081
h)DEM-River-North	-0,0162	0,9363	0,0049	0,7508
i)Asp-South-West	-0,0198	0,9325	-0,0065	0,7093

The predictions of the combinations giving minimum mean errors are presented in Figure 13 and 14.

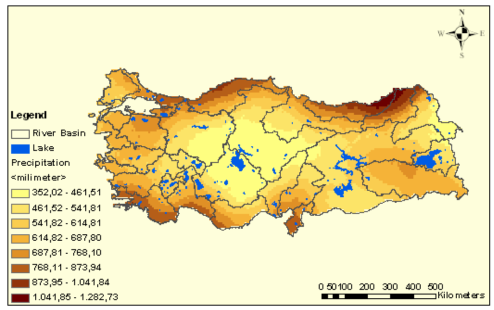

Figure 13. Precipitation prediction with “d” combination.

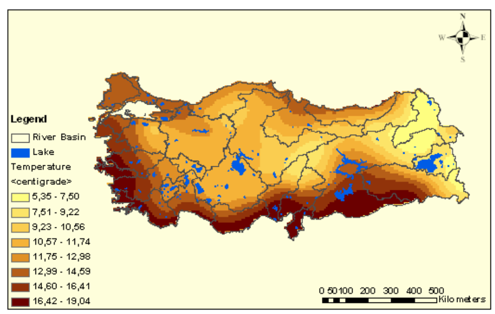

Figure 14. Temperature prediction with “c” combination.

The RMSE values should not be used alone in order to decide whether an interpolation method yields the best interpolation. Other issues such as the density and location of measurement points (bias) need to be considered (Carrera-Hernandez and Gaskin, 2006). For this purpose comparison between measured and predicted values obtained from co-kriging methods are made. As it is presented in Table (5), “b” and “d” variable combinations resulted with the highest r^2 values for precipitation estimation (0,418 and 0,425). These variable combinations also have minimum mean errors. So elevation, aspect and drop that are related with topography and nearest distance to river related with geography are made better approximations to derive the nature of spatial variation of precipitation.

Table 5. R^2 values between measurements and predictions.

	Precipitation	Temperature
a)	0,403	0,814
b)	0,418	0,814
c)	0,402	0,814
d)	0,425	0,814
e)	0,369	0,795
f)	0,4	0,808
g)	0,364	0,784
h)	0,402	0,807
i)	0,364	0,784

Similar to GWR analysis results, r^2 values for co-kriging analysis are obtained higher for temperature prediction than precipitation prediction. The highest ones are obtained with “a”, “b”, “c” and “d” combinations ($r^2 = 0,814$). Additionally “f” and “h” combinations have high r^2 values ($r^2 = 0,808$ and 0,807). Topographic factors (elevation, aspect, curvature, drop, roughness) and nearness to north coast and river are descriptive variables to infer the nature of spatial variation of temperature.

5. CONCLUSION

In this study**,** mean annual precipitation and temperature values measured at 225 meteorological observations over Turkey are used for visualization, exploration and modeling processes to reveal spatial distribution of mean annual precipitation and temperature by using secondary variables derived form DEM. Visualization which should be the first step in any spatial data analysis is performed with proportional symbols. According to visualization of meteorological observations, south, north and west coasts and south-eastern Anatolia have higher precipitation. On the other hand south, south-eastern, and west Anatolia have higher temperature values. Exploration for first order effects that gives the intensity variation of variables in the study region is performed with spatial moving averages method. Similar results can be observed with spatial moving averages and proportional symbols. Exploration for second order effects explores the spatial dependence of deviations in attribute values from their mean. Correlation among precipitation values is obtained as low even in small distances, whereas spatial dependence of temperature values is higher than the precipitation values. In modeling part, GWR and Co-kriging analysis are performed. GWR has provided a means of investigating spatial non-stationary in linear regression models (Brundson et al., 2000). From the outputs of GWR, predicted, local r-square and RMSE values are used to evaluate GWR results. These outputs are interpolated with kriging operation and cross validation results are analyzed. Mean and RMS errors are low for both meteorological variables. Also input data set is very appropriate when extracting temperature spatial distribution for whole Turkey due to high local r-square values. The maps of the local r^2 indicate that, as the relation varies locally, the benefits in using secondary data to provide accurate estimation will vary locally. In respect of relationship between measured and predicted values, GWR gives very truthful predictions for both precipitation and temperature.

Co-kriging method takes into account the spatial cross-correlation from two or more variables. Different variable combinations are analyzed and cross validation results are evaluated. Temperature prediction mean and RMS errors are lower than precipitation mean and RMS errors. Also the coefficient of determination, r^2, between measured and predicted values for temperature is very high than precipitation.

Generally GWR gives better predictions for two variable set in respect of r^2 values between predictions and measurements for Turkey. However, there are a variety of issues that should be explored further. Different sources of information such as data from synoptic stations may increase the accuracy of precipitation prediction. Different secondary variables can be considered to improve the precipitation and temperature modelling.

6. REFERENCES

Brundson, C., Fotheringham, S., Charlton, M., Geographically Weighted Regression as a Statistical Model, 2000.

Brundson, C., McClatchey, J., Unwin, D.J., Spatial Variations In The Average Rainfall–Altitude Relationship In Great Britain: An Approach Using Geographically Weighted Regression, Int. J. Climatol. 21: p.455–466 (2001).

Carrera-Hernandez, J.J, Gaskin, S.J., Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico, Journal of Hydrology 336: p.231-249 (2007).

Heuvelink, G.B.M., Incorporating process knowledge in spatial interpolation of environmental variables. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (2006).

Lloyd, C.D, 2005. Assessing the Effect of Integrating Elevation Data Into the Estimation of Monthly Precipitation in Great Britain. Journal of Hydrology 308: p.128–150 (2005).

Love, P., Future trends: water resources: meeting future demand. The journal of futures studies, strategic thinking and policy, vol.01, no.03, p.275-278 (1999).

Pfeiffer, D.U., Issues related to handling of spatial data, Australian Veterinary Association Second Pan Pacific Veterinary Conference, p.83-105 (1996).

Rigol, J.P, Jarvis, C.H., Stuart, N., Artificial neural networks as a tool for spatial interpolation, International Journal of Geographical Information Science, , vol.15, no. 4, p.323-343 (2001).

Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J., Haddix, M.L, Fine-resolution mapping of soil organic carbon based on multivariate secondary data, Geoderma, v132, 3-4, p.471-489 (2005).

Tobler, W.R., A computer movie simulating urban growth in the Detroit region, Economic Geography 46: p.234-240 (1970).