SPATIO-TEMPORAL ANALYSIS OF PRECIPITATION AND TEMPERATURE DISTRIBUTION OVER TURKEY

P. A. Bostan [a], Z. Akyürek [b]

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

"14th International Symposium on Spatial Data Handling"

KEY WORDS: Precipitation, Temperature, Co-kriging, Ordinary Least Square, Geographically Weighted Regression

ABSTRACT:

In this study**,** mean annual precipitation and temperature values observed at 225 meteorological observations over Turkey are used to disclose 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 SRTM and river network. Modeling the spatial distribution of data sets is implemented with Co-kriging (COK), Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) techniques with using secondary variables such as elevation, aspect, distance to river, roughness, drop (elevation differences between station and grid), sd-grid (standard deviation of 5*5 km grid), and plan-profile curvature. Correlations among the listed variables were analyzed and highly correlated ones were removed from the analysis. The study found a presence of high spatial non-stationary in the strength of relationships and regression parameters. The co-kriging interpolation method gave strong relationship for temperature (r^2= 0.823) but comparatively weak relationship for precipitation (r^2= 0.542). OLS method resulted with lower relationships for temperature (r^2= 0.68) and for precipitation (r^2= 0.3). The highest adjusted r^2 values were obtained with GWR method; 0.96 for temperature and 0.66 for precipitation.

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). At the areas that haven't observed values discrete precipitation values are needed. For this purpose thiesen polygons, regression analysis, inverse distance weighted method, trend analysis, isohyets curves etc. can be used for prediction. However these methods are generally inadequate and tendentious and they do not give variation information about predictions (Çetin and Tülücü, 1998). 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 (CO) technique, global ordinary least square regression technique known as OLS and a local regression technique known as Geographically Weighted Regression (GWR) methods. By using the spatial relationships between meteorological observations and variables derived from elevation, 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 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.

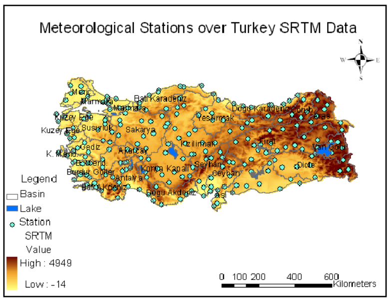


Figure 1: Study area and meteorological stations within the basins

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 elevation (SRTM 3 arc second spatial resolution) and river network digitized from 1/250000 scaled map.

Variable	Description
Elevation	Height of meteorological stations
Aspect	Function of aspect derived from elevation
Curvature	Degree of curvature derived from elevation
Roughness	Elevation cell height minus 5 km grid mean height
Drop	Elevation cell height minus 5 km grid minimum height
Sd_Grid	Stan. deviation of elevation in each 5 km*5 km grid
River	Distance to nearest river, calculated by using the Euclid. distance computation method

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

3. METHODOLOGY

i. 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).

ii. Ordinary Least Squares (OLS) provides a global model of the variable to predict by creating a single regression equation to express that process. It can be regarded as the starting point for all spatial regression techniques (ESRI website).

Theoretical background is summarized below:

Suppose that X is an independent and Y is a dependent variable. We make n number of observations on two variables. The relationship between Y and X can be regressed using OLS as follows:

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is a vector of the observed dependent variable, X is a known model matrix including a column of 1 (for intercept) and n independent variables, β is a vector of unknown fixed-effects parameters, and ε is a random error term. The OLS estimate of β is obtained by the least-squares method as shown in Eq. (2).

$$\beta = (X^T X)^{-1} X^T Y \tag{2}$$

where superscript T denotes the transpose of a matrix. The relationship represented by Eq. (1) is assumed to be universal or constant across the geographic area (Zhang et al, 2005).

iii. 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".

The simple linear model usually fitted by ordinary least squares (OLS) methods is given in Eq. (3).

$$P = Co + C1(H) + C2(A) + e \tag{3}$$

- 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 Eq. (4).

$$P = Co(x,y) + C1\,(x,y)(H) + C2\,(x,y)(A) + e \tag{4}$$

This revised model as seen in Eqn (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).

The assessment of methods is performed with statistical measures of RMSE and r^2.

4. ANALYSIS AND RESULTS

i. Co-Kriging Application

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

From the cross-validation table RMSE values were obtained for each variable set (Table 2). For precipitation (h) variable combination (aspect, drop and standard deviation of grid) had the minimum RMSE values (197 mm). For temperature (f) variable combination (SRTM, drop and standard deviation of grid) had the minimum RMSE values (1,532 C°).

	Prec. RMSE (mm)	**Temp.** RMSE (C°)
a)SRTM-Asp-Curv	222,2	1,561
b)SRTM-Asp-Drop	215,4	1,555
c)SRTM-Asp-Rough	222,04	1,560
d)SRTM-Asp-River	222,9	1,593
e)SRTM-Asp-Sd-Grid	213,8	1,545
f)SRTM- Drop -Sd-Grid	207,6	1,532
g)SRTM-River-Rough	223,04	1,594
h)Asp –Drop- Sd-Grid	197,06	1,535

Table 2. RMSE values of variable combinations.

The predictions of the combinations giving minimum errors are presented in Figure 2a and 2b. For precipitation, estimated values of "h" variable combination are used to obtain prediction map (Figure 2a). According to the Figure north-west, north-east, south and south- east regions of Turkey have more precipitation than other regions (≈1600 mm).

In order to obtain temperature prediction map "f" variable combination estimated values are used (Figure 2b). According to the figure at the west, south and south-east regions of Turkey average annual temperature is higher than the other regions (highest value is 19 C°).

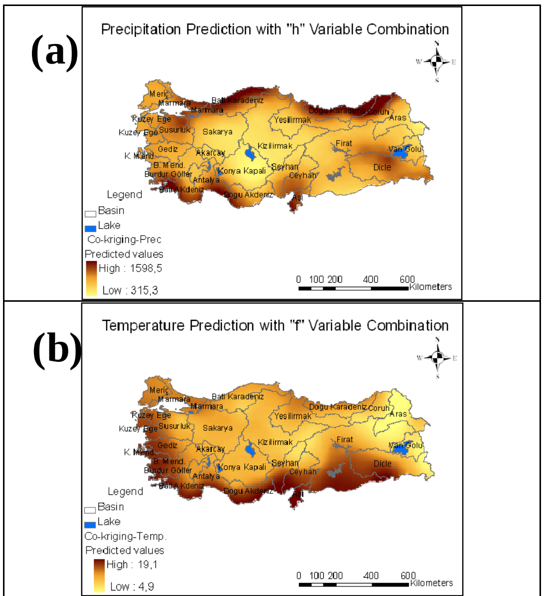


Figure 2. Co-Kriging prediction maps. In (a), precipitation; in (b) temperature prediction maps respectively obtained from "h" and "f"variable combinations are represented.

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 (3), "h" and "f" variable combinations resulted with the highest r^2 values for precipitation estimation (0,542 and 0,499). According to r^2 values of temperature "f" variable combination had the highest r^2 value (0,823). It can be said that parameters that show differences of topography have made better approximations to derive the nature of spatial variation of precipitation and temperature.

	Precipitation	**Temperature**
a)	0,424	0,818
b)	0,461	0,819
c)	0,425	0,818
d)	0,42	0,812
e)	0,469	0,821
f)	0,499	0,823
g)	0,419	0,812
h)	0,542	0,820

Table 3. R^2 values between measurements and predictions.

Interpolation residuals of precipitation and temperature are illustrated in Figure 3a and 3b. Generally precipitation prediction error is high at north and north-east regions of Turkey (Figure 3a). Underestimation and also overestimation is very high at these regions (-1067 and 557 mm). According to the temperature interpolation residuals, underestimation takes place at the south and north coasts of Turkey (Figure 3b).

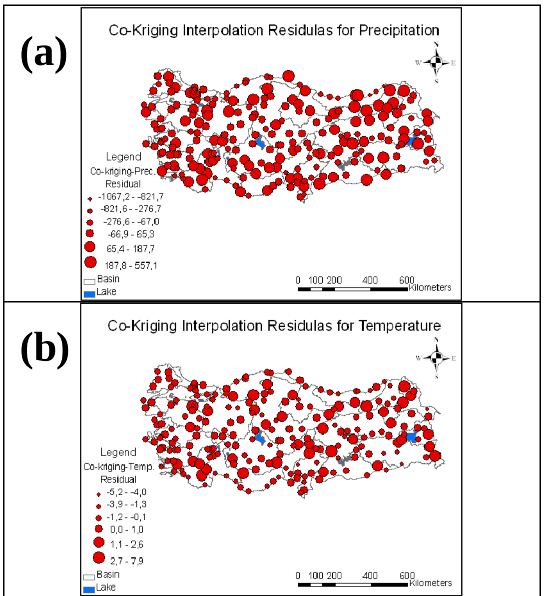


Figure 3. Co-kriging interpolation residuals. In (a), residuals of precipitation prediction; in (b), residuals of temperature prediction respectively obtained with "h" and "f" variable combinations are represented.

ii. OLS Application

OLS method is applied to meteorological variables to make predictions. By using seven secondary variables precipitation prediction is made. According to the t statistic only elevation and Sd_grid variables are statistically significant. Therefore OLS method is applied again by using these two variables. Adjusted r^2 is 0,3.

For temperature data set firstly all independent variables are used in OLS application. According to the t statistic only elevation variable is statistically significant parameter. So in second application only elevation is used as independent variable. Adjusted r^2 is 0,68.

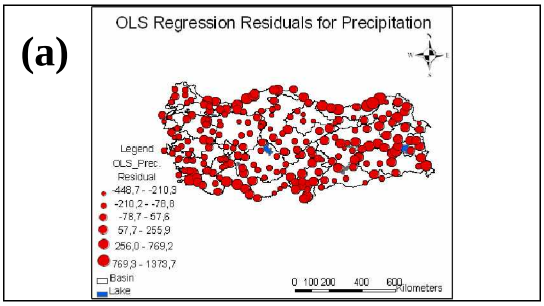
Moran's I technique is applied to precipitation and temperature residuals to control if the residuals are spatially autocorrelated or not. According to the precipitation residuals, index value is 0,23 and and Z score is 5,8. For temperature, index value is 0,73 and Z score is 17,2 (Table 4). For two datasets Moran's I index value indicates a strong clustering. Therefore null hypothesis of randomness is rejected. It can be easily concluded that according to the OLS results, there is statistically significant spatial autocorrelation for both datasets in the study area.

Global Moran's I Summary		
	Precipitation	**Temperature**
Moran's Index	0,238148	0,734636
Expected Index	-0,004464	-0,004464
Variance	0,001748	0,001830
Z Score	5,802919	17,278317

Table 4. Global Moran's I results for precipitation and temperature OLS regression.

In Figure (4a), regression residuals of precipitation prediction are illustrated with graduated symbols. In general, high residuals are located at the north-east Anatolia.

In Figure (4b), regression residuals of temperature prediction are represented. Overestimation is commonly concentrated at west and south regions; however underestimation is concentrated at north regions of Turkey. Figure also supports Moran's I index value with visually.



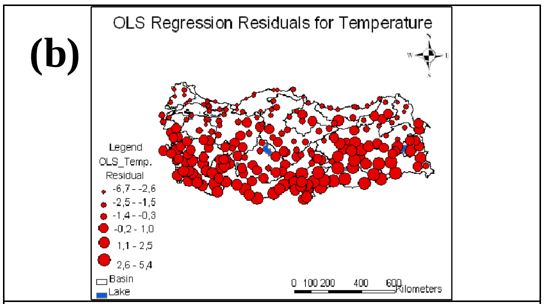


Figure 4. OLS regression residuals. In (a), residuals of precipitation estimation; in (b), residuals of temperature estimation are illustrated.

When there is statistically significant spatial autocorrelation of the regression residuals, the OLS regression is considered to be unreliable. In such circumstances using local regression techniques to model non-stationary variables instead of global techniques can be used to improve predictions. GWR is a frequently used local regression technique.

iii. GWR Application

The output obtained from GWR can be voluminous. Predicted precipitation, temperature values, local r-square values, root-mean-square errors (RMSE) and *t*-values of parameters for each meteorological station are used for discussion. In table (5), RMS errors are shown for two data sets.

Precipitation	**Temperature**
RMSE	RMSE
141,4 mm	0,6 C^o

Table 5. RMSE values of GWR predictions

Measured and predicted values obtained from GWR are compared for two meteorological variables. Having high r^2 values, 0,67 for precipitation and 0,96 for temperature, indicates truthful predictions are obtained with GWR.

Moran's I index values are calculated for GWR residuals (Table 6). For precipitation 0,004 Moran's I index and 0,15 Z score shows there is no spatially correlation. Residuals are randomly distributed across the study area. For temperature -0,11 Moran's I index and -2,0 Z score express tendency toward dispersion.

Global Moran's I Summary		
	Precipitation	**Temperature**
Moran's Index	0,004167	-0,116688
Expected Index	-0,004464	-0,004464
Variance	0,003077	0,003139
Z Score	0,155598	-2,002920

Table 6. Global Moran's I results for precipitation and temperature GWR applications.

The predicted precipitation and temperature values obtained from GWR are mapped and shown in Figure 5a and 5b.

(a)

(b)

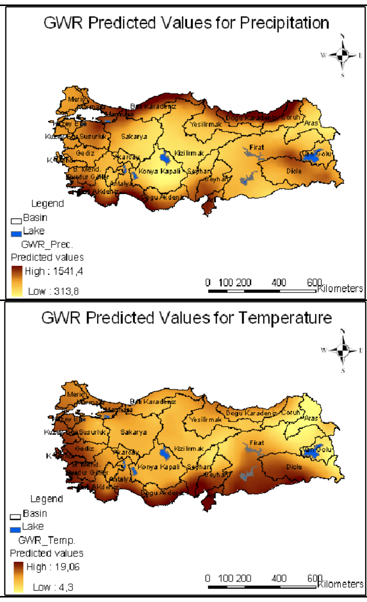


Figure 5. Prediction maps. In (a), precipitation; in (b), temperature prediction maps obtained with GWR application are shown.

North, south, and west coasts and south-eastern of Turkey have more precipitation. Average annual precipitation is comparatively lower than other regions in Central Anatolia (Figure 5a). In respect of Figure (5b), south, south-eastern and west coasts of Turkey have higher mean annual temperature values than other regions.

Local r-square values calculated for each meteorological station are interpolated with kriging operation (Figure 6a and 6b).

(a)

(b)

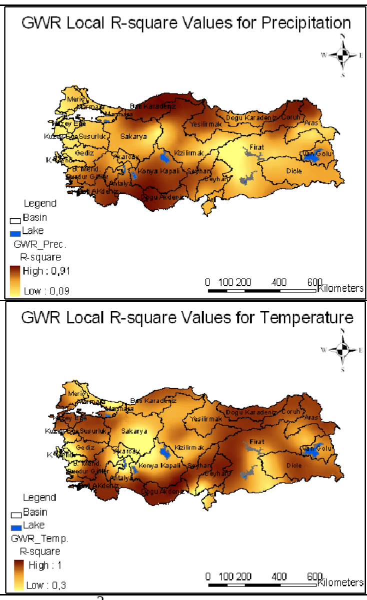


Figure 6. Local r^2 values. In (a), precipitation; in (b), temperature local r^2 values obtained with GWR application are shown.

For precipitation high r-square values were observed at the north-east, north-west and south regions of Turkey (Figure 6a). This indicates that the model best fits these regions when predicting the precipitation values. 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 were considerably high for all Turkey except some regions in central Anatolia (Figure 6b). At south parts, r-square values exceed to 0,99. Elevation parameter is very suitable when extracting temperature spatial distribution for whole Turkey according to local r^2 values of meteorological stations'.

Distribution of prediction errors of precipitation and temperature are mapped to highlight inaccurately estimated regions (Figure 7a and 7b). Both error values are lower than Co-kriging and OLS methods. In Figure 7a precipitation prediction errors are illustrated. Generally error is high at north-east regions of Turkey. In Figure 7b, temperature prediction errors are shown. At west, south-east and central Anatolia error is higher than the other regions.

(a)

(b)

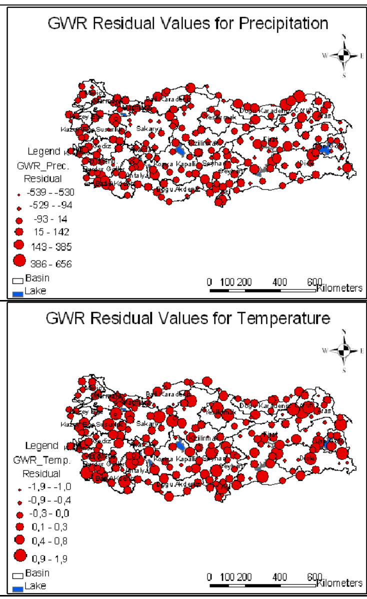


Figure 7. GWR residuals. In (a), precipitation; in (b), temperature residuals obtained with GWR application are shown.

5. CONCLUSION

In this study, mean annual precipitation and temperature values measured at 225 meteorological observations over Turkey are used for revealing spatial distribution of mean annual precipitation and temperature by using secondary variables derived from elevation and river network.

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.
OLS provides a global model of the variable to predict by creating a single regression equation to express that process. According to the results r^2 values are lower for two datasets than Co-kriging and GWR (0,3 for precipitation, 0,68 for temperature). Residuals' Moran's I index value show significant positive autocorrelation. This means that either key secondary variable is missing or the global model is not suitable for this dataset.

GWR has provided a means of investigating spatial non-stationary in linear regression models (Brundson et al., 2000). From the outputs of GWR, predicted values, local r-square values and RMSE of each meteorological stations, and *t*-values of parameters are used to evaluate GWR results. These outputs are interpolated with kriging operation and results are analyzed. The lowest RMS errors are obtained by GWR analysis for both data sets (Table 5). Also it is understood that input data set is very appropriate when extracting temperature spatial distribution for whole Turkey due to high local r-square values according to GWR results (Figure 6b). 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. According to the Moran's I index values there is no spatial autocorrelation for precipitation, for temperature residual pattern is dispersed (Table 6).

According to the three methods, generally GWR gives better predictions for two variable sets in respects of r^2 values and RMS error values between predictions and measurements and Moran's I index values. Also this method can be applied to areas that have no observation with respect of leave-one-out cross validation method.

Spatial distribution of meteorological variables (precipitation and temperature) over Turkey has been defined. In future studies, the temporal variation and distribution of these variables by estimating temporally varying coefficients α_{st}, β_{st} for each meteorological station will be studied in future studies.

4. 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).

Çetin, M., Tülücü, K., Doğu Akdeniz Bölgesinde Aylık Yağışların Yersel Değişimlerinin Jeoistatistik Yöntemle İncelenmesi, Tr. Journal of Engineering and Environmental Science, 22 (1998), p. 279-288.

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).

Jerosch, K., Schlüter, M., Pesch, R., Spatial analysis of marine categorical information using indicator kriging applied to georeferenced video mosaics of the deep-sea Håkon Mosby Mud Volcano, Ecological Informatics, vol 1, issue 4, p. 391-406, 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).

Mas, J. F., Puig, H., Palacio, J. L., Sosa-López, A., Modelling deforestation using GIS and artificial neural networks, Environmental Modelling & Software, vol: 19, issue: 5, p. 461-471, 2004.

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).

Weir-Smith, G., Schwabe, C.A., Spatial interpolation versus neural network propagation as a method of extrapolating from field surveys, GIS Centre, HSRC (Human Sciences Research Council), Pretoria, 2000.

Zhang, L., Gove, J. H., Heath L. S., Spatial residual analysis of six modeling techniques, Ecological Modelling 186 p. 154–177, 2005.

ESRI website:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20OLS%20regression%20works (accessed 25 August 2009).