Spatial analyses of groundwater levels using universal kriging

Kemal Sulhi Gundogdu[1,*] and Ibrahim Guney[2]

[1]*Department of Agricultural Structures and Irrigation, Faculty of Agriculture, Uludag University, Bursa 16059, Turkey.*

[2]*Department of Mathematics, Faculty of Arts and Sciences, Uludag University, Bursa 16059, Turkey.*

[*]*e-mail: kemalg@uludag.edu.tr*

For water levels, generally a non-stationary variable, the technique of universal kriging is applied in preference to ordinary kriging as the interpolation method. Each set of data in every sector can fit different empirical semivariogram models since they have different spatial structures. These models can be classified as circular, spherical, tetraspherical, pentaspherical, exponential, gaussian, rational quadratic, hole effect, K-bessel, J-bessel and stable. This study aims to determine which of these empirical semivariogram models will be best matched with the experimental models obtained from groundwater-table values collected from Mustafakemalpasa left bank irrigation scheme in 2002. The model having the least error was selected by comparing the observed water-table values with the values predicted by empirical semivariogram models. It was determined that the rational quadratic empirical semivariogram model is the best fitted model for the studied irrigation area.

1. Introduction

Keeping the water-table at a favourable level is quite significant for a sustainable irrigation project. Rising of the water-table for various reasons can cause adverse effects on human health and environment as well as crop production. In order to observe water-table continuously, groundwater observation wells are used and monthly measurements are normally recorded (Coram *et al* 2001).

In groundwater observations, it is assumed that the measured values can be applicable for a certain area. The more frequent the measurement network is, the more accurate would be the measurement of water-table. In a scattered groundwater observation net, geostatistical methods can be used to determine the values for the points where measurements are not made or are not feasible to measure due to economic consideration. Spatial interpolation of population characteristic values from data that are limited in number and obtained at irregularly arranged points is an important process for further understanding geostatistical structure in the natural fields. Geostatistics provides a set of statistical tools for analyzing spatial variability and spatial interpolation. A semivariogram is used to describe the structure of spatial variability. Kriging provides the best linear unbiased estimation for spatial interpolation. Nowadays, geostatistics has become a popular means to describe spatial patterns and to interpolate the attribute of interest at unsampled locations. Geostatistical methods have been increasingly used in many disciplines, such as mining, meteorology, hydrology, geology, remote sensing, soil science, ecology and environmental science (Chirlin and Dagan 1980; Bastin *et al* 1984; Hill and Alexandar 1989; White *et al* 1997 and Duc *et al* 2000).

In a study performed by Kumar and Ahmed (2003) monthly water-level data from a small watershed in a hard rock region of southern India have been analyzed geostatistically. They have

Keywords. Universal kriging; semivariogram; groundwater; water-table.

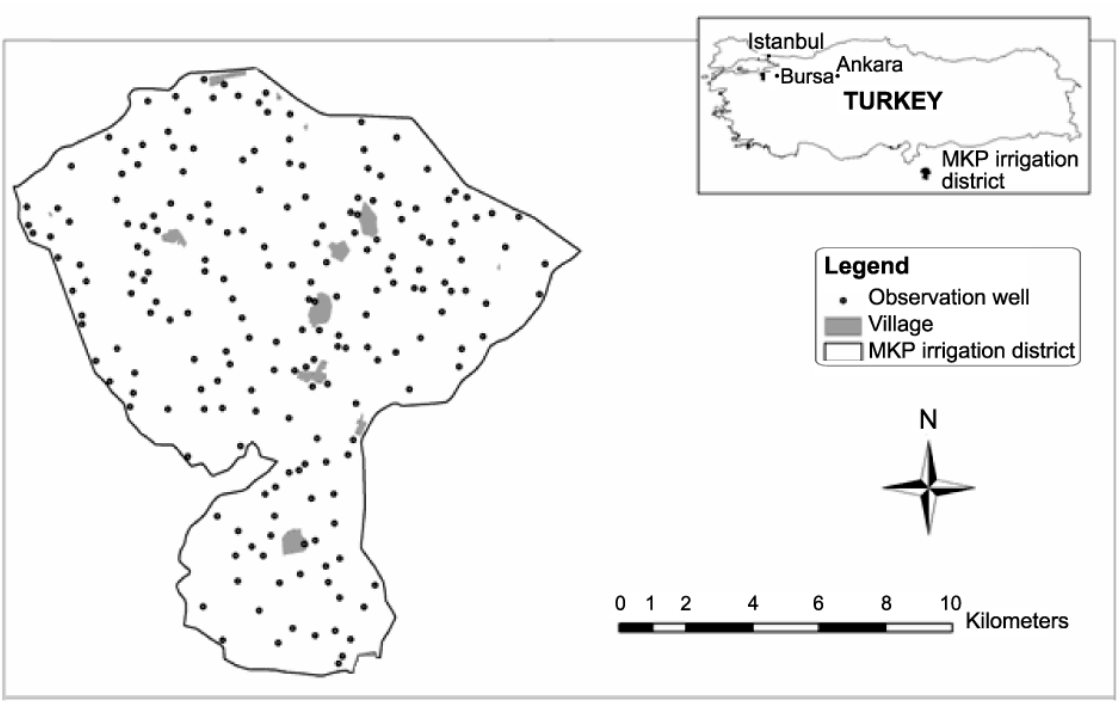

Figure 1. MKP left bank irrigation area and groundwater observation wells.

attempted to make a common variogram(s) for different time periods in a year to estimate water levels on the grids of an aquifer model for calibration purposes.

The semivariogram plays a central role in the analysis of geostatistical data using the kriging method. Before kriging is performed, a valid semivariogram model has to be selected and the model parameters have to be estimated. Determination of the spatial dependence structure of the considered variables and the semivariogram model as a measure of this dependence are the base of geostatistics (Vieira *et al* 1983). Objective and minimum variance predictions can be made by geostatistical methods considering the positions (co-ordinates) of observation points and the correlation between observations. Universal kriging geostatistical method has various semivariogram models. These models are: circular, spherical, tetraspherical, pentaspherical, exponential, gaussian, rational quadratic, hole effect, K-bessel, J-bessel and stable. The selected model influences the prediction of the unknown values, particularly when the shape of the curve nearby the origin differs significantly. The steeper the curve nearby the origin, the more influence the closest neighbours will have on the prediction.

The suitable semivariogram model has to be used in the studies of multi-year observation and evaluation of the water-table. This study is an attempt to find out which of the semivariogram models, from those mentioned for universal kriging with linear drift, will give acceptable results in predicting the water-table values based on the monthly water-table observations for the year 2002 in Mustafakemalpasa (MKP) irrigation scheme.

2. Subject area and data

MKP left bank irrigation scheme is the biggest irrigation system in Marmara region (Turkey) which covers an area of 15,000 ha in the north–west Anatolia. MKP irrigation lies between 25°22′E longitude, 40°12′N latitude and 28°31′E longitude, 40°02′N latitude (figure 1). MKP irrigation project started operation in 1967. Main slope direction in MKP irrigation area is south to north and average slope is ranged from 0 to 1%. Soil characteristic in the project area is mainly young alluvial; 54.9% of which is clayey-loamy and clayey, 34.3% is loamy sand-loam and 10.8% is sandy clay, sandy (Anonymous 1967). There are mesozoik and tertiary aged formations in the study area. Jura aged limestones represent the mesozoik formation. These limestones are more fractured and fractures were generally filled by calcite crystals. Above this layer, tertiary aged formations were placed. Neogene aged formations represent the tertiary. Neogene formations give the mostra in the large area. Below the neogene is the konglomera. Konglomera consists of jura aged limestone gravels. Above the konglomera poor cemented sandstones were placed. White lake limestones, marl and clay series were partially placed in this layer. The upper level of tertiary formations belongs to miyosen and consists of gravel and sandstones. Gravel and sandstones are cross layered. Partially clay bands were placed between sand and gravel. These layers are generally impermeable. In kuaterner, the greater part of neogene land was covered with alluvion due to materials (sand, gravel, clay) transportation by river (Anonymous 1976). Groundwater aquifer is homogeneous and unconfined aquifer.

The study area has mediterranean climate characteristics. Therefore humid temperate climate is dominant in the area. Annual average minimum and maximum temperatures were 1.8°C and 30.9°C, respectively while annual average temperature was 14.2°C. Annual average relative humidity was 71.6%. Rainfall is observed in all seasons. Maximum rainfall has been observed in winter, while minimum rainfall in summer. Seasonal distribution of rainfall is greatly differed: 25.1% of total rainfall is observed in spring, 8.9% in summer, 25.9% in autumn and 40.1% in winter. Average annual rainfall is 690.2 mm.

The main water source of the MKP irrigation system is the Mustafakemalpasa River. The electrical conductivity of the river water ranges from 0.48 to 0.72 dS/m and the corresponding range in sodium absorption ratio is from 0.65 to 2.00. Water is diverted to the irrigation area by means of a dam which diverts water into two main canals. There are drainage canals parallel to the irrigation system; drainage water is removed from the project area through pumping stations. There are 163 groundwater observation wells in MKP left bank irrigation area (Anonymous 2000). Monthly observations of water-table values done by the State Hydraulic Works of Turkey (SHW) in 2002 were used in this study. ArcGIS 8.2 software was used in applying the universal kriging method and the estimation of empirical semivariogram models. The monitoring wells are all bore wells and water levels have been monitored using a graded tape that provides sound signals when it touches water in the well, with an accuracy of 2 mm. Water levels data was measured on the last day of every month in all wells in the minimum possible time. These monitoring wells are only used for measuring water levels.

3. Models

Universal kriging assumes the model

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s})$$

where, $Z(\mathbf{s})$ is the variable of interest, $\mu(\mathbf{s})$ is some deterministic function and $\varepsilon(\mathbf{s})$ is random variation (called microscale variation). In figure 2, the

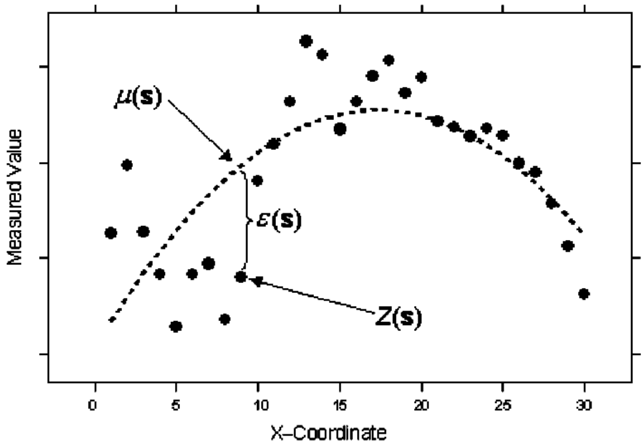

Figure 2. Universal kriging model.

observed data are given by the solid circles. The symbol $\mathbf{s}$ simply indicates the location (containing the spatial x- (longitude) and y- (latitude) co-ordinates). A second-order polynomial is the trend – long dashed line – it is $\mu(\mathbf{s})$. If the second-order polynomial is subtracted from the original data, the errors $\varepsilon(\mathbf{s})$ is obtained, which are assumed to be random. The mean of all $\varepsilon(\mathbf{s})$ is 0. Conceptually, the autocorrelation is modeled from the random errors $\varepsilon(\mathbf{s})$. Of course, a linear trend, a cubic polynomial, or any number of other functions can fit. In fact, universal kriging is a regression that is done with the spatial coordinates as the explanatory variables. However, instead of assuming that the errors $\varepsilon(\mathbf{s})$ are independent, they are modeled to be autocorrelated. There is no way to decide on the proper decomposition based on the data alone (Johnson *et al* 2001). Above-mentioned $\mu(\mathbf{s})$ is defined as drift in some references.

The drift is a simple polynomial function that models the average value of the scattered points. If the surface is not stationary, the kriging equations can be expanded so the drift can be estimated simultaneously, in effect removing the drift and achieving stationarity. Kriging will estimate the drift and residuals from the drift at every grid node so they may be mapped. The best estimate of the original surface is created by combining the surfaces representing the estimated drift and the estimated residuals. In theory, no other method of grid generation can produce better estimates (in the sense of being unbiased and having minimum error) in the form of a mapped surface than kriging. In practice, the effectiveness of the technique depends on the correct specification of several parameters that describe the semivariogram and the model of the drift. However, because kriging is robust, even with a naive selection of parameters the method will do no worse than conventional grid estimation procedures.

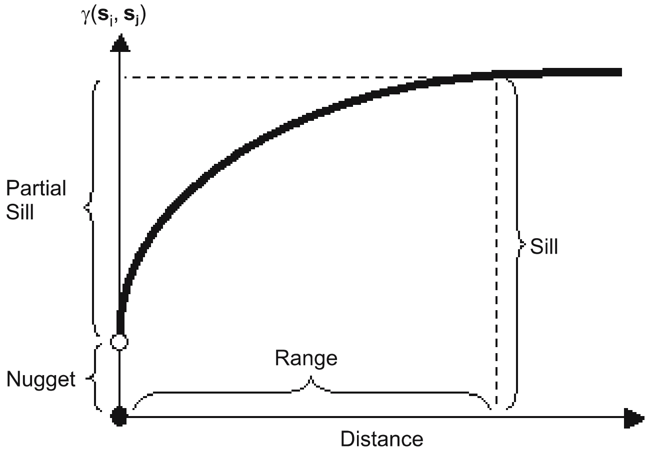

Figure 3. The anatomy of a typical semivariogram.

The two general classes of techniques for estimating a regular grid of points on a surface from scattered observations are methods called ‘global fit’ and ‘local fit’. As the name suggests, global-fit procedures calculate a single function describing a surface that covers the entire map area. The function is evaluated to obtain values at the grid nodes. In contrast, local-fit procedures estimate the surface at successive nodes in the grid using only a selection of the nearest data points.

Trend surface analysis is the most widely used global surface-fitting procedure. The mapped data are approximated by a polynomial expansion of the geographic coordinates of the control points, and the coefficients of the polynomial function are found by the method of least squares, ensuring that the sum of the squared deviations from the trend surface is at a minimum. Each original observation is considered to be the sum of a deterministic polynomial function of the geographic coordinates plus a random error (Anonymous 2001).

Like other kriging methods, various semivariogram models can be used with universal kriging. These models are the mathematical models which are fitted with the variation of the data set which consists of observations in distance dimension.

A semivariogram is defined as

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \mathrm{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))/2,$$

where var is the variance. Here, if two locations, $\mathbf{s}_i$ and $\mathbf{s}_j$, are close to each other in terms of the distance measure of $d(\mathbf{s}_i, \mathbf{s}_j)$, then we expect them to be similar, and so the difference in their values, $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$, will be small. As $\mathbf{s}_i$ and $\mathbf{s}_j$ get farther, they become less similar and so the difference in their values, $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$, becomes larger. This can be seen in figure 3, which shows the anatomy of a typical semivariogram.

Notice that the variance of the difference increases with distance, so the semivariogram can be thought of as a dissimilarity function. There are several terms that are often associated with this function, and they are also used in the geostatistical analysis. The height that the semivariogram reaches when it levels off is called the sill. It is often composed of two parts: a discontinuity at the origin, called the nugget effect, and the partial sill, which when added together gives the sill. The nugget effect can be further divided into measurement error and microscale variation. The nugget effect is simply the sum of measurement error and microscale variation and, since either component can be zero, the nugget effect can be comprised wholly of one or the other. The distance at which the semivariogram levels off to the sill is called the range (Johnson *et al* 2001).

Results of circular, spherical, tetraspherical, pentaspherical, exponential, gaussian, rational quadratic, hole effect, K-bessel, J-bessel and stable semivariogram models were evaluated in this study (Johnson *et al* 2001). A theoretical variogram is fitted, which is given least root mean square error (RMSE) value, for every semivariogram model by trial-and-error method.

RMSE can be used to compare the performance of several interpolation methods. RMSE is a kind of generalized standard deviation. It pops up whenever you look for differences between subgroups or for other effects or relationships between variables. It is the spread left over when you have accounted for any such relationships in your data, or (same thing) when you have fitted a statistical model to the data. Hence its other name, residual variation. RMSE is defined as the square root of an average squared difference between the observed and predicted values:

$$\mathrm{RMSE} = \sqrt{\frac{\mathrm{SSE}_i}{n}},$$

where SSE is sum of errors (observed − estimated values) and n is the number of pairs (errors). RMSE is frequently used in evaluating errors in GIS and mapping. It was tested whether the differences between the lowest RMSE value and the others are important. The method that yields the smallest value of RMSE is considered as the best fitted one.

4. Results and discussion

Universal kriging can be used efficiently with spatial data which have a normal distribution. Skewness values of the histograms of the spatial water-table data were used in order to check

Table 1. *Skewness states of monthly values in MKP, 2002.*

Month	Skewness: Without transformation	Skewness: With log-transformation
January	0.8900	−0.0800
February	0.9000	−0.0058
March	0.9100	−0.0300
April	0.9100	−0.0060
May	0.9100	−0.0222
June	0.9100	−0.0390
July	0.9100	−0.0420
August	0.9000	−0.0650
September	0.9000	−0.0700
October	0.8700	−0.1000
November	0.8800	−0.0730
December	0.9000	−0.0050

whether the available data match the normal distribution. Histograms show the frequency of monthly water-table values measured in 163 groundwater observation wells. Skewness values of the histograms regarding all months in 2002 are given in two columns in table 1. The skewness values in the first column were obtained without making any transformation on the water-table values. If these values are close to zero, this means there is no skewness, that is, it matches the normal distribution. In the table, skewness values in the first column are close to 1. In order to adjust the water-table values to the normal distribution, LOG transformation was made and the histogram was formed again. The *log transformation* is often used for data with a skewed distribution and a few very large values. These large values may be localized in the study area, and the *log transformation* will help to make the variances more constant and also normalize the data. For *log transformation*, the predictions are automatically back-transformed to the original values before a map is produced by ArcGIS software (Johnson *et al* 2001). Skewness values for the obtained histogram were listed in the second column (table 1).

Thus, it was concluded that using log transformation, the data match the normal distribution. Hence, log transformation of the data was made and their semivariograms were calculated. It was attempted to find the method which has the least error by comparing the predicted values in the semivariogram models with the real values. RMSE values obtained for each semivariogram model were given in table 2.

Estimation is performed with local polynomial interpolation in some semivariogram models, others with global polynomial interpolation.

Table 2. *RMSE values obtained in various semivariogram models, 2002.*

Models	RMSE Values: January	February	March	April	May	June	July	August	September	October	November	December
Circular	0.8461	0.826	0.8312	0.8424	0.8526	0.8488	0.849	0.8527	0.8604	0.8431	0.8459	0.8248
Spherical	0.8463	0.8261	0.8312	0.8424	0.8526	0.8485	0.8487	0.8524	0.8602	0.8432	0.8461	0.8250
Tetraspherical	0.8465	0.8262	0.8313	0.8425	0.8527	0.8485	0.8488	0.8524	0.8602	0.8433	0.8464	0.8252
Pentaspherical	0.8467	0.8264	0.8315	0.8427	0.8529	0.8488	0.8490	0.8526	0.8604	0.8435	0.8466	0.8255
Exponential	0.8721	0.8557	0.8560	0.8644	0.8714	0.8559	0.8573	0.8572	0.8699	0.8708	0.8745	0.8541
Gaussian	0.8921	0.8754	0.8847	0.8951	0.9038	0.9028	0.9024	0.9102	0.9125	0.8911	0.8919	0.8746
Rational quadratic	0.8415	0.8206	0.8271	0.8396	0.8480	0.8462	0.8462	0.8504	0.8573	0.8376	0.8405	0.8212
Hole Effect	0.8880	0.8694	0.8794	0.8898	0.8985	0.8966	0.8963	0.9042	0.9070	0.8869	0.8877	0.8697
K-bessel	0.8837	0.8663	0.8756	0.8861	0.8949	0.8940	0.8937	0.9013	0.9040	0.8824	0.8833	0.8658
J-bessel	0.8858	0.8632	0.8731	0.8835	0.8923	0.8907	0.8903	0.8980	0.9011	0.8807	0.8815	0.8635
Stable	0.8901	0.8732	0.8825	0.8930	0.9017	0.9007	0.9003	0.9080	0.9105	0.8891	0.8899	0.8725

Table 3. *Results of analysis of variance for RMSE.*

Source	DF	SS	MS	F
Models	10	0.0659	0.0065	50.96
Error	121	0.0156	0.0001	
Total	131	0.0815		

Table 4. *LSD test results of RMSE values of the semivariogram models (a, b, c and d mean within a column followed by the same letter do not differ significantly ($P < 0.05$)).*

Models	RMSE
Gaussian	0.89471 a
Stable	0.89262 ab
Hole effect	0.88945 ab
K-bessel	0.88592 ab
J-bessel	0.88364 b
Exponential	0.86327 c
Pentaspherical	0.84388 d
Tetraspherical	0.84366 d
Circular	0.84358 d
Spherical	0.84355 d
Rational quadratic	0.83968 d
LSD (0.05)	0.00913

Global polynomial interpolation was used for estimation in Gaussian, hole effect, K-bessel, J-bessel and stable semivariogram models while local polynomial interpolation was used in the other models.

As seen in table 2, the rational quadratic model has the least RMSE in all months. For this reason, it can be said that rational quadratic semivariogram model is the most suitable model for completing the missing data in water-table measurements and forming a water-table surface net. The differences between RMSE values of the models were found statistically important ($P < 0.01$) (table 3).

Groundwater level maps were created for every month without using kriging and the rational quadratic semivariogram method which has the lowest RMSE value (figure 4a and b). Both maps were generally similar, as it can be seen from figure 4. However, the groundwater level map created without the use of kriging had sharp lined curves because of intensive distribution of observation wells in the study area while the map created with kriging had more smooth lines.

According to the results of least significant difference (LSD) test given in table 4, the rational quadratic model had a lower average RMSE value (0,83968) than other models. But, according to results of LSD test, rational quadratic, spherical, circular, tetraspherical and pentaspherical semivariogram models are in the same group. The

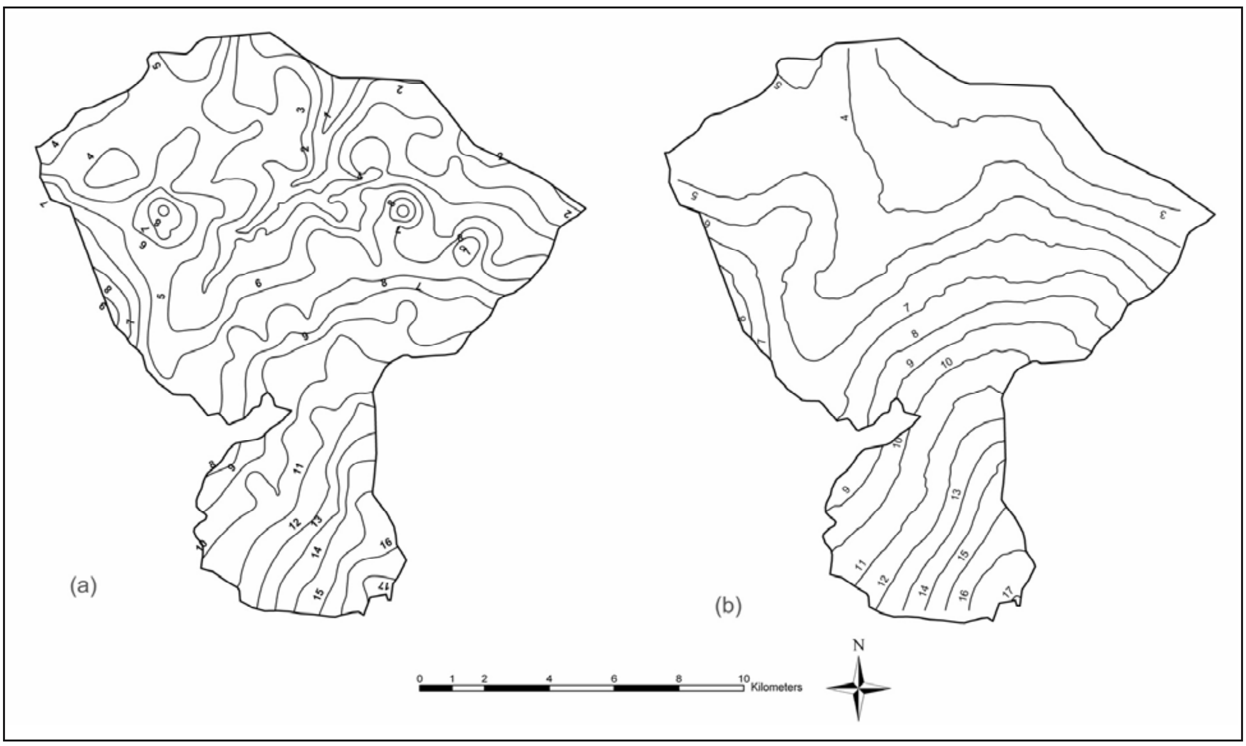

Figure 4. Water level (m) map above mean sea level for the month of January 2002, **(a)** map without kriging and **(b)** kriged map using rational quadratic semivariogram model.

gaussian semivariogram model had the highest RMSE value.

5. Conclusion

It can be concluded that the use of the rational quadratic model is the most suitable for the irrigation area. However, spherical, circular, tetraspherical and pentaspherical semivariogram models can give nearly the same water-table surface maps. Therefore water-table values that cannot be measured can be adequately described by these models which are required for the groundwater system planning and management for the region.

References

Anonymous 1967 Mustafakemalpasa Sulama Projesi Tesis Tanitma Foyu, Devlet Su Isleri, Karacabey Sube Mudurlugu, Bursa (Turkish).

Anonymous 1976 Mustafakemalpasa Projesi, MKP Ovasi Detayli Arazi Tasnif ve Drenaj Raporu, Cilt 1, Devlet Su Isleri Genel Mudurlugu 1. Bolge Mudurlugu, Bursa (Turkish).

Anonymous 2000 Mustafakemalpasa Sulamasi Taban suyu Raporu (Ekim 1999–Eylül 2000). Devlet Su Isleri Genel Mudurlugu 1. Bolge Mudurlugu, Bursa (Turkish).

Anonymous 2001 Surface III, Kansas Geology Survey at the University of Kansas, http://www.kgs.ku.edu/Tis/surf3/.

Bastin G, Lorent B, Duque C and Gevers M 1984 Optimal estimation of the average rainfall and optimal selection of raingage locations; *Water Resour. Res.* **20(4)** 463–470.

Chirlin G R and Dagan G 1980 Theoretical Head Variogram for Steady Flow in Statistically Homogeneous Aquifers; *Water Resour. Res.* **16(6)** 1001–1015.

Coram J, Dyson P and Evans R 2001 An Evaluation Framework for Dryland Salinity, report prepared for the National Land & Water Resources Audit (NLWRA), sponsored by the Bureau of Rural Sciences, National Heritage Trust, NLWRA and National Dryland Salinity Program, Australia.

Duc H, Shannon I and Azzi M 2000 Spatial distribution characteristics of some air pollutants in Sydney; *Mathematics and Computers in Simulation* **54(1–3)** 1–21.

Hill M and Alexandar F 1989 Statistical Methods Used in Assessing the Risk of Disease Near a Source of Possible Environmental Pollution: a Review. *J. Roy. Stat. Soc.* **152** 353–363.

Johnson K, Ver Hoef J M, Krivoruchko K and Lucas N 2001 Using ArcGIS Geostatistical Analyst, GIS by ESRI, Redlands, USA.

Kumar D and Ahmed S 2003 "Seasonal Behaviour of Spatial Variability of Groundwater Level in a Granitic Aquifer in Monsoon Climate"; *Curr. Sci.* **84(2)** 188–196

Vieira S R, Hatfield J L, Nielsen D R and Biggar J W 1983 "Geostatistical Theory and Application to Variability of Some Agronomical Properties", *Hilgardia* **51(2)** 1–75, Davis, California.

White J G, Welch R M and Norvell W A 1997 Soil Zinc Map of the USA Using Geostatistics and Geographic Information Systems; *Soil Science Society of America Journal* **61(1)** 185–194.

MS received 26 July 2006; revised 22 September 2006; accepted 26 September 2006