PREDICTION OF UNMEASURED PET DATA USING SPATIAL INTERPOLATION METHODS IN AGRICULTURAL REGION

Ju Young Lee[1] **and Ganeshi Krishinamurshy**[2]

[1,2] School of Environmental and Water Resource Engineering, Department of Civil Engineering, Texas A&M University, College Station, TX 77840

Abstract: This paper describes the use of spatial interpolation for estimating seasonal crop potential evapotranspiration (PET) and irrigation water requirement in unmeasured evaporation gage stations within Edwards Aquifer, Texas using GIS. The Edwards Aquifer area has insufficient data with short observed records and rare gage stations, then, the investigation of data for determining of irrigation water requirement is difficult. This research shows that spatial interpolation techniques can be used for creating more accurate PET data in unmeasured region, because PET data are important parameter to estimate irrigation water requirement. Recently, many researchers are investigating intensively these techniques based upon mathematical and statistical theories. Especially, three techniques have well been used: Inverse Distance Weighting (IDW), spline, and kriging (simple, ordinary and universal). In conclusion, the result of this study (Table 1) shows the kriging interpolation technique is found to be the best method for prediction of unmeasured PET in Edwards aquifer, Texas.

Keywords: Kriging, Edwards Aquifer, Potential Evapotranspiration, Spatial Interpolation

1. INTRODUCTION

Estimation of potential evapotranspiration (PET) is essential for role in the agricultural water resource planning and management using ArcGIS 8.3. Crop evapotranspiration (ETc) is calculated mathematically as:

$$ET_c = K_c \times PET \qquad (1)$$

where Kc is the crop coefficient having different value according to the genotype and its growing stage. However, Edwards aquifer area has insufficient PET gage station and data with short observed records. As figure 1 shown, two PET gage stations are just installed in Edwards aquifer area namely, Kinppa and Bexar PET gage stations. Moreover, its agricultural area has several geological variables as elevation which range from 200 m to 390 m. Thus, it is not appropriately to predict precisely PET data and use substitute from data of nearby known point PET gage station.

Therefore, various geostatistical interpolation techniques create predictable PET data from spatially sparse observational PET data. Three techniques are used for this study: Inverse Distance Weighting (IDW), spline, and kriging (simple, ordinary and universal). Also, this study shows which interpolation methods are more precise through making a comparing among them. Of these techniques, this study shows that spline and IDW are not suitable for prediction of creating PET data. Also, this result falls into Cressie's result which kriging (Cressie, 1991) is the best suitable technique. Moreover, Hosseini (2001) applied to the temperature and evaporation using Kriging (simple, ordinary and universal) interpolation. Hutchinson (1993) and Voltz and Goulard (1994) applied for climatic data and soil retention curve.

2. MATERIALS AND METHODS

2.1 Description of the Study Area

The Edwards aquifer is about 160 miles as measured length and range from 4 to 40 miles as width from Brackettville to Kyle and it consists of six following major counties: Kinney, Uvalde, Medina, Bexar, Comal, and Hays. Especially, the interesting counties within Edwards Aquifer are Kinney, Uvalde, Medina, and Bexar. The agricultural crop land distributed over Edwards Aquifer within 4 counties. Corn, Cotton, wheat, Oat, and Sorghum as major crops are harvested. This area belongs to semi-arid and semi-humid climate zone, thus, agricultural water use; particularly irrigation water use can be insufficient.

2.2 Database Development

The Texas Evapotranspiration Network provides the historical daily, monthly, and annual PET data at PET gage stations in Texas. Monthly PET data are used for this study. The 19 and 40 PET gage stations are used for potential evapotranspiration measured in Texas. But PET gage stations do not have sufficient in Edwards aquifer, Texas. More than 50% of PET gage stations have short observed data within 1 or 3 years. Most of all, there is just two PET gage stations.

2.3 Spatial Interpolation Methods

This study describes the application and comparing of three spatial interpolation techniques. Three methods such as IDW, spline and kriging (simple, ordinary and universal) are used for this study. Of these, Table 1 shows that kriging technique is more precise than IDW and spline. These results are similar to Delfiner and Delhomme's conclusion (1975) and Cressie's result (1991).

Kriging defines that it assigns all of weights from measured data to predict data at unmeasured data. Kriging method depends on mathematical and statistical aspects. Its basic concept is autocorrelation and distance. Spatial PET gage stations simulate their distances and their monthly PET data are computed as autocorrelation response to a function of distance. It is expressed statistically:

$$Z(s) = \mu(s) + \varepsilon(s) \tag{2}$$

where $Z(s)$ is data value, comprised of deterministic trend term $\mu(s)$ and error term ($\varepsilon(s)$). Error term decomposed into white noise and autocorrelated errors. "s" refers to location. The interpolated values which get through three different kriging interpolation techniques can be diverse according to trend term ($\mu(s)$). If μ is unknown, it is ordinary kriging technique based upon spatially correlated location. The quantification of spatially correlated locations is expressed statistically by semivariance:

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}(Z(s) - Z(s+h))^2 \qquad (3)$$

where h is lag according to distance; n is the number of PET gage station. Universal kriging technique assumes that μ is following a first order or second order trend.

$$\mu(s) = \alpha_1 x_i + \alpha_2 y_i \qquad (4)$$

$$\mu(s) = \alpha_1 x_i + \alpha_2 y_i + \alpha_3 x_i^2 + \alpha_4 y_i^2 \qquad (5)$$

Meanwhile, if μ is constant or known, it is simple kriging. The derivation of ordinary, simple and universal kriging is given in detail by Cressie (1993). The empirical semivariance values of PET data are fitted as semivariogram as all PET gage stations in Texas. Semivariogram model shows the spatially autocorrelated relationship among the PET gage stations. The general formula for fitting model is:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} W_i Z(s_i) \qquad (6)$$

where $\hat{Z}(s_0)$ is predict value; N is the number of measured PET data; W_i are the unknown weights for the associated value. Weights (W_i) are obtained from the modeled semivariogram values between all data in Texas PET gage stations and the fitted semivariance in unknown PET data area. Kitanidis (1997) shows how it is mathematically and statistically derived. However, literature review of other techniques as IDW and Spline is not referred in this paper.

2.4 Validation and Cross Validation Test

Validation and cross validation test provides how well interpolation method predicts PET data or which interpolation method better at unmeasured PET area. 30 observed PET stations distribute in Texas. The principle of cross validation is to compare predicted PET data to observed PET data. Cross validation omits data in one observed PET station, but it calculates data in 29 observed PET stations. This protocol is repeated. Cross validation is calculated statistically following formula (ESRI, 2001):

1) Mean prediction errors,

$$\frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - Z(s_i))}{n} \qquad (7)$$

2) Root mean square prediction errors,

$$\sqrt{\frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - Z(s_i))^2}{n}} \qquad (8)$$

3) Average kriging standard errors,

$$\sqrt{\frac{\sum_{i=1}^{n}\hat{\sigma}(s_i)}{n}} \qquad (9)$$

4) Mean standardized prediction errors,

$$\sqrt{\frac{\sum_{i=1}^{n}[(\hat{Z}(s_i) - Z(S_i)/\hat{\sigma}(s_i)]^2}{n}} \qquad (10)$$

For validation test, one of 30 PET observed stations, Kinppa gage station in Uvalde, Texas, omits to compare predicted data to observed data how well the model is precise. Most of all, the main reason that Kinppa PET station is omitted is due to be installed within Edwards aquifer region.

2.5 Methodology

Before creating predicting data in unmeasured area, the exploratory spatial data analysis is performed whether observed monthly PET data are necessary transformation for normal distribution to analysis or not. Methodology is histogram, normal QQ plot, or Skewness-Kurtosis analysis. If observed monthly PET data require Transformation, observed dataset are manipulated by using Box-Cox, arcsine, or log transformations. It shows in column 2 of Table 1.

In next step, detrending work is necessary to creating dataset. Then, the local trend can be analyzed through voronoi map or Thiessen Polygon through observed PET gage stations. It also can be computed to assign different weights to neighbors relative to unmeasured polygon. Also, before three interpolation methods are applied, trend analysis is essential to remove trend. It shows in column 3 and 4 of Table 1. In the geostatistics, the removing trend provides the modeling of random variation to obtain precise prediction.

Therefore, the spatial neighbors PET data analysis should be based on geological aspects. In general, the Edwards aquifer area has anisotropic formation, For example, its elevation ranges from 167 m to 592m. The elevation of northern area has higher than southern, and the western has higher than eastern. Understanding the spatial locations is important to get reasonable prediction.

3. RESULTS AND DISCUSSION

Geostatistical analysis of evapotransporation data has important role in the study how spatial interpolation methods are important for more precise prediction. Of these techniques, especially, kriging technique, it is comprised of 3 kinds of detailing methods. 3 kriging techniques are more suitable for this study namely, simple kriging, ordinary kriging, and universal kriging. However, the results by using spline and IDW have too much difference comparing to actual data observed. Table 1 shows results summarized using kring techniques (simple, ordinary and universal), spline, and IDW. Especially, column (circle) 12 and dark part indicate that it is the most suitable technique in the Table 1.

In the exploratory spatial data analysis, most of 30 observed monthly PET data are not following normal distribution, for example, Figure 2 shows that observed PET data for February need to transform to obtain precise result. Log transformation is used for February dataset. In the histogram analysis for February, skewness and kurtosis are -0.17072 and 1.8192. The negative skewness means that its distribution has long left tail in the small value, and kurtosis less than 3 means that its distribution has thin tail. Normality test is simulated for different months.

First of all, the ordinary kriging is chosen to choose good technique for February PET data prediction at the Edwards aquifer. To analyze the spatial correlation, Figure 3 shows spatially empirical semivarogram/covariance computed through all observed PET data and they are fitted by the spherical model.

Table 1. Several interpolation methods for Edwards aquifer region

Month	1	2	3	4	5	6	7	8	9	10	11	12
Feb	**Ordinary**	**Log**	**Local 100%**	**None**	**0.02**	**0.54**	**0.59**	**-0.02**	**1.01**	**2.60**	**2.538**	O
	Ordinary		Local 100%	1	0.02	0.54	0.54	-0.01	1.12	2.60	2.94	
	Simple	Log			0.03	0.56	0.60	0.05	0.94	2.60	2.46	
	Universal	Log	Local 100%	Const	-0.04	0.57	0.43	-0.12	1.41	2.60	1.88	
	IDW									2.60	1.40	
	Spline									2.60	1.70	
March	Ordinary	Log		None	0.00	0.68	0.67	-0.02	1.06	3.50	4.20	
	Simple	**Log**			**0.00**	**0.59**	**0.61**	**0.01**	**0.96**	**3.50**	**3.72**	O
	Universal	None	Local 100%	1st	-0.27	1.35	0.92	-0.12	1.15	3.50	4.19	
	IDW									3.50	3.12	
	Spline									3.50	4.30	
April	Ordinary	None		none	-0.21	4.69	4.74	-0.05	1.00	4.40	5.17	
	Simple	Log			-0.21	4.29	2.15	0.00	2.00	4.40	5.50	
	Simple	**Box Cox**			**0.00**	**4.29**	**4.29**	**0.00**	**1.00**	**4.40**	**5.10**	O
	IDW									4.40	3.30	
	Spline									4.40	5.72	
May	Ordinary	none	Local 100%	3rd	-0.19	1.48	0.98	-0.19	1.51	6.20	6.80	
	Simple	N Score			-0.12	1.29	1.21	-0.09	1.04	6.20	6.50	
	Universal	**Box-Cox**	**Local 100%**	**3rd**	**-0.23**	**2.41**	**1.57**	**0.00**	**1.52**	**6.20**	**6.28**	O
	IDW									6.20	5.42	
	Spline									6.20	6.72	
June	Universal	None	Local 100%	3rd	-0.13	2.53	1.67	0.04	1.55	7.00	7.12	
	Universal	**Log**	**Local 100%**	**3rd**	**0.00**	**2.74**	**2.21**	**-0.29**	**1.64**	**7.00**	**6.91**	O
	IDW									7.00	6.52	
	Spline									7.00	7.52	
July	Ordinary	Log	Local 100%	3rd	1.53	9.42	2.82	-0.52	2.95	6.00	6.63	
	Universal	**Log**	**Local 100%**	**1st**	**0.09**	**1.33**	**1.52**	**0.06**	**1.02**	**6.00**	**6.14**	O
	IDW									6.00	5.14	
	Spline									6.00	7.00	
Aug	Ordinary	None		None	0.03	1.23	1.49	0.02	0.85	7.60	7.85	
	Ordinary	**None**	**Local 100%**	**1st**	**0.02**	**1.29**	**1.21**	**0.02**	**1.07**	**7.60**	**7.64**	O
	IDW									7.60	7.10	
	Spline									7.60	8.10	
Sept	Universal	None	Local 100%	3rd	-0.26	1.72	1.21	-0.10	1.36	5.20	5.73	
	Universal	**Box-Cox**	**Local 100%**	**3rd**	**0.14**	**1.21**	**1.21**	**0.44**	**1.17**	**5.20**	**5.40**	O
	IDW									5.20	4.32	
	Spline									5.20	5.69	
Oct	Universal	Box-Cox	Local 100%	3rd	-0.26	1.49	1.15	-0.13	1.23	3.70	4.30	
	Universal	**Box-Cox**	**Local 100%**	**3rd**	**0.13**	**0.92**	**1.10**	**0.25**	**0.93**	**3.70**	**3.80**	O
	IDW									3.70	3.20	
	Spline									3.70	4.10	
Nov	Simple	Box cox			-0.14	0.52	1.52	-0.10	0.36	2.30	2.49	
	Universal	**none**	**Local 100%**	**3rd**	**-0.12**	**0.64**	**0.50**	**-0.19**	**1.25**	**2.30**	**2.15**	O
	IDW									2.30	1.75	
	Spline									2.30	3.20	
Dec	Universal	Box-Cox	Local 100%	3rd	-0.16	0.57	0.47	-0.39	1.28	2.07	2.06	
	Universal	**Box-Cox**	**Local 100%**	**3rd**	**0.05**	**0.54**	**0.71**	**0.06**	**1.07**	**2.07**	**2.08**	O
	IDW									2.07	2.00	
	Spline									2.07	2.20	

1: The applied interpolation methods
2: The applied transformations
3: Detrending works
4: Order of trend removal
5: Mean prediction errors
6: Root mean square prediction errors
7: Average kriging standard errors
8: Mean standardized prediction errors
9: Root mean square standized prediction errors
10: Measured PET value at Knippa station, Texas
11: Predicted PET value at Knippa station, Texas

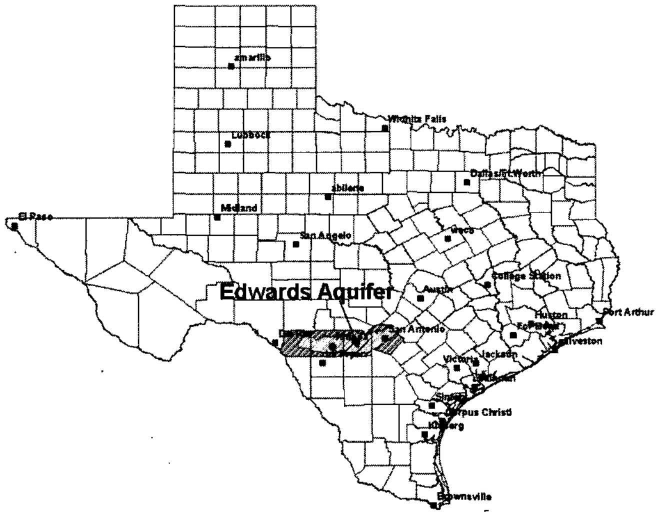

Figure 1. PET stations and crop land in Edwards aquifer region

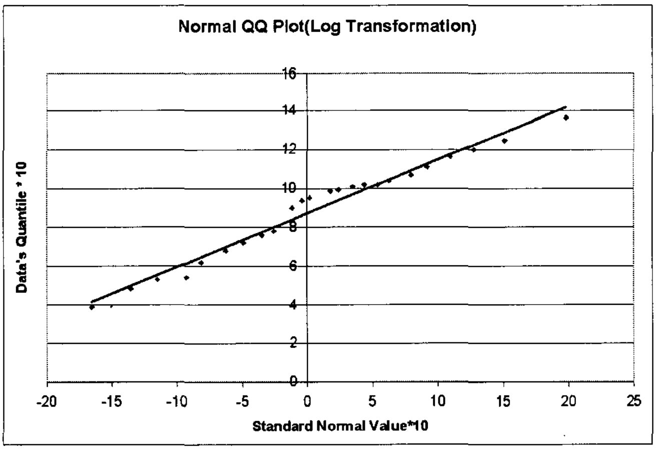

Figure 2. Normal QQ plot of PET for February

In the figure 3, major range is about 11.1 decimal degrees, and partial sill and nugget effect are 0.2632 and 0.21713 respectively. Range, sill and nugget effect provide information that semivariance value is 0.2632 when February data at all stations is spatially autocorrelated within approximately 700 miles (11.1 decimal degrees), but it is not suitable at unmeasured Edwards aquifer. Therefore, as shown Figure 4, considering of weights for eight measured PET stations is to obtain precise results including Edwards Aquifer region, but kinppa PET station is omitted for validation test. The range is about 150 miles (2.399 decimal degrees) including study area.

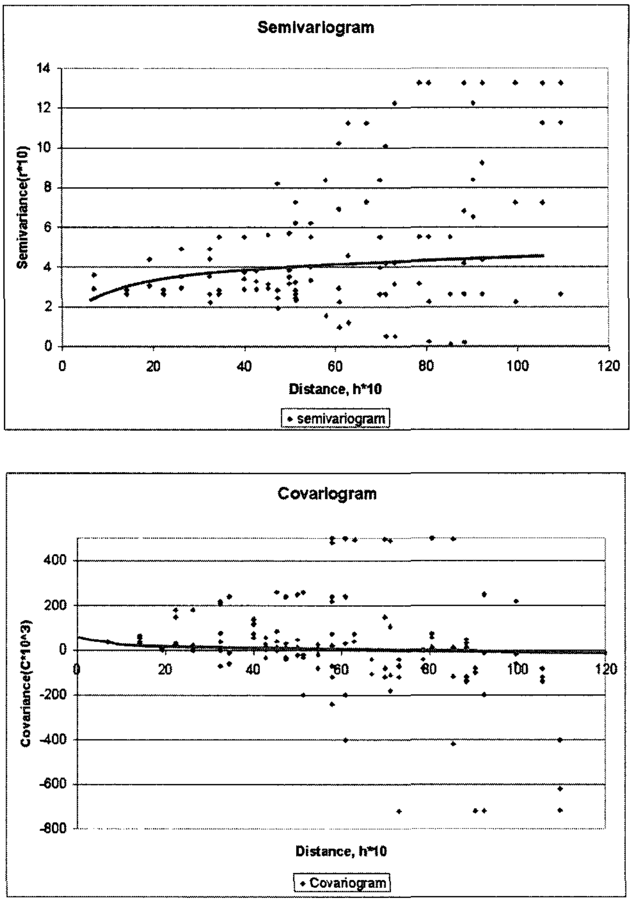

Figure 3. Empirical and fitted semivariogram/covariance for February

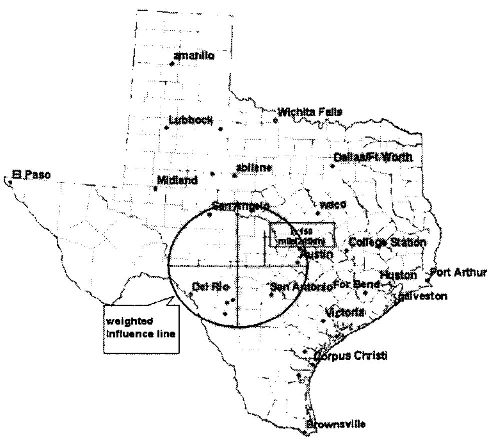

Figure 4. Determining of influence region

In the cross validation, as shown figure 5, there are five statistics values of prediction errors: the mean, the root mean square, the average standard error, the mean standardized, and the root mean square standardized. In figure 5, the mean is 0.0154. It is close to 0 (unbaisedness). The root-mean Square and average standard error are 0.5441 and 0.5907 respectively. They should be small. That means that the predictions are close to the measured values, and the root mean square standardized is close to 1(ESRI, 2001). That means that the standard errors are accurate between measured and predicted data. It shows in column 5, 6, 7, 8, and 9 of Table 1. In the validation test, observed and predicted value is 2.6 inch/Feb. and 2.54 inch /Feb. at Knippa PET station. Two results are similar.

Moreover, figure 5 shows that small values seem to be overpredicted and large values seem to be underpredicting on scatter chart, because kriging interpolates among extreme values. Therefore, ordinary kriging is the best method for February PET prediction. Table 1 shows that spatial kriging methods are used for this study are summarized.

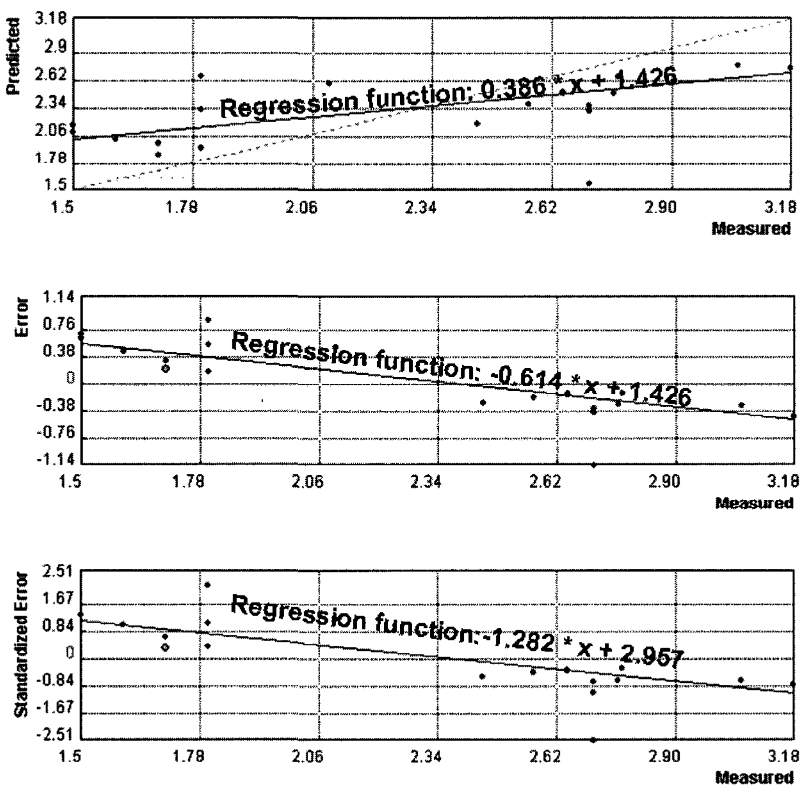

Figure 5. Cross validation for February using ordinary kriging

4. CONCLUSION

In this study, interpolation procedures were evaluated for precise prediction of potential evapotransiration (PET) data which is used for seasonal crop water requirement at unmeasured Edwards aquifer region, Texas. Moreover, these techniques provide direct prediction from measured PET data for irrigation water availability and water requirement. In this study, kriging interpolation methods are suitable to generate precise PET data in Edwards aquifer region but this study is not shown about results of IDW and spline. Kriging methods calculate autocorrelation of the measured data from PET stations and prediction values in unmeasured region from measured data. In summary, this study should be handled carefully following aspects:

First, exploratory spatial data are analyzed through several statistical methods. Mostly, biasedness between measured data and predicted data is due to insufficient exploratory spatial data analysis. Second, determining of measured neighbors has an important role. That means that PET data having the closest value is more weighted. In results, if all stations are considered to obtain prediction data in unmeasured Edwards aquifer region, the prediction data can be useless because geological variability and local trend can be negligent. Third, further, this study should take account for nugget effect because meteorological parameters can be affected by geological variability. Finally, validation and cross validation test would be much of help for this study to obtain precise results. Here, five sensitivity analyses are introduced. In table 1, kriging method is used and the best interpolation technique is chosen through five sensitivity test.

REFERENCE

Cressie, N., 1991. Statistics for spatial data, John Wiley and Sons, New York, N.Y.

Cressie, N., 1993. Statistics for Spatial Data, revised ed. John Wiley and Sons, New York. pp 900

Delfiner P. and Delhomme, J.O., 1975. Optimum interpolation by Kriging, John Wiley and Sons, New York. pp 96-104.

Huthcinson M.F., 1993. On the thin plate spline and kriging, Interface Foundation of North America, University of California, Berkeley. Vol.25: pp 104-113.

Kitanidis, P.K., 1997. Introduction to Geostatistics, Cambridge University Press, pp 41-109.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K., and Lucas, N., 2001. Using ArcGIS Geostatistical Analyst. ESRI.

Voltz, M. and Goulard, M., 1994. Spatial interpolation of soil moisture retention curves. Geoderma 62: 109-123.

INTERNET SOURCE

http://www.atmos.umd.edu/~owen/CHPI/IMAGES/EA-location.html

[http://www.geocomputation.org/2001/papers/mahdian.pdf

School of Environmental and Water Resource Engineering, Department of Civil Engineering, Texas A&M University, College Station, TX 77840

(E-mail: wayincrezki@hanmail.net)