# Applications of Spatial Interpolation for Climate Variables Based on Geostatistics: A Case Study in Gansu Province, China

WenzeYue, Jianhua Xu, Hongjuan Liao, and Lihua Xu

Department of Geography, East China Normal University, Shanghai, 200062, P. R. China

**Abstract**

Based on reviewing the origin, development and basic principles of Geostatistics, this article mainly introduces two interpolation methods: Ordinary Kriging and Cokriging. As an optimal one among so many methods to spatial interpolation for climate variables is not available, the article discusses Geostatistics-based interpolation methods by using 30-year average precipitation and evaporation data in Gansu province from 1961 to 1990. According to different semivariogram theory models, we adopt Ordinary Kriging and Bivariate Cokriging interpolation methods, and compare research results. We draw the following conclusions: (1) Both 30-year average precipitation and evaporation present obvious gradient change on space, in a great range. But the former's is larger than the latter's. 30-year average precipitation decreases gradually from southeast to northwest, however, evaporation increases gradually from southeast to northwest. (2) According to semivariogram cloud plots and experiment variance minimum principle, we select suitable theoretical semivariogram models based on Geostatistics interpolation, which can simulate the spatially continuous distribution patterns of the special regionalized variables in a better way. Compared with Ordinary Kriging, Cokriging considers the influence of altitude on precipitation and evaporation and thereby has higher interpolation accuracy. (3) Though Geostatistics methods can better reflect the general space patterns of climate variables, their interpolations precision is not high as we expect and can be improved further.

## I. INTRODUCTION

Traditional statistical methods are based on sample frequency distribution or average variance relation and corresponding determination criteria to determine sample distribution patterns and their relations. But they neglect spatial location and direction of the sample so as not to easily distinguish the differences between different spatial patterns. Geostatistics, as a method studying the spatial distribution patterns of features, considers not only the sample value, but also their spatial locations and distances between each other (Wang, 1999). Geostatistics theory was put forward and founded in 1962 by G. Matheron, a famous French mathematician, on the base of the work of South African geological engineer D· G· Krige and other achievements (Matheron, 1963). Generally, Geostatistics is accepted as a science to study the random and structural character or spatial autocorrelation and spatial heterogeneity of natural phenomena with the support of Regionalized Variables Theory and variograms (Webster, 1985; Wang and Guo, 1987; Wang and Hu, 1988; Issakes and Srivastuva, 1989; Hou and Huang, 1990). Therefore, Geostatistics is widely applied to many fields: structural character and randomicity of spatial data, spatial autocorrelation or dependence, spatial pattern and heterogeneity, optimal and unbiased estimate to simulate the discreteness or fluctuation of spatial data etc, for example: Mining, Geology, Agrology, Hydrology, Meteorology, Ecology, Environment and so on (Wang, 1999). An important purpose of Geostatistics is to interpolate in such locations without samples. In addition, Geostatistics includes many interpolation methods and some of them have already been applied in practical research (Wang and Guo, 1987; Chaolunbagen et al., 1994; Zhang et al., 1998).

Although Geostatistics is widely applied to interpolate with climate variables, data used in many previous cases such as precipitation or evaporation are single variables. This lowers interpolation precision as precipitation and evaporation variables are mostly subject to regional environment influences, for example, topography factor, especially altitude influence. Wu (2000), Zhang (1998), Thomas(2000)and other scholars have proved that there is close correlation between precipitation or evaporation and altitude. So this paper uses single variable such as annual average precipitation or annual average evaporation with elevation variable to interpolate and compares research findings by two different methods.

## II. PRINCIPLES OF SPATIAL INTERPOLATION METHODS FOR CLIMATE VARIABLES BASED ON GEOSTATISTICS

The spatial distributions of geography phenomena have intrinsic laws. Generally, there are two basic laws: Firstly, everything is related to everything else, but near things are more related than distant things (Tobler, 1979). Secondly, while causal factors form an essential part of the underlying pro-

1082-4006/03/0901~2-71$5.00

cess in a geographical system, the inclusion of the spatial components is likely to be important to an understanding of the problem analyzed (Cliff and Ord, 1973; Anselin and Getis, 1992). So we often regard the variables that have double attributes of random and structural character as regionalized ones. By definition, Geostatistics is a branch of statistics. It accounts for the spatial relationship of a regionalized variable. Considering the location, direction and distance of samples, Geostatistics directly calculates the correlation and dependence of spatial pattern and measures the spatial distribution and variation regular of regionalized variables (Hohn et al., 1993).

Climate features are typically regionalized variables because their spatial patterns have apparent structural characters due to the planetary wind belt and air pressure belt of large scale, and randomicity due to environment factors (terrain, physiognomy) and human activities in small scale. Moreover, the climate variables in long time scale used as important driving factors together with the climate variables value of every grid of the study area are needed to research environmental evolvement and the relation between human and nature. But such variable values can only be obtained through spatial interpolation by using limited climate data of the research region because meteorological stations are very limited and scatter randomly in space (Lin et al., 2002). Traditional interpolation methods of climate variables include inverse distance weighting interpolation, polynomial interpolation and radial basis function interpolation and so on. A common defect of these methods is that they can't explore the intrinsic essence of regionalized variables. So we discuss the method of regional climate variables interpolation based on Geostatistics to enhance interpolation precision.

## The measure for regionalized variable: Semivariograms

Firstly, regionalized variable is a stochastic variable and has local, random and exceptional attributes. Secondly, regionalized variable has generic or average structure character, namely there is some autocorrelation between Z (x) and Z (x+h). Here Z (x) and Z (x+h) are respectively observation values of points x and x+h, and h is the spatial distance from point x to x+h. Covariance and semivariograms, as two essential Geostatistics functions, are founded according to regionalized variables.

If regionalized variable Z (x), of which the expectation is m, satisfies the second order stationarity assumption and intrinsic assumption, then mathematically, its covariance $c(h)$ and semivariograms $\gamma(h)$ can be defined as the following (Matheron 1963; Hou and Huang 1990):

$$E[Z(x)] = m \tag{1}$$

$$c(h) = E[Z(x)Z(x+h)] - m^2 \tag{2}$$

$$\gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 \tag{3}$$

## Kriging interpolation based on Geostatistics

Both covariance and semivariograms can be used to estimate the values of a variable at unsampled locations. Kriging interpolation includes a series of methods such as Ordinary Kriging, Universal Kriging, Cokriging, etc. (Wang, 1999). Here we mainly introduce Ordinary Kriging and Cokriging.

### *(1) Ordinary Kriging*

Ordinary Kriging can be defined mathematically as follows (Xu, 2002):

$$Z_X^* = \sum_{i=1}^{n} \lambda_i Z(X_i) \tag{4}$$

where $Z_X^*$ is the estimated value, $\lambda_i$ is the corresponding weight of each observation $Z(X_i)$ on the estimation. These weights are calculated to ensure that the estimator is unbiased and the estimation variance is a minimum. The unbiased condition requires that:

$$\begin{cases} \sum_{i=1}^{n} \lambda_i C(X_i, X_j) - \mu = C(X_i, X^*) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{5}$$

where $C(X_i, X_j)$ is the covariance between sampled points i and j, $C(X_i, X^*)$ is the covariance between sampled point and estimated point, μis the Lagrange multiplier of minimum condition. Because the relation of covariance and semivariogram can be expressed as below:

$$\gamma(h) = c(0) - c(h) \tag{6}$$

Equations (5) can be defined by semivariogram:

$$\begin{cases} \sum_{i=1}^{n} \lambda_i \gamma(X_i, X_j) + \mu = \gamma(X_i, X^*) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{7}$$

### *(2) Cokriging*

Ordinary Kriging is applied to single regionalized variable but Cokriging applied to multiple regionalized variables. The latter is the developing result of the former and there are no essential differences between them. So we can develop the Cokriging method by using the same procedure as that is used in Ordinary Kriging method. Suppose there are K variables that make up coregionalization variable $Z_K(x)(K = 1,2,\cdots,K)$. If coregionalization variable $Z_K(x)$ satisfies second order stationarity assumption and intrinsic assumption, we can confirm that cross covariance and cross variogram are existent and they can be defined as follows:

$$E\{Z_K(x)\} = m_K \tag{8}$$

$$C_{KK'}(h) = E\{Z_K(x) \cdot Z_{K'}(x+h)\} - m_K m_{K'} \tag{9}$$

$$\gamma_{K'K}(h)=\frac{1}{2}E\{[Z_{K'}(x)-Z_{K'}(x+h)][Z_K(x)-Z_K(x+h)]\} \quad (10)$$

where $m_K$ is the expectation of $Z_K(x)$ and $C_{K'K}(h)$ is the cross covariance and $\gamma_{K'K}(h)$ is the cross variogram. Suppose $Z_X^*$ is determined by two coregionalization variables $Z(X_i)$ and $Y(X_j)$, then bivariables CoKriging Interpolation can be defined mathematically as given in Equation (11):

$$Z_X^* = \sum_{i=1}^{n} a_i Z(X_i) + \sum_{j=1}^{m} b_j Y(X_j) \quad (11)$$

where $a_i$ and $b_j$ are the corresponding weights of coregionalization variables $Z(X_i)$ and $Y(X_j)$. Equation (11) can be expressed by linear equations as follows:

$$\begin{cases} \sum_{i=1}^{n} a_i C(Z_i, Z_j) + \sum_{i=1}^{m} b_i C(Y_i, Z_j) + \mu_1 = C(Z^*, Z_j) \\ \sum_{i=1}^{n} a_i C(Z_i, Y_j) + \sum_{i=1}^{m} b_i C(Y_i, Y_j) + \mu_2 = C(Z^*, Y_j) \\ \sum_{i=1}^{n} a_i = 1 \\ \sum_{j=1}^{m} b_j = 0 \end{cases} \quad (12)$$

From equation (12), we can see Cokriging is more meaningful in practice to multiple variables.

The advantage of Ordinary kriging is to easily get resource data and to interpolate fast because this method only uses one single variable. However, low interpolation precision is also obvious as the method neglects influences of other factors. Compared with Ordinary kriging, Cokriging has higher precision and sampling efficiency and can get the variable value measured easily to measure variable values that are not easily attained by Ordinary Kriging. It also offers methods to solve spatial correlation and estimation between two variables, and plays an important role in spatial analyzation. However, the interpolation procedure is more difficult and should apply cross variogram under the hypothesis of minimum variance of the difference of two variables (Wang 1999).

## III. A CASE OF APPLICATION: THE SPATIAL INTERPOLATION OF MULTI-YEAR AVERAGE PRECIPITATION AND EVAPORATION IN GANSU PROVINCE

### Data

Data used in this article includes 30-year average precipitation and evaporation from 1961 to 1990 and altitudes of 53 meteorological observatories in some counties of Gansu Province. The spatial distribution of these meteorological observatories is shown in Figure 1. Firstly, we use Ordinary Kriging to interpolate the precipitation and evaporation data from observation stations. Secondly, we use Cokriging to interpolate precipitation and evaporation including altitude factor because Gansu Province lies in the transition area of three major plateaus (Qinghai-Tibet Plateau, loess plateau and the Inner Mongol plateau) in China and its terrain condition is very complicated.

### Semivariogram models

Choosing a semivariogram model is the precondition of spa-

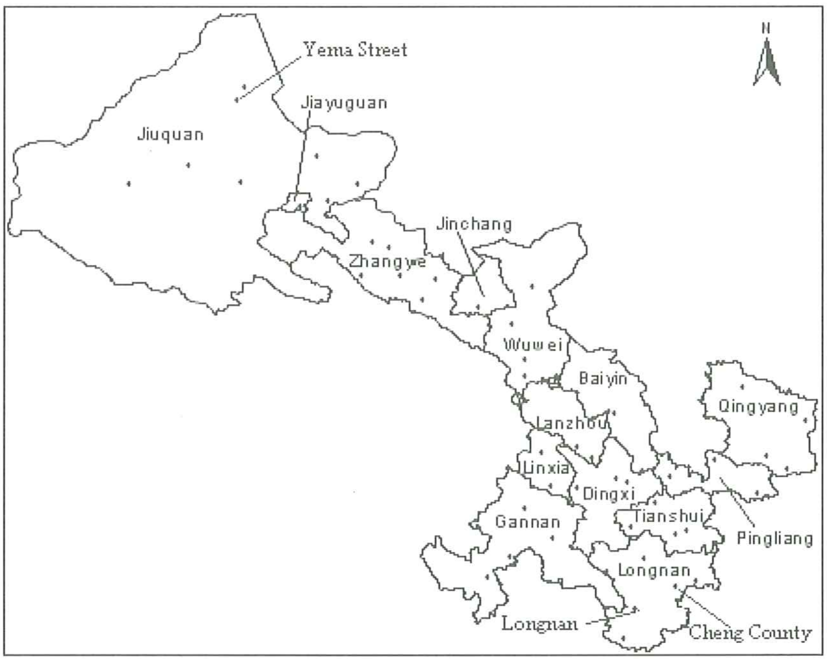


**Figure 1.** Spatial distribution of meterorological stations and the districts in Gansu province

tial interpolation, and semivariogram model determines the precision of spatial interpolation. In general, semivariogram model is chosen out from the existing theoretical models and based on the distributions of semivariogram cloud and the minimum estimation variance. Semivariogram cloud of 30-year average precipitation and evaporation from 1961 to 1990 in Gansu province is shown in Figure 2. In this figure, plot (1) represents precipitation semivariogram cloud and plot (2) represents evaporation semivariogram cloud.

From Figure 2 we can see that there are not obvious nuggets effects to 30-year average precipitation and evaporation because the sample value is observed at each meteorological observation station and thus the error of space resolution can be neglected. As the precipitation semivariogram cloud in plot (1) shows that the range is rather large, and the value of semivariogram increases slowly. When h is 6.2, the autocorrelation of precipitation fades away. In the semivariogram theory models, the exponential model is perfect to simulate the spatial pattern. Mathematically, exponential model can be expressed as follows:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C(1 - e^{-\frac{h}{a}}) & h < 0 \end{cases} \quad (13)$$

where, a is the range.

In plot (2), evaporation semivariogram cloud shows that the range is rather small. When h equals to about 3.9, the autocorrelation of evaporation does not exist. In the semivariogram theoretical models, the spherical model is defined as below:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c(\frac{3h}{2a} - \frac{h^3}{2a^3}) & 0 < h \le a \\ c_0 + c & h > a \end{cases} \quad (14)$$

## Spatial interpolation results

According to the semivariogram theory models of equations (13) and equations (14), we use Ordinary Kriging to interpolate 30-year average precipitation and evaporation in Gansu province from 1961-1990. Figures 3 and 4 show the interpolation results.

Ordinary Kriging interpolation only considers single factor autocorrelation but neglects the influence of other relative factors on the interpolation variables. In our study area, precipitation and evaporation change with altitude, so we should consider the factor of altitude. We use the method of Cokriging interpolation by adding the altitude of each meteorological observation station to interpolation. Figures 5 and 6 show spatial interpolation results.

## Discussion

We obtain the values of each grid of multi-year average precipitation and evaporation in Gansu province by Kriging spatial interpolation, these values are very important and they are the data source for other research (Lin et al., 2002). As the results show, both precipitation and evaporation have an obvious gradient change in space. Their change ranges are very wide and the precipitation's is much wider than the

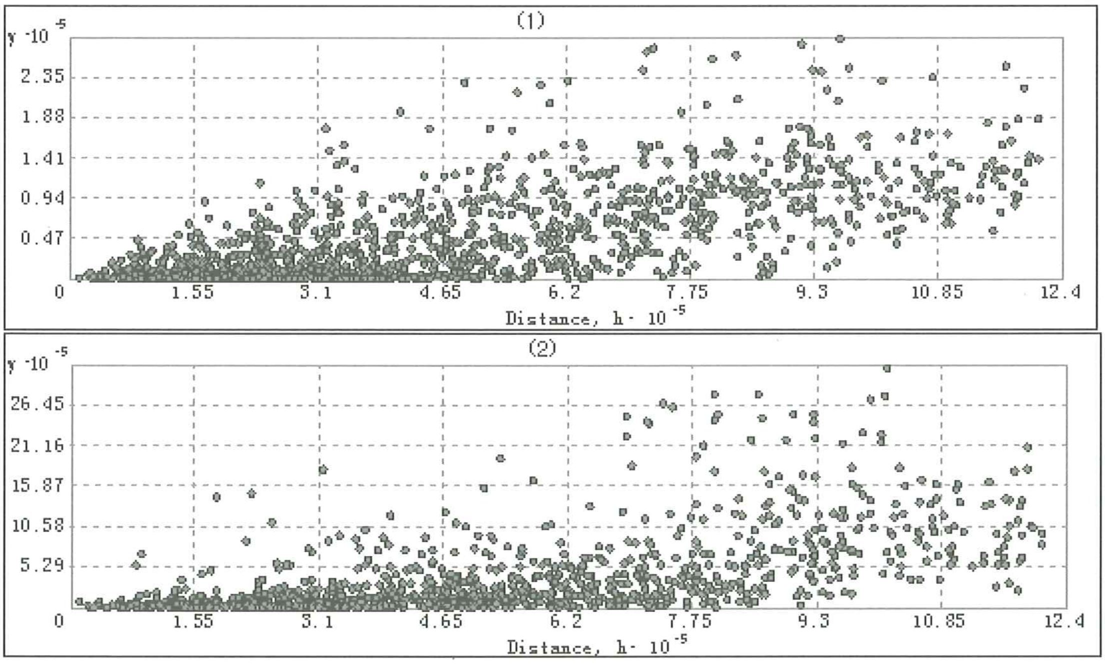


**Figure 2.** Semivariogram cloud of 30-year average precipitation and evaporation in Gansu province(1961-1990)

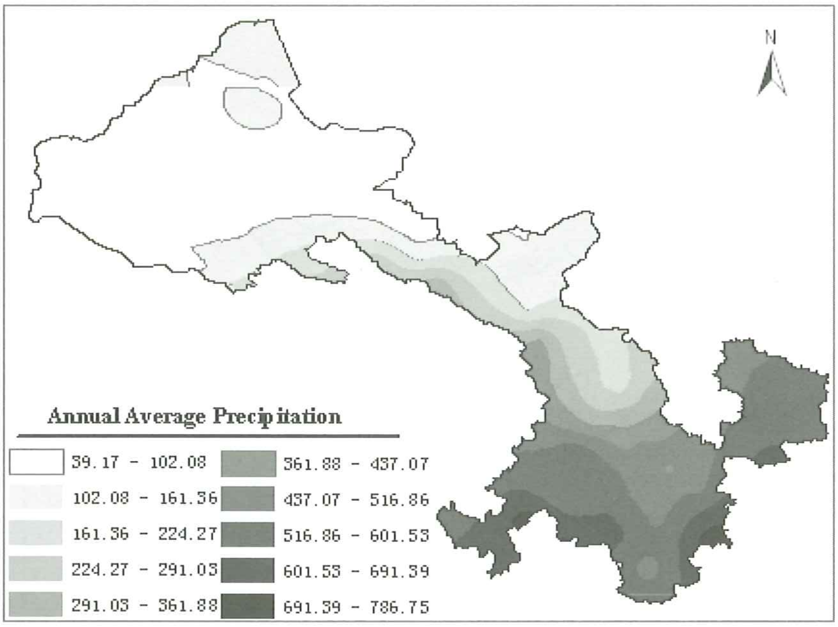


**Figure 3.** Spatial interpolation of annual average precipitation based on Ordinary Kriging

evaporation's. As to precipitation, it decreases gradually from southeast to northwest and evaporation increases from southeast to northwest.

Figure 3 shows that there is more precipitation in southeast than in northwest in Gansu province on the whole, and the gradient change is very evident; there is also more precipitation in mountains than in the plain from south to north. Precipitation decreases gradually from the south Qilian Mount to the north desert. The spatial difference of the precipitation is very distinct. Precipitation in the southeast is about 10 times as much as precipitation in the northwest. In the southeast of Gannan district (parts of Maqu country and Luqu country), Longnan district and Pingliang district (the southeast part of Lingtai country), precipitation is up to about 691.59-786.75mm in some years. In the northwest, precipitation decreases gradually. The 400mm contour of thirty-year average precipitation is located around Lanzhou; however, minimum contour around the almost whole Jiuquan district, Jiayuguan city, and part of Zhangye city located in the Northwest of Guansu province lies between 59.17mm and 102.08mm. From Figure 4, we can see that in Guansu province, the spatial pattern of 30-year

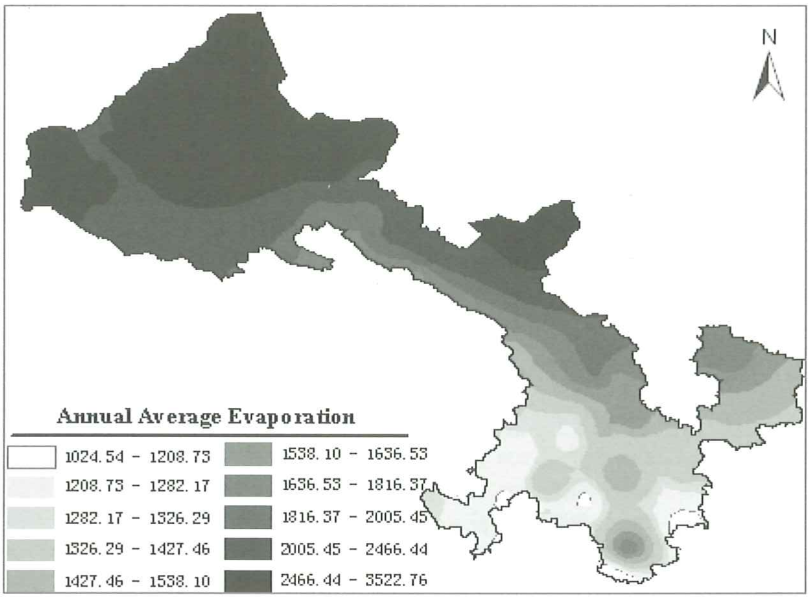


**Figure 4.** Spatial interpolation of annual average evaporation based on Ordinary Kriging

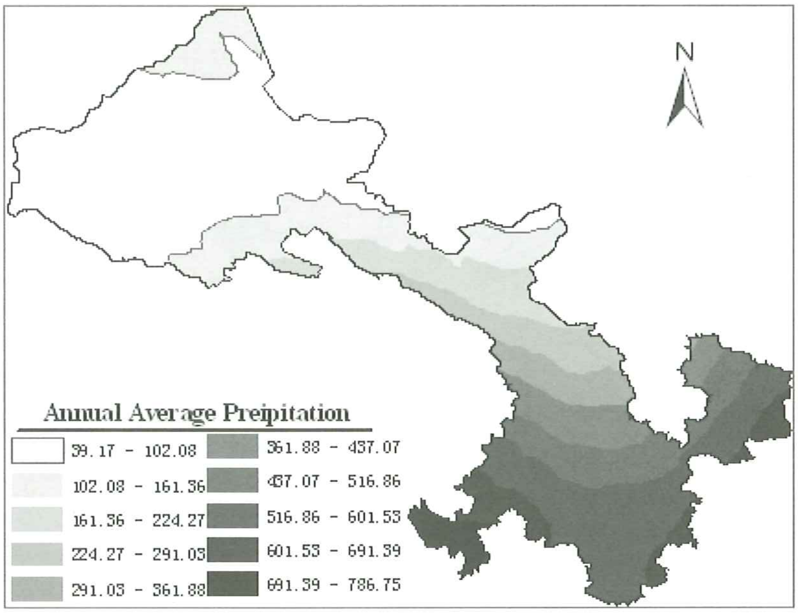


**Figure 5.** Spatial interpolation of annual average precipitation based on Cokriging

average evaporation is opposite to that of precipitation. 30-year average evaporation declines gradually from Northwest to Southeast and shows obvious gradient change feature. Though the change range of evaporation is smaller than that of precipitation, it is still very obvious. In almost all Jiuquan district and part of Tenggeli desert in the north of Minqin county, the 30-year average evaporation can reach 2931.30-3522.76mm, but in small part of Maqu county it only reaches 1024.54-1179.88mm. The 30-year average evaporation near to Lanzhou city is about 1389.77-1508.66mm. So the spatial patterns of 30-year average precipitation and evaporation obtained by interpolation are consistent with the fact of Gansu province.

However, altitude has a great influence on precipitation and evaporation. In Figures 3 and 4 the interpolation results may be not accurate in some regions because Ordinary Krging does not consider altitude factor. For example, as Figure 3 shows, Observation station in Cheng County in the south-east has the altitude of 970m and a lower precipitation, Longnan (has the altitude of 1079m) is similar, while Yema Street in the northwest with the altitude of 2159m has relatively more precipitation. As for evaporation, opposite to precipitation, in Figure 4 Cheng County has relatively more evaporation and

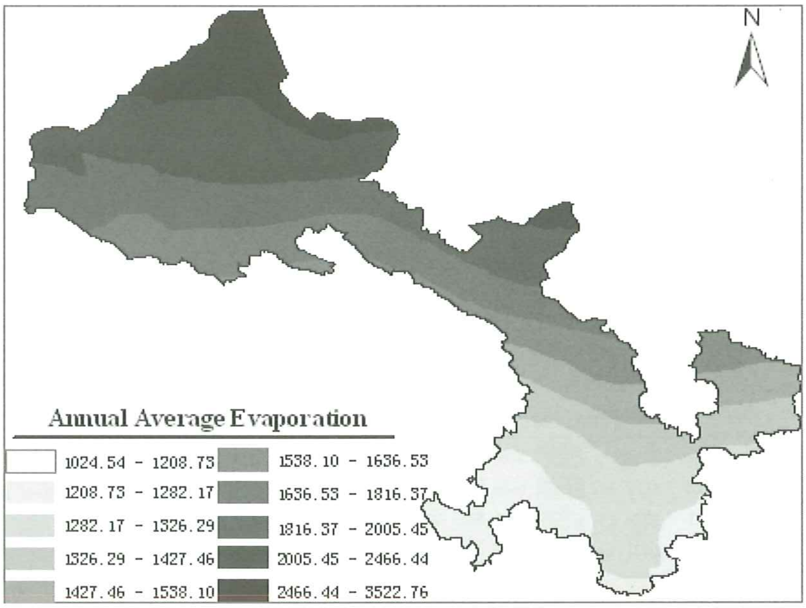


**Figure 6.** Spatial interpolation of annual average evaporation based on Cokriging

other districts also have the similar situation. Figures 5 and 6 show the results of Cokriging interpolation. On the general spatial patterns, Figures 5 and 3 are close, so are Figures 6 and 4. But the result obtained by the method of Cokriging interpolation is more reasonable because it considers altitude factor and reduces the local interpolation errors. As Figure 5 shows, the obviously low values of Cheng County meteorological observation station is corrected. In Figure 6, the high values of the same region are also rectified. However, the differences in interpolation results as showed in Figures 6 and 4 are not so apparent, which means altitude has a less influence on evaporation than on precipitation. The higher altitude is, the more obvious this influence is. Adding altitude to interpolation not only eliminates the spatial errors due to the altitude difference of meteorological stations, but also is more accurate. Comparing the results of Ordinary interpolation and Cokriging interpolation, we find the latter improves the precision and accuracy, which are shown in Figures 5 and 6.

Though the results of the above two methods can show the spatial patterns of climate features on the whole, interpolation results are not precise as we expect because 53 observations selected from Guansu province with an area of more than 440,000 $km^2$ are very limited and don't distribute uniformly in the research unit. What is more, in mountains, difference in altitude between adjacent two sampling points is more than 1,000m. Another problem is that observation equipment precision and technique of observers also have an influence on interpolation precision. So to make higher interpolation precision in some region, especially in small region like Jiuquan district, needs considering uniform and reasonable sample data except for perfect interpolation methods.

## IV.SUMMARY

(1) Geostatistics interpolation methods, which are based on semivariogram cloud and principle of the minimal variance, can simulate very well the spatially continuous distribution patterns of regionalized variables by choosing suitable semivariogram function models. Kriging interpolation fully considers the characteristics of regionalized variable and thereby greatly improves the interpolation precision. In addition, compare Ordinary Kriging with Cokriging, we find the latter has comparatively higher interpolation precision.

(2) Geostatistics interpolation can perfectly show the spatial distribution patterns of climate variables, but the precision of the above two methods is not so high as we need because of theoretic models, research region and data (broad region, complicated terrain, irregular shape, limited samples, uniform distribution and difference in observation precision). So we should pay more attention to the characters of the study area as well as the models when we check the results of Geostatistics interpolation. In addition, we can improve the interpolation precision by increasing samples, advancing observation accuracy and rectifying semivariogram theory models.

## ACKNOWLEDGMENTS

This paper is partially supported by Grant No. 40171069 from the National Natural Science Foundation, China. We want to thank the anonymous reviewer for the useful comments that helped to improve the quality of this paper. We are also grateful to Lu Y. for her assistance in English editing. And the work of the editor and the editorial staff is very much appreciated.

## REFERENCES

[1] Anselin, L, Getis, A., 1992, Spatial statistical Analysis and Geographic Information Systems[J]. *Annals of Regional Science*, 26(1): 19~25.

[2] Chaolunbagen, Hetai, Liu, T. X., 1994, A Study of Spatial Variability on Aquifer Hydraulic Conductivity K. *Acta Geologica Sinica*, 68(4): 358~367 [in Chinese].

[3] Cliff, A. D. and Ord, J. K., 1973, Spatial Autocorrelation[M]. London: Point Ltd.

[4] Hohn, M E, Liebhold, A. M, Gribko, L. S., 1993, Geostatistical Model for Forecasting Spatial Dynamics of Defoliation Caused by Gypsy Moth (Lepidoptera: Lymantriidae). *Environ Entomol*, 22(5):1066-1075.

[5] Hou, J R, Huang, J. X., 1990, *The Theory and Methods of Geostatistics*, Beijing: Geology Press [in Chinese].

[6] Issaks, E. H, Srivastava, R.M., 1989, *An Introduction to Applied Geostatistics*. Oxford Univ. Press.New York.

[7] Lin, Z .H, Mo, X. G, Li, H. X. et al., 2002, Comparison of Three Spatial Interpolation Methods for Climate Variables in China. *Acta Geogra Sin,* 57(1): 47-55 [in Chinese].

[8] Matheron, G., 1963, Principles of Geostatistics. *Economic Geoogy*. 58:1246—1266.

[9] Thomas , Axel, 2002, Seasonal and Spatial Variation of Evapotranspiration in the Mountains of Southwest China. *Journal of Mountain Science*, 20 (4):385-393.

[10] Tobler, W., 1979, Cellular geography. In *Philosophy in Geography*, ed. S. Gale and G. Olsson. Dordrecht: Reidel.

[11] Wang, G. D., Guo, C. L., 1987, Application of the Kriging Technique in Geography. *Acta Geogra Sin*, 42(4): 366-375 [in Chinese].

[12] Wang, R D., Hu, D. G., 1988, *Linear Geostatistics*. Beijing: Geology Press [in Chinese].

[13] Wang, Z. Q., 1999, *Geostatistics Applied in Ecology*. Beijing: Science Press. 109-126 [in Chinese].

[14] Webster, R., 1985, Quantitative Spatial Analysis of Soil in the Field. *Advanced in Soil Science*. 3:1-70.

[15] Wei, F.Y, Cao, H.,X., 2002, *Geostatistics and Its Application to Meteorological Studies*. Weather.

[16] Wu, G. H., 2000, *Physical Geography*. Beijing: Higher Education Press [in Chinese].

[17] Xu, J. H., 2002, *Mathematical Methods in Contemporary Geography*. Beijing: Higher Education Press[in Chinese].

[18] Zhang, C. S., Zhang, S., He, B. J., 1998, Spatial Distribution Characteristics of Heavy Metals in the Sediments of Changjiang River System—Spatial Autocorrelation and Fractal Methods. *Acta Geogra Sin*, 53(1): 87-96 [in Chinese].

[19] Zhang, Y. G., 1998, Several Issues Concerning Vertical Climate of the HengDuan Mountains. *Recourses Science*, 20 (3):12-19 [in Chinese].