atial zoning of precipitation by geo-statistical methods and DEM as auxiliary variable in the Northeast of Iran

S.H.Sanaeinejad, N. Siyabi and M. Ghaemi
Department of water engineering, Ferdowsi University of Mashhad, Mashhad, Iran
Corresponding author's email: sanaein@gmail.com

tract

of the most important factors in environmental studies is precipitation and its spatial variability ipitation data is collected in some individual points in different stations, while these point based data not often completed in a long term period. In arid and semi-arid area data incompleteness is more us. There are different methods for reconstruction and zoning of precipitation data, however, it is ntial to choose a reliable method that could matches the regional requirements and also provides uate precision for our applications. Geographical position and topographical factors strongly ence spatial pattern of precipitation. Multivariate geo-statistical method is used to estimate the spatial lation of two variables that are interdependent in a physical sense. In this paper Ordinary Kriging as a univariate method and Ordinary CoKriging (OCK) as multivariate method were used for polation precipitation. The purpose of this study was to show how the use of a digital elevation model nprove interpolation processes for mapping the mean annual and monthly precipitation amount in the East of Iran. As most of the study area is mountainous, Digital Elevation Model (DEM) was used as iary variable in the models. Cross validation is used to compare the prediction performances of the tatistical interpolation algorithms. In this study Mean Squared Error (MSE) values were used for aring the models. Cross-validation suggested that OK, which ignores the information on elevation, iger prediction errors in comparison with prediction errors produced by OCK specially in months low precipitation. For example MSE values were 0.47, 0.3, 0.25 and 0.5 for March, October, mber and December respectively when OCK was applied, while these values were 0.17, 1.8, 1.2 and spectively for the same months when OK was applied. Overall, the smallest value of MSE was 0.25 vember when OCK was applied which had the highest correlation coefficient as well. It could be uded that using of DEM as auxiliary variable improves the results, because of high variability in raphy in the study area. The results seem to favor the multivariate geo-statistical method including ary information (related to elevation). We conclude that OCK is a very reliable and robust olation method because it can take into account several properties of the landscape; it should ore be applicable in other mountainous regions, especially where precipitation is an important rphological factor.

vords: Precipitation, Geostatistic, Kriging method

duction

studies have been directed to analyze spatial variations of meteorological data such as itation. To achieve more accuracy in estimation of spatial precipitation distribution, very networks of measuring instruments are required, which entails large installation and ional costs (Goovaerts, 2000). It is also necessary to estimate point rainfall at unrecorded ns from values at surrounding sites. Different methods have been developed and applied in eld of study. Mathematical formulas and sample point values are used in these methods to te unmeasured values across the given surface. For the interpolation of unknown values a weight is assigned to each known value depending only on the distance between sample points and location of the interpolated point.

Spatial interpolation methods are believed to be the most suitable methods of interpolation for climatic data, including precipitation. Different Kriging methods have provided more accurate predictions in comparison of commonly used methods such as IDW (Goovaerts, 2000; Lloyd, 2005). In Kriging methods the sparsely sampled variables can be complemented by secondary attributes that are more densely sampled. Topography and weather-radar observations could be used as secondary information in these models. The ability of model for interpolation increases significantly if co-variables are used (Daly et al. 1994). A regression analysis between missing precipitation data and elevation was performed. Hevesi et al. (1992 a,b) and Goovaerts (1999) incorporated elevation into the analysis to estimate precipitation by CoKriging method. Cokriging is a multivariate geo-statistical method that is used to estimate the spatial correlation of two variables that are interdependent in a physical sense.

In this study, we used geo-statistical algorithms (Ordinary Kriging and Ordinary Co-Kriging.) along with monthly and annual precipitation data for interpolation and generating the corresponding maps. Among these methods, OK use only precipitation data recorded at the meteorological stations and in OCK algorithm precipitation data is combined with auxiliary data (elevation). The results of methods were compared by using cross validation. The higher the correlation coefficient between precipitation data and elevation, the better results obtained from the models.

Material and methods

The study area is Khorasan province (Northeast of Iran, longitude 55∘W to 61∘E and latitude 38∘S to 30∘N). The area is approximately 248,000 km2 in a semiarid climate. The monthly and annual precipitation has been averaged for the climate normal period of 1993 – 2009. We were very strict in data selection, only keeping weather stations with complete years. After assuring the raw data quality, monthly and annual precipitation averages were calculated.

The second source of information is the station elevation above sea level (in meters), estimated from a DEM with 1 km by 1 km resolution.

Interpolation methods

In this method, the result is calculated from the weighted average of the known data in a way that the deviation of the results would be minimal. This can be determined with the help of the so called variogram (semi-variogram) function which gives half of the square aggregate of the value differences gear to the distance between the points:

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(X_i + h) - Z(X_i)]^2 \qquad \textbf{(1)}$$

where γ *(h)* is the experimental semi-variance value at distance interval *h*, *N*(h) is the number of pairs within a given class of distance and direction, Xi and $Xi + h$ are sampling locations separated by a distance *h*, $Z(Xi)$ and $Z(Xi + h)$ are measured values of the variable *Z* at the corresponding locations. A model (theoretical variogram) should be fitted to the points, when an experimental variogram is defined (Cressie, 1985).

Ordinary Kriging (OK)

The ordinary Kriging estimator $\hat{Z}(x_0)$, of an unsampled site is a linear sum of weighted observations within a neighbourhood:

$$\hat{z}(x_0) = \sum_{j=1}^{n} \lambda_j . Z(x_j) \qquad (2)$$

re the weights λ_i are obtained by solving the OK system (Journel and Huijbregts, 1978).

inary CoKriging(OCK)

Cokriging technique is a modified version of the kriging technique. It is used to merge two or random variables. The primary variable is usually correlated to the secondary variables. It ld be noticed that cross-variogram models that is used in OCK may get smoother than gram models and improve the estimation (Journel and Huijbregts.1978). The cross variance is computed by equation (3).

$$\gamma_{zs}(h) = \frac{1}{2} E[(Z(x) - Z(x+h))(Y(x) - Y(x+h))] \qquad (3)$$

variogram of elevation has been computed based on the point measured data on the location synoptic weather stations. This is due to avoid possible inconsistencies in the subsequent ling of direct and cross-semivariograms (Goovaerts, 1997; Wackernagel, 1998).

ts

1 lists the descriptive statistics of the raw data including (mean, standard deviation, ium, maximum, median, skewness, kurtosis and correlation coefficient). The data was used precipitation analysis was described by classical descriptive statistics.

Table 1. General statistics for the monthly and annual precipitation data and also linear relation coefficient between precipitation and elevation in each month precipitation (mm)

od	Mean	Std.dev	Min	Max	Median	SK	KU	Correlation coefficient between P & E
	28.03	6.06	19.09	41.88	28.90	0.42	0.13	0.83
	29.37	7.38	14.08	43.72	30.61	-0.31	0.49	0.82
	40.93	13.49	17.58	63.91	43.80	-0.13	-0.61	0.84
	26.84	9.95	10.81	43.41	26.59	0.03	-0.98	0.83
	12.54	10.36	1.151	31.72	10.78	0.67	-0.90	0.83
	2.94	3.13	0.15	9.52	1.42	1	-0.32	0.78
	1.10	1.90	0	7.51	0.35	0.04	0.92	0.65
	0.51	1.15	0	4.81	0.064	0.56	1.1	0.64
	1.82	2.77	0.02	11.49	0.98	0.47	0.99	0.80
	4.35	3.51	0.82	12.68	2.96	0.12	0.009	0.83
	13.87	8.22	2.81	32.03	12.04	0.93	0.52	0.84
	20.73	6.41	10.44	38.98	19.92	0.11	0.52	0.83
l	183.02	62.45	88.06	297.91	182.54	0.22	-0.70	0.83

Standard deviation; MIN : Minimum value; MAX : Maximum value; SK : skewness; KU : Kurtosis; lation; E:Elevation

It was observed that monthly and annual mean and median values were similar and the coefficients of skewness were almost zero. Therefore it was concluded that the data can be assumed to have a normal distribution and there is no need to use any transformation function to fit them in a normal distribution.

Fluctuations of monthly and annual precipitations from one year to another will not be investigated because the subsequent analysis will be conducted on these averaged data. The correlation coefficient ranges from (0.64to 0.84) therefore it seems worth accounting for the complete secondary information into the mapping of rainfall. It can be seen that the correlation coefficients are high (%75) except for dry months. Table (2) shows linear correlation between monthly precipitation data. It can be seen that the average correlation between precipitation measured over two continuous months slightly decreases when the separation time increases. Experimental variograms were computed for monthly and annual data. Variograms and cross-variograms were calculated for the data sets and the parameters of variogram models for each data set were chosen through cross validation. Figure 1 shows the cross semivariogram of annual precipitation and elevation.

Table 2. Matrix of linear correlation coefficients between monthly precipitation data

Period	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1.0000	0.8193	0.6883	0.2996	0.2283	0.2408	0.0231	0.0596	0.0085	0.2865	0.3117	0.8897
Feb		1.0000	0.9170	0.6781	0.5587	0.4988	0.2381	0.1667	0.2830	0.5402	0.5540	0.8316
Mar			1.0000	0.8229	0.7874	0.7166	0.3761	0.2437	0.4131	0.7060	0.7231	0.7859
Apr				1.0000	0.9309	0.8683	0.6747	0.5472	0.6961	0.8753	0.8943	0.5384
May					1.0000	0.8988	0.7458	0.5578	0.7415	0.8849	0.8949	0.4247
Jun						1.0000	0.7296	0.5872	0.7441	0.9697	0.8732	0.4938
Jul							1.0000	0.9102	0.9821	0.7785	0.8308	0.1129
Aug								1.0000	0.9507	0.6231	0.6736	0.0901
Sep									1.0000	0.7857	0.8150	0.1529
Oct										1.0000	0.9059	0.5118
Nov											1.0000	0.4912
Dec												1.0000

Semivariogram values increase with the separation distance as expected. Thus the squared difference between the data measured in two adjacent points is smaller than those which are further apart.

The semivariograms of precipitation data for two of the wet months are shown in Figure 2. The nugget effect was estimated by extending the variogram along the vertical axis. The behavior of nugget effect near the origin is very important because of its influence on the interpolation

:ulation. Decreasing nugget effect can be interpreted as closer distance between the measured nts. This is only possible when more precipitation data from new weather stations are ilable.

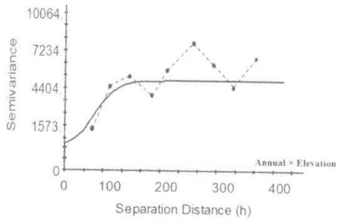


Figure 1. The cross semivariogram of elevation and annual precipitation

prediction at all un-sampled locations, by using univariate and multivariate interpolators, :h integrate the spatial correlation structures, is described with the variograms. After applying)K and OCK methods, the relevant maps obtained for annual and monthly precipitation data. re 3 shows the annual precipitation maps which illustrate the effect of each estimators and mpact of elevation on precipitation. The figure also shows that the map produced by using : method provides more details while the maps produced by applying OK. Reflect fewer ls which exhibit the needs for denser sampled data in the study area.

maps derived from multivariate interpolation methods, in which elevation is used as ancillary show better results and further details can be observed. Figure 4 shows some of the maps h produced by using OK method for April and May.

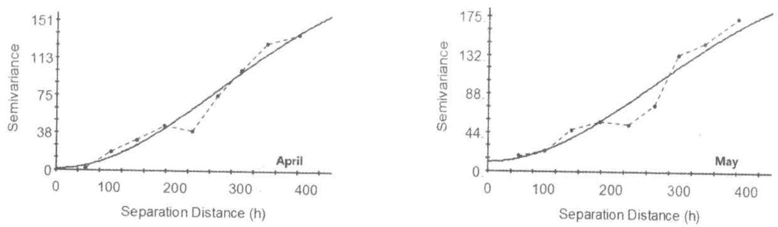


Figure 2. Experimental semivariograms of monthly precipitation

Interpolator performance

The interpolators were evaluated using cross-validation. In this method one precipitation data was removed at a time from the database and estimated from remaining data using different estimators. (e. g. Isaaks and Srivastava, 1989). Then the average square difference between the measured precipitations $Z(x_i)$ and its prediction $\widehat{z}(x_i)$, (MSE) was calculated and used as comparison criterion for all precipitation stations as following:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(z(x_i) - \widehat{z}(x_i)\right)^2 \quad \textbf{(4)}$$

In this study MSE was calculated for the precipitation data and used as comparison criterion. The value of this criterion should be close to zero, if the algorithm is accurate. For OK and COK methods the mean square errors of the prediction values was calculated using monthly (Jan–Dec) and annual (Ann) precipitation data. Figure 5 shows the mean square errors of prediction produced by OK and OCK methods.

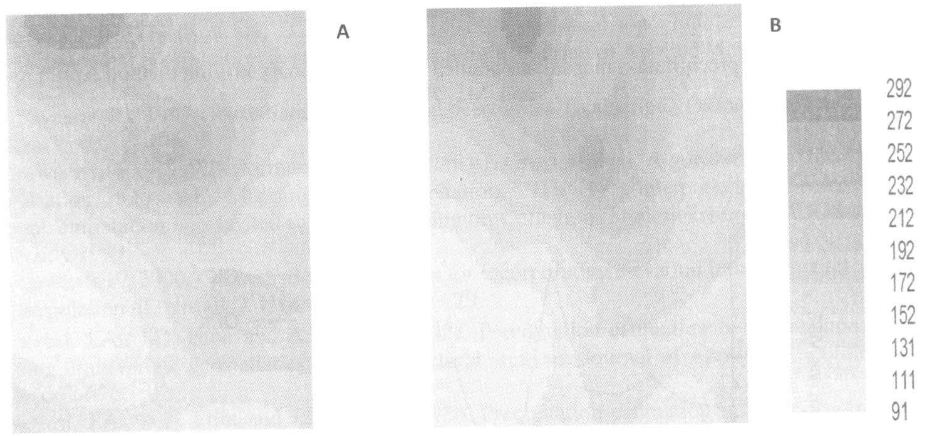
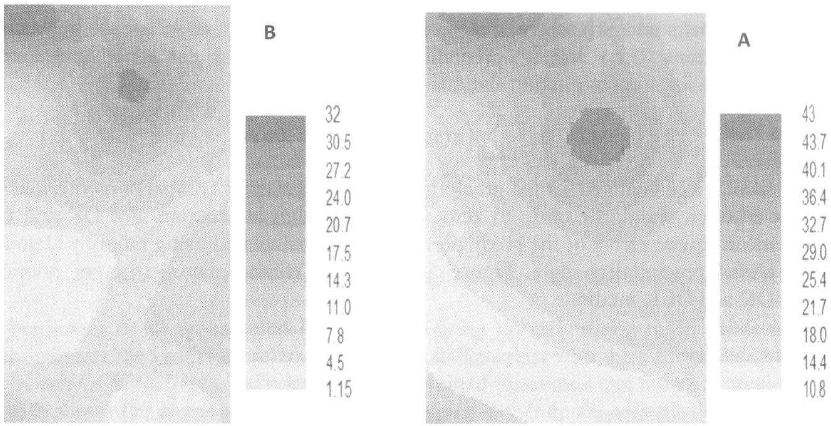
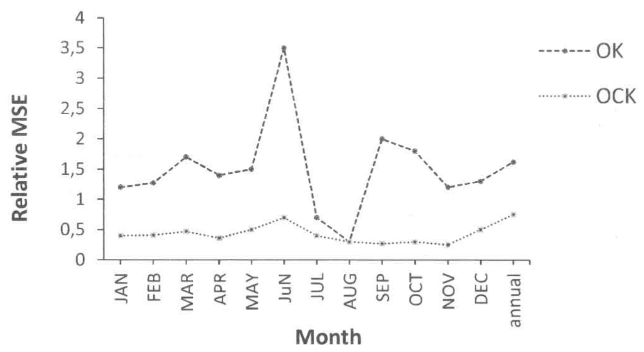


Figure 3. Annual precipitation maps which are produced by using OK (A), OCK (B) methods

As the charts show, maximum MSE for all months except August was for OK method and Maximum MSE (3.5) was derived for June. The lowest MSE (0.25) was also derived for November using the OCK method. The lowest MSE obtained by using COK for the months that have the highest correlation coefficients which is similar with the results obtained by Goovaerts (2000). In this study, when the correlation coefficient was low, i.e. in July and August, OCK and OK methods can be used similarly. Multivariate methods that use elevation as auxiliary data have etter performance than univariate methods; therefore OCK can provide better results in paring with OK method



gure 4. Monthly precipitation maps that obtained by using ordinary kriging method April (A) and May (B).



gure 5. Mean square error of prediction produced by the interpolation algorithms for monthly (Jan–Dec) and annual (Ann) precipitation.

Conclusions

Multivariate geostatistical methods perform better than others. Estimation can be improved if an ancillary data, such as elevation, is taken into account. Among the geostatistical methods that elevation is used as ancillary data, ordinary CoKriging produced slightly better results than the ordinary Kriging method. Cross validation showed that prediction performances can vary greatly among the algorithms. It was also concluded that ordinary Co-Kriging seems to give the most coherent results in this study. Without considering elevation larger estimation errors were obtained. This is more evident in this study as the area is mountainous. However the benefit of multivariate techniques can become marginal, if the correlation between precipitation and elevation is too small. Besides the correlation between elevation and precipitation, it is also important to look at their patterns of spatial continuity.

As the study area is mountainous, topography had a great influence on interpolation methods. It was concluded that if COK methods are used instead of OK, the estimation error is considerably reduced when correlation coefficient is higher than 0.78.

References

Cressie N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology **17**(5): 563–586.

Daly, C., R.P. Neilson and D.L. Phillips., 1994. A statistical topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology, 33 (2),pp. 140-158.

Goovaerts P., 1997.Geostatistics for Natural Resources Evaluation. Oxford University Press: New York.

Goovaerts, P., 1999.Performance Comparison of Geostatistical Algorithms for Incorporating Elevation into the Mapping of Precipitation. The IV International Conference on GeoComputation was hosted by Mary Washington College in Fredericksburg, VA, USA, on 25-28 July 1999.

Goovaerts P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol., 228, 113–129.

Hevesi, J.A., J.D. Istok and A.L. Flint., 1992a. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: Structural analysis. Journal of Applied Meteorology, 31, pp. 661-676.

Hevesi, J.A., A.L. Flint and J.D. Istok., 1992b. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: Isohyetal maps. Journal of Applied Meteorology, 31, pp. 677-688.

Isaaks, E. H., Srivastava, R. M. 1989.*An Introduction to Applied Geostatistics*. Oxford University Press: Oxford.

Journel, A.G. and Rossi, M.E., 1989. 'When do we need a trend model in kriging?',*Math. Geol.*, **21**(7), 715–739.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, New York.

Lloyd CD., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. Journal of Hydrology 308: 128–150.

Wackernagel, H., 1998. (completely revised). Multivariate Geostatistics. 2nd ed, Springer, Berlin.