Global Journal of Engineering Science and Research Management

# MULTIVARIATE GEOSTATISTICS FOR MAPPING METEOROLOGICAL PRECIPITATION

**Ismail Bulent Gundogdu**
Geomatics Engineering Department, Engineering Faculty, Selcuk University.
*Correspondence Author: bgundogdu@selcuk.edu.tr

**Keywords:** Multivariate geostatistics, linear regression, inverse square distance, ordinary co-kriging, precipitation map, water erosion.

## Abstract

Modelling of meteorological data is very important both for the evaluation of meteorological events and for the observation of its effects on the environment. Generally, classical interpolation methods are insufficient but multivariate geostatistical methods can be more effective, especially when studying secondary variables, because secondary variables might affect directly the model precision. In this study, the mean annual and mean monthly precipitation data from 264 meteorological stations in Turkey have been used for producing maps predicting meteorological precipitation. In addition to using linear regression (LR), the methods of inverse square distance (ISD) and ordinary co-kriging (OCK) were used and elevation, slope and aspect data for each point were added to the database as secondary variables. Cross-validation indicates that OCK yields the smallest prediction error. Standard errors verified that the best model could be produced with aspect as a secondary variable. Consequently, an aspect standard error map was produced to evaluate which points are more effective in the model. It is concluded that OCK is a very flexible method because it can account for several properties of the landscape. Therefore, it should be applicable in similar regions and a wider context, especially where precipitation is an important factor in water erosion.

## Introduction

The effects of climate on erosion are caused by precipitation, temperature and wind with precipitation the most effective factor. The type, severity, duration and regime of rainfall affect erosion significantly. In Turkey, soil is eroded annually at a rate of 600 tons/km2, compared with a global value of 142 tons/km2. Therefore, the precise mapping of precipitation amounts is the most important factor in prioritizing precautionary measures.

The purpose of the study is to research the environmental effects of meteorological precipitation and to determine the quantitative effects of elevation, slope and aspect on the described precipitation models.

This study investigates the importance of secondary variables and the identification of which of them are most effective for increasing the precision of modelling.

The available interpolation methods include geostatistics, e.g., kriging, inverse-distance weighting (IDW) and smoothing splines. Kriging has been used for the interpolation of temperature and precipitation regression residuals across a wide region, including the Mediterranean by Agnew and Palutikof (2000), the UK by Perry and Hollis (2005) and Iran by Alijani et al. (2008). Hold away (1996) and Erxleben et al. (2002) used residual kriging for the interpolation of temperature in a forest area and Goovaerts (2000) used co-kriging to incorporate elevation into the mapping of rainfall.

Ordinary kriging for the interpolation of monthly temperature anomalies has been used by almost all researchers; however, Brown and Comrie (2002) preferred to use inverse-distance weighting for precipitation.

A distance-weighted method has also been used by Perry and Hollis (2005), Attorre et al. (2007), and by Laguardia (2011) for interpolating monthly global climate anomalies.

The comparison of the surface interpolation methods has been the subject of many environmental and meteorological studies (Saito et al., 2000; Caeiro et al., 2003; Mouser et al., 2005; Milillo et al., 2008; Vienneau et al., 2009; Bajat et al., 2013). Generally, among the climate variables, rainfall attracts the greatest attention. According to Goovaerts (2000) and Burrough (2001), a geostatistical approach is more effective because it has the advantage of using the spatial context and external information as a random quantity to determine the most appropriate change of scale and possibly, to improve the predictions or simulations required.

Some recent studies have used topographical elevation as a source of secondary information for hydro meteorological variables (Pearce et al., 2009; Krakauer, 2012; Iaco 2013).

It is important to compare the statistical results obtained by using alternative methods and different secondary variables applied to the same dataset or area, because no single method or secondary variable(s) is optimal for all regions or for all seasons.

For estimating the long-year-average annual and monthly precipitation in Turkey, OCK has been compared with two alternative methods: linear regression (LR) and inverse square distance (ISD). It is concluded that OCK is a robust and flexible interpolation method because it can take into account auxiliary information in the form of smoothing the digital elevation map, in addition to secondary variable values.

Global Journal of Engineering Science and Research Management

## Study area and methods

Currently, terrestrial climate conditions are effective in the area. In severe winters, soil freeze to a depth of 50 cm. Because of the loosening soil following warmer days, precipitation becomes the most important factor in erosion. Mean annual precipitation is 720 mm and the maximum difference of elevation is 2500 m.

Precipitation records from 264 meteorological stations over a 30-year period have been obtained from the Turkish State Meteorological Service and input into a database using Microsoft Access to establish the Geographical Information Systems (GIS)-based application. In this process, in addition to the monthly and mean annual precipitation values, details of time, elevation, slope and aspect were recorded in the main dataset. A digital elevation map (250 × 250 m) has been used to obtain values of secondary variables. Figure 1 shows the study area and the network of stations.

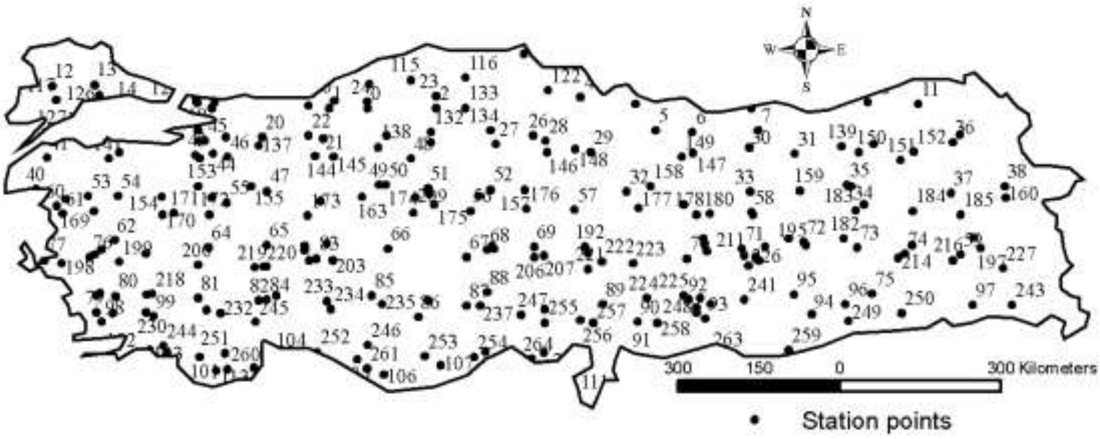


***Figure 1. Spatial distribution of the 264 meteorological stations.***

For estimating the 30-year-average annual and monthly precipitation in the area, data have been controlled by histogram values. Where a transformation has been necessary, the OCK surface has been compared with that of LR and ISD after controlled prediction errors.

## Methodology

### Linear regression

Diodato & Ceccarelli (2005) wrote a linear regression relation between elevation and precipitation in their paper as fallows;

$$z^*_{LR}(s_0) = a + bz_2(s_0)$$

(z) Is the precipitation value at the given grid node.

### Inverse square distance

$$z^*_{ISD}(s_0) = \frac{\sum_{i=1}^{n(s)} z_1(s_i) d_i^{-2}}{\sum_{i=1}^{n(s)} d_i^{-2}}$$

Where; $z_1(s_i)$ :the data points, $d_i$ : the distances between the station locations $s_i$ and the unknown point $s_0$ ..

### The multivariate geostatistical approach

$$Z_{OCK}(s_0) = \sum_{i=1}^{n} \lambda_i z_i(s_i) + \sum_{j=1}^{m} \lambda_j z_2(s_j)$$

Global Journal of Engineering Science and Research Management

(3)

Where; $z_i$ : a vector of the primary precipitation data selected in the $s_0$ neighborhood observation $s_i$,

$z_2$ **:** Secondary data selected in the $s_0$ which observed $s_j$,

$\lambda_i$ And $\lambda_j$ : weights associated with the distance between $s_0$ and $s_i$.

## Assessment of data

***Table 1. Statistical values to describe mean annual precipitation, elevation, aspect and slope***

| ***Statistics*** | ***Mean annual precipitation (cm)*** | ***Elevation (m)*** | ***Aspect*** | ***Slope*** |
|---|---|---|---|---|
| ***Min value*** | 2.05 | 1.62 | 1 | 2 |
| ***Max value*** | 15.27 | 2400 | 357 | 90 |
| ***Mean value*** | 6.07 | 707.69 | 195.87 | 79.32 |
| ***Std. dev.*** | 2.57 | 583.56 | 10.23 | 19.38 |
| ***Skewness*** | 1.02 | 0.908 | 0.975 | 1.07 |
| ***Kurtosis*** | 3.98 | 2.07 | 2 | 5.09 |
| ***1st quartile*** | 4.09 | 7.25 | 122.75 | 67.75 |
| ***Median value*** | 5.46 | 775 | 195 | 81 |
| ***3rd quartile*** | 7.53 | 1112.5 | 281.25 | 85 |

Table 1 shows the general statistical information of the data. These values must be controlled if these are normal distributions before geostatistical application. If the mean and median values are close, it can be said that the data have a normal distribution.

Table 2. Statistics of the experimental errors computed from mean annual precipitation data (LR: Linear regression, ISD: Inverse square distance, OCKE: Ordinary co-kriging with elevation, OCKA: with aspect, OCKS: with slope and OCKEAS: Ordinary co-kriging with elevation, aspect and slope)

| ***Statistics*** | ***Mean annual precipitation (cm)*** | ***Elevation (m)*** | ***Aspect*** | ***Slope*** |
|---|---|---|---|---|
| ***Min value*** | 2.05 | 1.62 | 1 | 2 |
| ***Max value*** | 15.27 | 2400 | 357 | 90 |
| ***Mean value*** | 6.07 | 707.69 | 195.87 | 79.32 |
| ***Std. dev.*** | 2.57 | 583.56 | 10.23 | 19.38 |
| ***Skewness*** | 1.02 | 0.908 | 0.975 | 1.07 |
| ***Kurtosis*** | 3.98 | 2.07 | 2 | 5.09 |
| ***1st quartile*** | 4.09 | 7.25 | 122.75 | 67.75 |
| ***Median value*** | 5.46 | 775 | 195 | 81 |
| ***3rd quartile*** | 7.53 | 1112.5 | 281.25 | 85 |

 ISSN: 2349-4506

Global Journal of Engineering Science and Research Management

Generally, the best model is the one that has the standardized mean nearest to zero, the smallest root-mean-square prediction error, the average standard error nearest the root-mean-square prediction error and the standardized root-mean-square prediction error nearest one Johnson et al. (2003). The best statistical results are in the OCKA line of Table 2.

Figure 2a–2c shows the variogram for precipitation variables, the variogram for aspect variables and the cross-covariance function of the precipitation variable and aspect, respectively. Ideally, the value of the semi-variogram for precipitation should be zero when the separation vector is zero. However, because measurement error exists, this is not true in this study. As shown in Figure 3e, annual precipitation is significant along the north and south borders of the area.

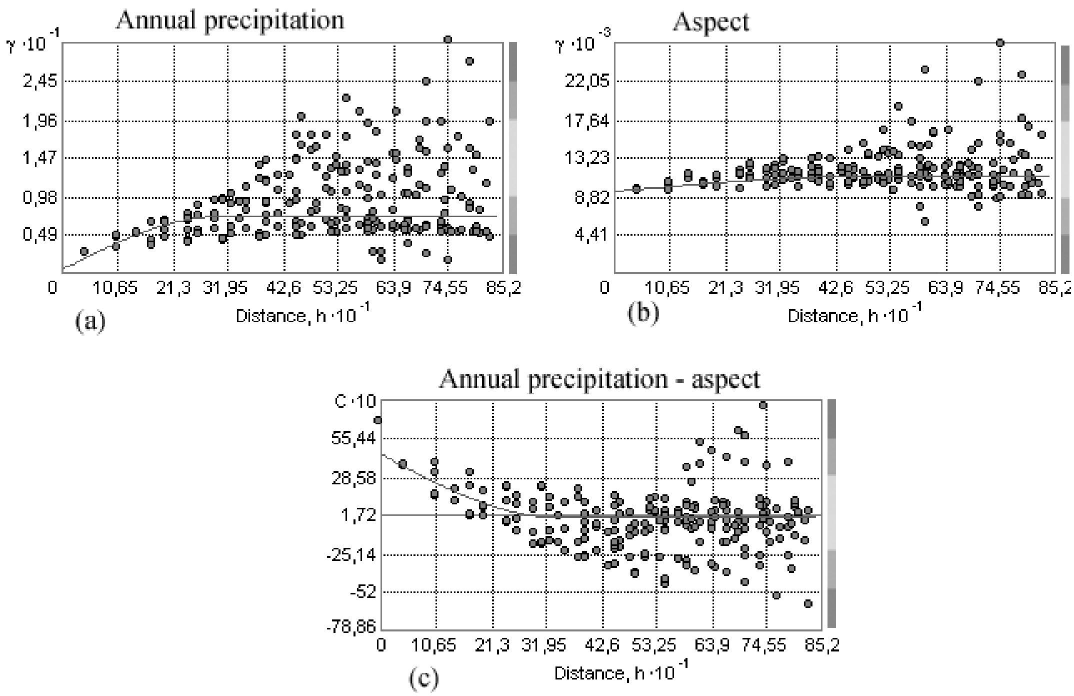


***Figure 2. Experimental semi-variance and cross-covariance function (dots) and their co-regionalization model (continuous line) for (a) annual precipitation, (b) aspect and (c) annual precipitation-aspect.***

In this phase, the absolute error values were re-estimated using seasonal precipitation (Table 3).

***Table 3: Mean absolute error prediction (cm) computed from seasonal precipitation data***

| Algorithms | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| **LR** | 2.170 | 2.190 | 2.160 | **1.960** |
| **ISD** | 2.164 | 2.173 | 2.133 | **1.977** |
| **OCK$_A$** | **1.122** | **1.149** | **1.127** | **1.062** |

The influence of the topographic condition is essentially sensitive on mean, root mean square and average standard errors. Root mean square is close to the average standard errors mean, correctly assessing the variability in the prediction (Table 2, OCKA method). There is clearly a significant improvement in the estimation performance when taking into account the aspect as a secondary variable through OCK; the mean absolute error decreases from 2.820–2.915 to 1.004 cm.

## Result maps and conclusions

The techniques were illustrated by using the annual and monthly precipitation observations collected at 264 climatic stations in Turkey. In the case study, cross-validation was used to compare the prediction performances of the three methods. OCKA gave the best performance in the statistical sense. Five types of map were produced for the result maps figure (Figure 3) to compare visually the precipitation distribution from the three methods. OCK was used with elevation OCKE, slope OCKS and OCKA aspect.

Global Journal of Engineering Science and Research Management

Following the decision that OCK is the best method for precipitation modelling, the secondary variables were used singly, doubly and as a trio (Elevation, aspect, slope, elevation-aspect, elevation-slope, aspect-slope and elevation-aspect-slope). However, prediction error values of OCKS and OCKA approximate each other and OCKA yields the best performance. Cross-validation indicates that the inverse distance interpolation, which ignores the information on aspect, yields the largest prediction errors. When comparing the relative performance of the classical algorithm, the LR or ISD techniques might struggle to improve the final maps. The aspect standard error map was produced from the prediction standard errors at each station point. It is indicates those regions for which priority should be given regarding taking precautions against water erosion. Indeed, the spatial distribution of station points and their number is also effective in the performance of the precipitation models.

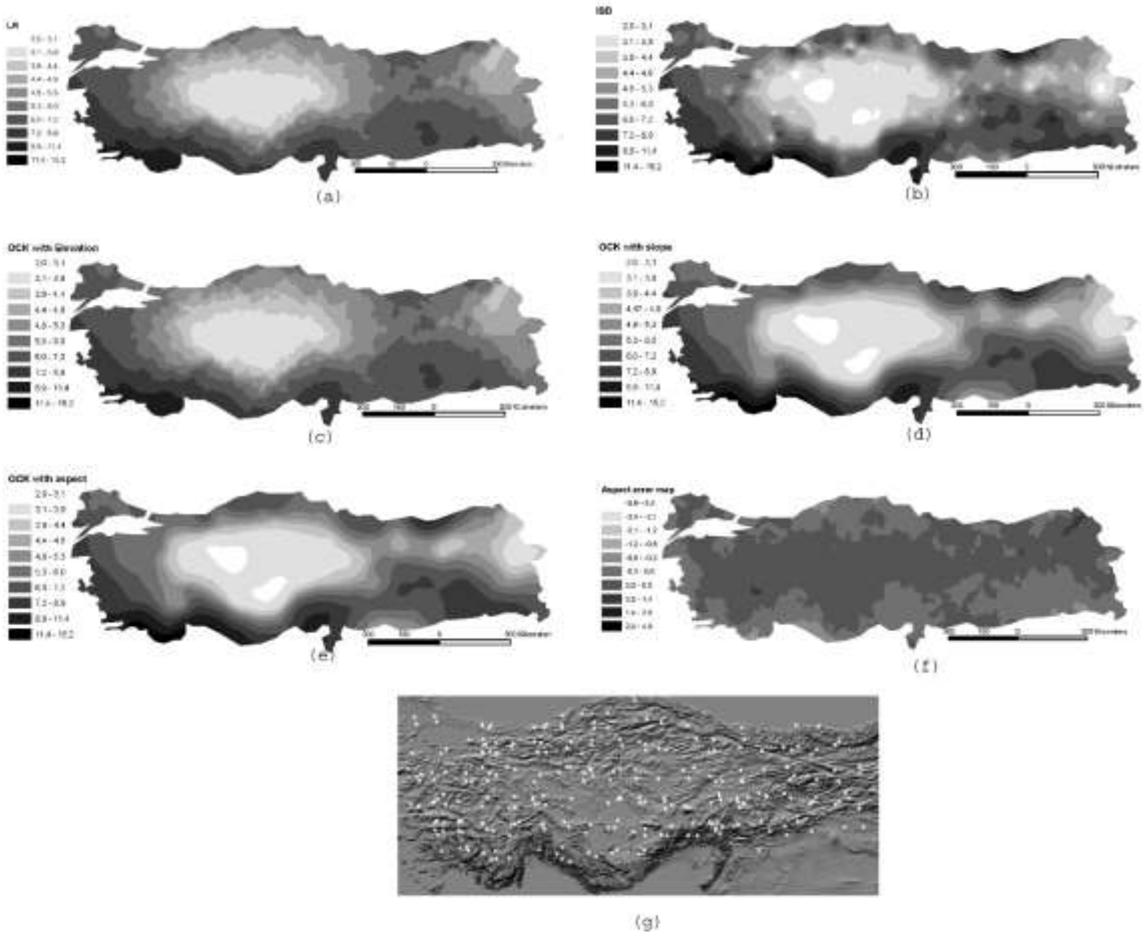

***Figure 3. (a) Mean annual precipitation map produced by LR, (b) mean annual precipitation map produced by ISD, (c) mean annual precipitation map produced by OCK with elevation, (d) mean annual precipitation map produced by OCK with slope, (e) mean annual precipitation map produced by OCK with aspect, (f) aspect standard error map, (g) digital elevation model (250 × 250 m resolution)***

Turkey is affected by cold air masses from the north and warm air masses from the south. The study area is in the central climatic zone, which has four distinct seasons. However, areas in the warm climatic zone have some different climatic properties. The reasons for these differences are:
-they are surrounded by sea on three sides.
-they lie parallel to the mountains and coastline.
-the mountains prevent sea effects reaching the interior area.
-increasing elevation from west to east.
According to Turkish meteorologists, severe unpredictable rainy days will occur following severe hot climatic conditions and the number of these days will increase during the next five years. This suggests a scenario of future climate in Turkey, where aridity will threaten southern regions and torrents will affect regions in the north.

Global Journal of Engineering Science and Research Management

## *References*


1. *Agnew MD, Palutikof JP. 2000. GIS-based construction of baseline climatologies for the Mediterranean using terrain variables. Climate Research 14: 115–127.*
2. *Alijani B, Ghohroudi M, and Arabi N. 2008. Developing a climate model for Iran usin GIS. Theoretical and applied Climatology, 92, 103-112*
3. *Attorre F, Alfo M, Sanctis DM, Francesconi F and Bruno F 2007. Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. International journal of climatology 27, 1825-1843.*
4. *Bajat, B., Pejović, M., Luković, J., Manojlović, P., Ducić, V., Mustafić, S. 2013. Mapping average annual precipitation in Serbia (1961-1990) by using regression kriging. Theoretical & Applied Climatology. 112-2, pp1-13.*
5. *Burrough PA. 2001. GIS and geostatistics: essential partners for spatial analysis. Environmental and Ecological Statistics 8: 361–377.*
6. *Brown DP, Comrie AC. 2002. Spatial modelling of winter temperature and precipitation in Arizona and New Mexico, USA. Climate Research 22: 115–128.*
7. *Caeiro S, Goovaerts P, Painho M, and Costa MH 2003 Delineation of Estuarine Management Areas Using Multivariate Geostatistics: The Case of Sado Estuary Environ. Sci. Technol., 37 (18), pp 4052–4059*
8. *Diodato N, Ceccarelli M. 2005 Interpolation processes using multivariate geostatistics for mapping of climatological precipitation mean in the Sannio Mountains (southern Italy), Earth Surface Processes and Landforms. Pp.259-268*
9. *Erxleben J, Elder K, Davia R, 2002. Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rock Mountains. Hydrological Processes, 16 3627-3649.*
10. *Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology 228(1–2): 113–129.*
11. *Hold away MR. 1996. Spatial modelling and interpolation of monthly temperature using kriging. Climate Research 6: 215–225.*
12. *Iaco, S., Palma, M., Posa, D. 2013. Prediction of Particle Pollution through Spatio-temporal Multivariate Geostatistical Analysis , AStA: Advances in Statistical Analysis, 97-2, pp. 133-50*
13. *Johnston, K., Ver Hoef, J.M., Krivoruchko, K., and Lukas, N., 2003. Using ArcGis Geostatistical Analyst, ESRI, 300 pp.*
14. *Laguardia, G. 2011 representing the precipitation regime by means of Fourier series. International Journal of Climatology. 31- 9, p1398-1407*
15. *Krakauer, N.Y. 2012.Estimating Climate Trends: Application to United States Plant Hardiness Zones. Advances in Meteorology. pp1-9.*
16. *Milillo TM and Gardella J. (2008) Spatial Analysis of Time of Flight−Secondary Ion Mass Spectrometric Images by Ordinary Kriging and Inverse Distance Weighted Interpolation Techniques. Anal. Chem. pp 4896–4905*
17. *Mouser PJ, Rizzo DM, Röling, WFM and Breukelen BM. (2005) A Multivariate Statistical Approach to Spatial Representation of Groundwater Contamination using Hydrochemistry and Microbial Community Profiles Environ. Sci. Technol., 2005, 39 (19), pp 7551–7559*
18. *Pearce, JL; Rathbun, SL; Aguilar-Villalobos, M; Naeher, LP. 2009 Characterizing the spatiotemporal variability of PM2.5 in Cusco, Peru using kriging with external drift*
19. *Atmospheric Environment, 43; 12; p2060-p2069*
20. *Perry M, Hollis D (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. International journal of climatology. pp.1041-1054*
21. *Saito H and Goovaertsm P. 2000 Geostatistical Interpolation of Positively Skewed and Censored Data in a Dioxin-Contaminated Site. Environ. Sci. Technol. , 34 (19), pp 4228–4235*
22. *Vienneau, D.; de Hoogh, K.; Briggs, D. 2009. A GIS-based method for modelling air pollution exposures across Europe. Science of the Total Environment. 408-2, p255-266.*