

Comparison of Spatial Interpolation Methods for Precipitation in Ningxia, China

Wu Hao^{1,2}, Xu Chang^{1,2}

¹School of Environmental Science and Engineering, Chang'an University, 126, Yanta road, Xi'an, Shaanxi, 710054, China

²Key Laboratory of Subsurface Hydrology and Ecological Effect in Arid Region of Ministry of Education, Xi'an, Shaanxi, 710054, China

Abstract: As basic data, the reliability of precipitation data makes a significant impact on many results of environmental applications. In order to obtain spatially distributed precipitation data, measured points are interpolated. There are many spatial interpolation schemes, but none of them can perform best in all cases. So criteria of precision evaluation are established. This study aims to find an optimal interpolation scheme for rainfall in Ningxia. The study area is located in northwest China. Meteorological stations distribute at a low density here. Six interpolation methods have been tested after exploring data. Cross-validation was used as the criterion to evaluate the accuracy of various methods. The best results were obtained by cokriging with elevation as the second variable, while the inverse distance weighting (IDW) perform worst. Three types of model in cokriging were compared, and Gaussian model is the best.

Keywords: precipitation, rainfall, interpolation, cross-validation, kriging, IDW.

1. Introduction

Rainfall data, as a climatic factor, is important parameters in environmental studies, such as climate modeling and hydrological modeling. It can be used to estimate runoff, predict meteorological disaster like rainstorm, drought and flood [1]. Continuous precipitation data, as input, is essential for accurate modeling. However, missing data, systematic and random errors are universal due to various reasons [2]. And meteorological stations are usually sparse especially in developing northwest of china. It is uneconomical and difficult to layout high density meteorological stations network because of sparse population and complex terrain in these areas. Therefore, precipitation in no monitoring areas must be estimated by numerical analysis.

Spatial interpolation is a common method to estimate a new date point which is in the range of a series of known isolated points [3]. Interpolation techniques are divided into two main categories: deterministic and geostatistical [4]. Deterministic interpolation methods, such as inverse distance weighted (IDW) and radial basis function (RBF), capitalize on known points to create surfaces, according to the degree of similarity (e.g. IDW) or smoothing (e.g. RBF). Geostatistical interpolation methods, such as kriging, estimate the value of an unknown point utilizing the statistical properties of neighboring known points and taking the spatial autocorrelation among known points into consideration [5]. Each method has advantages and disadvantages, and none of them can perform best in all cases. The best method in a specific case is closely related to the characteristics of the discrete data sets [6].

2. Spatial interpolation schemes

2.1 Kriging and cokriging

Spatial variation observed can be modeled by random processes with spatial autocorrelation which is assumed by kriging based on regionalized variables [7]. It considers both

the degree and distance of variation among measured points.

2.2 Inverse distance weighting

Inverse distance weighting (IDW) assumes that every known point has an influence on the predicted point. The weight is proportional to the p -th power of inverse distance [8]. The optimal value of p can be obtained by minimizing the root mean square prediction error (RMSPE) which is a statistic calculated during cross-validation.

2.3 Radial basis functions

Radial basis functions (RBF) methods are a series of splines to the multivariate case. It includes several basis functions, such as thin-plate spline function, Gaussian function, multiquadric function and inverse multiquadric function [9].

2.4 Polynomial interpolation

Polynomial interpolation is described by a polynomial function. Polynomial interpolation includes global polynomial interpolation (GP) and local polynomial interpolation (LP). They have smooth surface, but unlike other interpolation methods, they are inexact interpolation methods.

3. Materials and methods

3.1 Study area and available data

Ningxia Hui Autonomous Region (66,400 km²), located in the northwest of China, belongs to continental climate of arid and semi-arid with temperature from about -10 to 25 °C. The south part belongs to the Loess Plateau, while the north part is consisted of the Yellow River alluvial plain, Helan Mountains and Ordos platform. Its elevation ranges from about 1000 to 3000 m. The layout arises from north to south due to several mountains in the south Loess Plateau. It benefits the invasion of cold air and lifting of air. The precipitation arises from north to south along the terrain. Maximum rainfall which is over 600 mm was recorded in Liupan Mountains; while minimum rainfall is present in Shizuishan with average annual precipitation is 150 mm [10].

The precipitation data in this study were measured by

meteorological stations in Ningxia. Annual mean precipitations of 20 years from 1991 to 2010 at 25 stations are used to interpolation research. The study area and distribution of meteorological stations are shown in Figure 1.

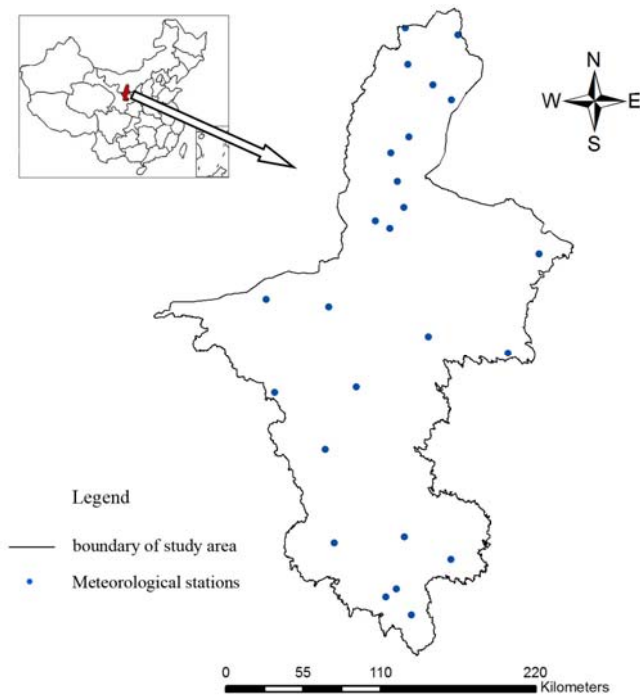


Figure 1: The location of study area and positions of meteorological stations

3.2 Explore Data

Before calculate the annual mean precipitation of 20 years, global and local outliers should be detected by histogram, semivariogram, covariance cloud and Voronoi maps. Meanwhile, semivariogram and covariance cloud will explore the spatial autocorrelation. Trend analysis may instruct to use global or local polynomial interpolation and remove the trend in kriging [11].

Most interpolation schemes do not require the data to be approximate Gaussian distribution except some certain kriging methods. The histogram and normal QQ plot are available for checking the Gaussian distribution. If the data obey Gaussian distribution, the mean should be approach the median, the skewness should be close to zero, and the kurtosis should approximate to three. To avoid lopsided data, utilizing data transformation may convert data to Gaussian distribution [12].

3.3 Interpolation and cross-validation

In this study, six spatial interpolation schemes were explored and compared. Cross-validation is usually used to compare the accuracy of interpolation methods. It removes a data point each time from the data sets and predicts values using the other measured points. Then predicted value at the removed point can be compared with the measured value. This procedure will be carried out for all the measured points. Various evaluation criteria have been used in cross-validation. They have varying sensitivity to different data sets. In this study, mean relative error (MRE) and mean absolute error

(MAE) [13] are used as validation criteria. There are expressed as

$$MRE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Z_i - \bar{Z}}{Z_i} \right| \times 100\% \tag{1}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z_i - \bar{Z}| \tag{2}$$

The interpolation method with smaller values of MAE and MRE is the better schemes.

4. Implementation and discussion of results

4.1 Interpolation

Annual mean precipitation of 20 years from 1991 to 2010 was calculated after outlier detection. From Figure.2, the trend can be seen clearly by three-dimensional perspective. It is obvious that the rainfall increase significantly from north to south and that in middle part is a little bigger than the east and west. The trend will be removed in kriging method.

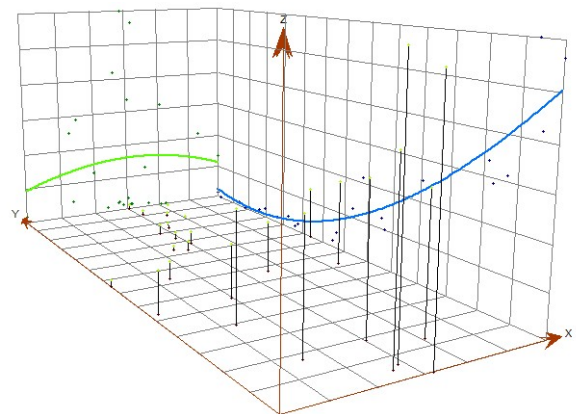


Figure 2: The spatial distribution trend of annual mean precipitation of 20 years

Data characteristic can be described by statistics, such as mean, median, skewness, kurtosis and so on. Table 1 shows the data characteristic and that after data transformation. Obviously, it is more approximate Gaussian distribution after transformation. Mean is closed to median in log transformation, and skewness is closer to 0, therefore, log transformation is adopted.

Table 1: Data characteristic of rainfall data

Statistics	None	Log	Box-Cox
Min	166.9	5.1174	23.838
Max	647.9	6.4737	48.908
Mean	285.11	5.551	30.889
Median	202.1	5.3088	26.432
Std. Dev.	144.16	0.4406	7.8219
Skewness	1.2513	0.7692	0.9953
Kurtosis	3.4957	2.2364	2.777

Std. Dev. is standard deviation. Parameter λ in Box-Cox transformation is 0.5.

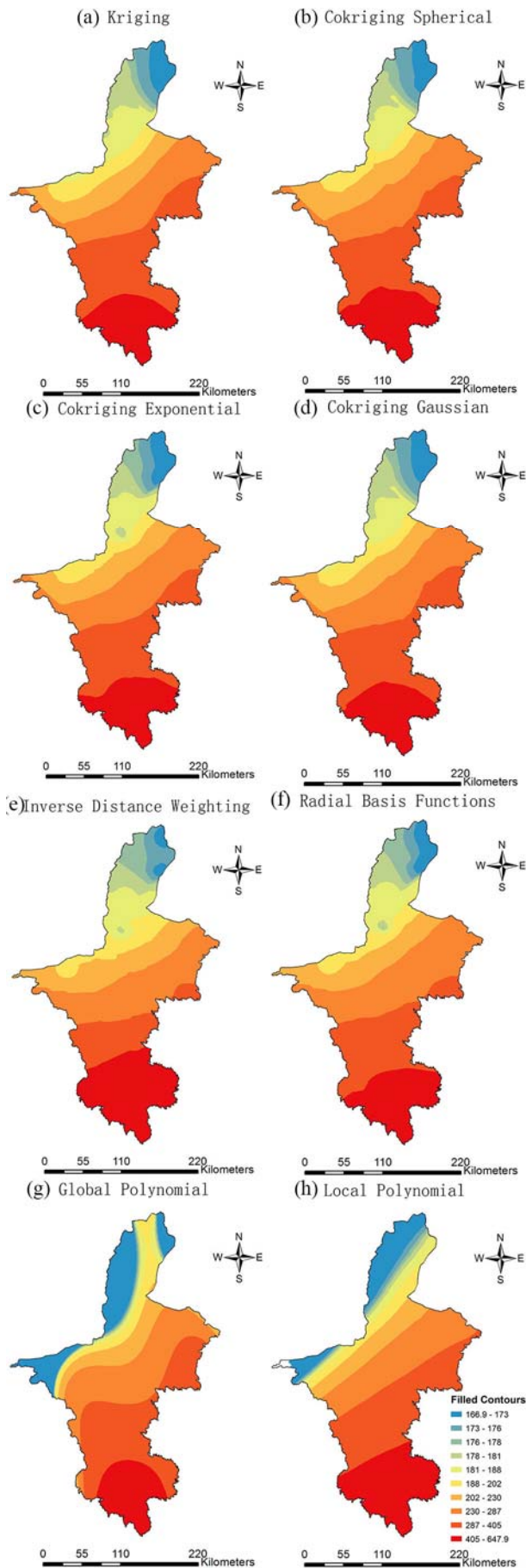


Figure 3: Interpolation surfaces by six schemes and three models for cokriging

4.2 Discussion of results

Figure 3 shows interpolation surfaces interpreted by six methods namely kriging, cokriging, IDW, RBF, GP and LP. In cokriging model, rainfall data is the estimated variable, elevation which have an important influence on precipitation is the second variable. IDW, RBF and LP were mapped with optimal parameters, for example, p equal to 1.3761 was chosen ideally in IDW scheme.

Cross-validation was carried out for six methods and three types of model (spherical, exponential, Gaussian) for cokriging. After a series of test calculation, MAE and MRE which are sensitive to this data were chosen as validation criteria for comparing interpolation accuracy. Table 2 show that IDW with maximum MAE and MRE perform worst in this case, because IDW never predict the maximum and minimum which is unsuited for this case. Cokriging with Gaussian model perform best, because it considers the altitude which has a significant impact on rainfall. Three cokriging models get different results widely demonstrate model in cokriging affect the accuracy of cokriging heavily. Polynomial interpolation is not an exact interpolator. It cannot pass through many measured points. The accuracy of LP is higher than GP obviously because GP takes global trend into account. However, GP has smoother surface than LP.

Table 2: MAE and MRE for six methods

Methods	MAE	MRE
Kriging	31.4588	0.0835
Cokriging Spherical	30.4198	0.0804
Cokriging Exponential	31.6738	0.092
Cokriging Gaussian	30.0135	0.081
IDW	34.224	0.0961
RBF	31.0998	0.0879
GP	31.6738	0.092
LP	30.2338	0.084

5. Conclusion

The six interpolation methods of annual mean precipitation of 20 years in Ningxia, China was compared. The accuracy of interpolation was determined by cross-validation. MAE and MRE were chosen as validation criteria because of high sensitivity in this study. Visual results were given in Figure 3. From the comparison of cross-validation criteria, it has been observed that the best interpolation scheme is cokriging with Gaussian model and elevation as the second variable. However, IDW method is less efficient scheme with biggest MAE and MRE. The maximum and minimum values only occur at the measured points, so the accuracy of IDW is not high. The types of model in cokriging have a significant impact on accuracy. So it is necessary to compare models in cokriging or kriging. Gaussian model perform best, while exponential model perform worst in this study.

References

- [1] S.R. Durrans, S.J. Burian, & R. Pitt, "Enhancement of precipitation data for small storm hydrologic prediction," *Journal of Hydrology*, 299(3), pp. 180-185, 2004.
- [2] R.S.V. Teegavarapu, M. Tufail, & L. Ormsbee, "Optimal functional forms for estimation of missing precipitation data," *Journal of Hydrology*, 374(1), pp. 106-115, 2009.
- [3] D. Kurtzman, & R. Kadmon, "Mapping of temperature variables in Israel: a comparison of different interpolation methods," *Climate Research*, 13, pp. 33-43, 1999.
- [4] V. Sharma, & S. Irmak, "Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: part I. Precipitation and Reference evapotranspiration," *American Society of Agricultural and Biological Engineers*, 55(3), pp. 907-921, 2012.
- [5] D.L. Philips, J. Dolph, & D. Marks, "A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain," *Agricultural and Forest Meteorology*, 58(1-2), pp. 119-141, 1992.
- [6] M. Keblouti, L. Ouerdachi, & H. Boutaghane, "Spatial interpolation of annual precipitation in Annaba-Algeria – comparison and evaluation of methods," *Energy Procedia*, 18, pp. 468-475, 2012.
- [7] P. Goovaerts, "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall," *Journal of Hydrology*. 228(1), pp. 113-129, 2000.
- [8] L. de Mesnard, "Pollution models and inverse distance weighting: Some critical remarks," *Computers & Geosciences*. 52, pp. 459-469, 2013.
- [9] S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation," *Advances in Computational Mathematics*, 11, pp. 193-210, 1999.
- [10] P.Y. Li, & J.H. Wu, "Water resources status and change law in Ningxia," *Water Conservancy Science and Technology and Economy*. 16(4), pp. 405-407, 2010.
- [11] ESRI Inc, "ArcGIS Help 10," ESRI Inc., Redlands, USA.
- [12] R. Webster, & M.A. Oliver, "Geostatistics for Environmental Scientists (Second Edition)," John Wiley & Son, Ltd, Hoboken, New Jersey.
- [13] F. Li, Y. Sun, & J.J. Zheng, "Spatial interpolation of precipitation in Anhui province," *Research of Soil and Water Conservation*, 17(5), pp. 183-186, 2010.

Author Profile



Wu Hao received the bachelor degree in groundwater science and engineering from Chang'an University in 2012. Presently he studies for the master degree in engineering from Chang'an University.



Xu Chang received the bachelor degree in groundwater science and engineering from Chang'an University in 2012. Now, He is studying for a master's degree of Hydraulic Engineering in Chang'an University.