Precipitation Interpolation by Multivariate Bayesian Maximum Entropy Based on Meteorological Data in Yun-Gui-Guang region, Mainland China

Chaolin Wang¹, Shaobo Zhong¹*, Fushen Zhang¹, Quanyi Huang¹
¹Department of Engineering Physics/ Institute of Public Safety Research, Tsinghua University, Beijing 100084, China
*Corresponding author: zhongshaobo@tsinghua.edu.cn

Abstract. Precipitation interpolation has been a hot area of research for many years. It had close relation to meteorological factors. In this paper, precipitation from 91 meteorological stations located in and around Yunnan, Guizhou and Guangxi Zhuang provinces (or autonomous region), Mainland China was taken into consideration for spatial interpolation. Multivariate Bayesian maximum entropy (BME) method with auxiliary variables, including mean relative humidity, water vapour pressure, mean temperature, mean wind speed and terrain elevation, was used to get more accurate regional distribution of annual precipitation. The means, standard deviations, skewness and kurtosis of meteorological factors were calculated. Variogram and cross-variogram were fitted between precipitation and auxiliary variables. The results showed that the multivariate BME method was precise with hard and soft data, probability density function. Annual mean precipitation was positively correlated with mean relative humidity, mean water vapour pressure, mean temperature and mean wind speed, negatively correlated with terrain elevation. The results are supposed to provide substantial reference for research of drought and waterlog in the region.

1. Introduction

Precipitation has close relation to some meteorological and environmental factors. Humidity, water vapor pressure, temperature, wind and terrain elevation etc. are potential meteorological factors that affect the spatial distribution of precipitation. However, the observations of meteorological stations are discrete and sparse in time and space. Hence, spatial interpolation for meteorological factors is necessary to obtain regionalized distribution of annual mean precipitation [1]. Precipitation interpolation has been a hot area of research for many years. Bayat et al. [2] compared Bayesian maximum entropy (BME) and ordinary kriging methods for spatial and spatiotemporal variations of annual precipitation prediction, and cross-validation was used to assess their performances. Shi et al. [3] employed BME model taking TRMM satellite data as soft data, observations of meteorological stations as hard data, to estimate the precipitation of Fujian Province, China. Many researches have demonstrated that an appropriate covariate can improve the estimation accuracy largely [4, 5]. Zhang et al. [6] compared four BME models to estimate the spatial distribution of mean annual precipitation using meteorological, elevation and the nearest distance to coastline variables. The results were assessed by cross validation and proved that multivariate BME method was optimal. Most methods and models were only based on indexes using hard data. However, soft data not only can provide some useful information, but also can save calculation cost and improve the predictions. The application of spatial distribution of precipitation was helpful to risk assessment of drought and inundation, which could map disaster risk accurately combined with disaster index (say, drought index).

This paper took precipitation data as main variable, mean relative humidity, water vapor pressure, mean temperature, mean wind speed and terrain elevation as auxiliary variables to study the spatial...
interpolation of precipitation in Yunnan, Guizhou provinces and Guangxi Zhuang autonomous region during 1964-2013. The means, standard deviations, skewness and kurtosis of meteorological factors were calculated. Variogram and cross-variogram were fitted between precipitation and auxiliary variables. The results are supposed to provide substantial reference for research of drought and waterlog in the region.

2. Materials and methods

2.1. Study area

The study area is Yunnan, Guizhou provinces and Guangxi Zhuang autonomous region (Yun-Gui-Guang region, 97°E-113°E, 20°N-29°N), which is located in the Yunnan-Guizhou Plateau, southwest of Mainland China, eastern along the border of China, Vietnam, Laos and Myanmar and coastline of South China Sea (Figure 1). Mean terrain elevation of the region is 1000~2000 meters. This region covers an area of 795,300km², and contains 39 big cities, such as Kunming, Guiyang and Nanning. The climate of this region is subtropical monsoon climate. Drought happens easily in spring because of poor rainfall and strong evaporation. There are 129.77 million resident people in the region. Guizhou and Guangxi are relatively richer in water resource than Yunnan. And all this region is undeveloped provinces compared to east China coastal area.

![Figure 1. The terrain of Yun-Gui-Guang region and distribution of meteorological stations.](image)

2.2. Data

Data used in this paper was acquired from the annual China’s terrestrial climate data set, which was published by China Meteorological Data Sharing Service System (http://cdc.nmic.cn). We took 50 years (1964-2013) of annual precipitation as main variable, mean relative humidity, water vapor pressure, mean temperature and mean wind speed as covariates to study the spatial distribution of precipitation from 91 meteorological stations in or around the study area. However, Observations are not recorded completely at 19 meteorological stations which are replaced by -32766 or 32766. In this paper, they were replaced by soft data, precipitation probability density function(pdf) which was matched by kernel density estimation method by MATLAB software using precipitation data set from 1951 to 2015. These 19 stations were called meteorological station of soft data; other 72 stations were called meteorological stations of hard data. All stations were marked in map in Figure 1. Digital Elevation Model(DEM) data were also used in this paper, called ASTER GDEM v2(Figure 1). The spatial resolution is 0.1°.

2.3. BME method

Christakos (1990) established the BME method. BME method takes many types of data and different types of knowledge bases into spatial interpolation. These data and information are divided into general knowledge(\( \kappa_{G} \)) and site-specific knowledge(\( \kappa_{S} \)) [7]. The steps for calculating BME are as follows.
**Prior stage.** The aim of prior stage is to figure out the pdf, based on general knowledge \( K_G \), called prior pdf. Information entropy is defined as equation (1) by Shannon information measure [8]:

\[
\text{Info}_G(Z_{\text{map}}) = - \log f_G(Z_{\text{map}})
\]

Where, \( Z_{\text{map}} \) is the stochastic variable in the study area, \( Z_{\text{map}} = (Z_{\text{hard}}, Z_{\text{soft}}) \), \( Z_{\text{hard}} \) and \( Z_{\text{soft}} \) indicate the value of hard data, soft data and location for estimating, respectively. \( f_G(Z_{\text{map}}) \) indicates the pdf based on general knowledge \( K_G \). Based on these constraints and Lagrange multiplier approach, we can get the prior pdf:

\[
f_G(Z_{\text{map}}) = A^{-1} \exp \left( \sum_{\alpha=1}^{N} \mu_{\alpha} g_{\alpha}(Z_{\text{map}}) \right)
\]

Where, \( \mu_{\alpha} \) indicates Lagrange’s multiplier, \( g_{\alpha}(Z_{\text{map}}) \) is the known function associated with \( Z_{\text{map}} \) based on \( K_G \), \( A \) indicates normalization coefficient:

\[
A = \int \exp \left( \sum_{\alpha=1}^{N} \mu_{\alpha} g_{\alpha}(Z_{\text{map}}) \right) dZ_{\text{map}}
\]

**Pre-posterior stage.** The aim of pre-posterior stage is to collect and organize additional auxiliary information in appropriate forms to produce site-specific knowledge \( K_S \). And then they will be used in BME model. Hard data has been incorporated into prior stage indirectly, and will be used directly at this stage.

**Posterior stage.** The aim of posterior stage is to update prior pdf based on Bayesian conditional probability theorem and \( K_S \). So we can get posterior pdf. When the distribution of hard data and soft data is certain, posterior pdf of spatial variable \( Z \) at location \( x_0 \) is:

\[
f_K(Z_0) = f_G \left( Z_0 \mid Z_{\text{hard}}, Z_{\text{soft}} \right) = f_G \left( Z_0, Z_{\text{hard}}, Z_{\text{soft}} \right) \times \left( f_G \left( Z_{\text{hard}}, Z_{\text{soft}} \right) \right)^{-1}
\]

Where, \( Z_{\text{hard}} = [x_1, \ldots, x_n] \), \( Z_{\text{soft}} = [x_{n+1}, \ldots, x_m] \), \( n, m \) indicate the number of hard data and soft data within the scope of maximum distance \( d_{\text{max}} \) to estimation point, respectively.

### 3. Numerical results and plots

Table 1 shows the means, standard deviations, skewness and kurtosis of meteorological factors during 1964 to 2013 in Yun-Gui-Guang region. Figure 2 and figure 3 show the pdf and the cumulative density function(CDF) of meteorological factors in study area, respectively.

**Table 1.** The basic statistical indicates of meteorological factors during 1964 to 2013 in Yun-Gui-Guang region.

|                      | Annual Mean Precipitation(0.1mm) | Mean Relative Humidity(%) | Water Vapor Pressure(0.1hPa) | Annual Mean Temperature(0.1℃) | Annual Mean Wind Speed(0.1m/s) |
|----------------------|----------------------------------|---------------------------|------------------------------|--------------------------------|--------------------------------|
| Means                | 12974                            | 95.07                     | 184.78                       | 194.30                         | 98.73                          |
| Standard Deviations  | 1.91×107                         | 5.94×105                  | 5.92×105                     | 5.91×105                       | 2.68×106                       |
| Skewness             | 1.1716                           | 42.382                    | 42.215                       | 42.234                         | 19.922                         |
| Kurtosis             | 2.76                             | 1.79×103                  | 1.78×103                     | 1.79×103                       | 394.87                         |
The constitutive property and randomness of annual mean precipitation can be described by variogram. We set distance lag as 1° to compute variogram value in interval (0, 10). The directional variograms of precipitation along the N-S (stars), the W-E (circles) directions and omnidirectional variogram of precipitation (red circles) are computed and shown in Figure 4. The calculated variogram value is shown by blue line in Figure 4. Then Nested model is defined with Nugget model and Gaussian model. Then we can get the equation after fitted by least square method:

$$\gamma(h) = 1.115 \times 10^7 + 1.501 \times 10^7 \times \left[1 - \exp\left(-3h^2 / 10^2\right)\right], \quad h > 0$$

(5)

Where, first half part of the equation is Nugget model, and the other is Gaussian model. Fitted variogram curve is shown by red line in Figure 4. It can be seen that the fitted curve coincides with the calculated value of variogram well. Increasing of variogram with distance lag indicates that the correlation of annual precipitation increases with distance lag.

In this paper, we take mean relative humidity, water vapor pressure, mean temperature, mean wind speed and terrain elevation as cross variates. Cross-variograms between annual precipitation and cross variates describe the correlation of spatial distribution. The nested models are defined same to variogram. Cross-variograms between any two variates are fitted and are shown in Figure 5. As we can see, the correlations of cross-variogram between mean relative humidity, water vapor pressure, mean temperature and annual precipitation are positive, terrain elevation is negative. And mean wind speed is unstable.
Figure 4. Variogram of annual precipitation.

Figure 5. Cross-variogram of Variates ($h > 0$). (1) Annual mean precipitation; (2) Mean relative humidity; (3) Mean water vapor pressure; (4) Mean temperature; (5) Mean wind speed; (6) Terrain elevation.

Based on the achievement in the past, multivariate BME method were employed to estimate the spatial distribution of annual precipitation in Yun-Gui-Guang region. We studied year by year from 1964 to 2013 at a 0.05° spatial resolution. Then the annual precipitation distribution images were mapped in ARCGIS 10.2.2 software. Figure 6 shows the interpolation annual mean precipitation distribution maps of Yun-Gui-Guang region. Overall, there is more rain in the south and southeast because of being adjacent to South China Sea, and it is rainless in the north-west where is deep in Mainland China. In addition, there is more rain in south than north of the study area, east than west.
4. Conclusions
This paper takes annual precipitation with auxiliary variables, mean relative humidity, water vapor pressure, mean temperature, mean wind speed and terrain elevation into consideration for spatial interpolation. The PDF and CDF are analyzed for multivariate BME method, which are used to get more accurate regional distribution of annual mean precipitation. The variogram of annual precipitation and cross-varioigram of mean relative humidity, water vapor pressure, mean temperature, mean wind speed and terrain elevation are studied to get the correlation between mean annual precipitation and covariates. Finally, the spatial distribution of annual precipitation was mapped in Yun-Gui-Guang region. The results show that:

1. Annual mean precipitation was positively correlated with mean relative humidity, mean water vapor pressure, mean temperature and mean wind speed, negatively correlated with terrain elevation.
2. There is more rain in the south and southeast, and it has a trend that there is more rain in south than north of the study area, east than west.

Some natural disasters, such as drought, waterlog and so on, have close relation to precipitation. The precipitation map is obtained in this paper by BME method of the study area of interest. The results are supposed to provide substantial reference for disaster prevention and mitigation of drought and waterlog in the region. However, the mapping procedure of precipitation is also affected by the factor of time. The space-time estimation method of precipitation should be paid more attention in further research.

Acknowledgments
The authors would like to thank the support of the National Natural Science Foundation of China (Grant No. 91224004), the Project in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (Grant No.2015BAK12B03). The authors appreciate support for this paper by the Collaborative Innovation Center of Public Safety.

References
[1] Yu, H., Chen, J., Christakos, G. and Jerrett, M. 2009 Environ Health Perspect 117 537-544.
[2] Bayat, B., Zahraie, B., Taghavi, F. and Nasseri, M. 2013 Theoretical and applied climatology, 113 429-444.
[3] Shi, T., Yang, X., Zhang, T., Liu, L. and Tian, W. 2014 Journal of Geo-information Science 16 470-481.
[4] Yu, X., Ma, J., Zhu, J. and Sun, S. 2011 Science of Surveying and Mapping 36 83-85+55.
[5] Jost, G., Heuvelink, G. B. M. and Papritz, A. 2005 Geoderma 128 258-273.
[6] Zhang, F., Yang, Z., Zhong, S. and Huang, Q. 2016 Advances in Meteorology 2016 1-13.
[7] Douaik, A., Van Meirvenne, M. and Tóth, T. 2005 Geoderma 128 234-248.
[8] Shannon, C. E. 1948 Bell System technical Journal 27 379-423 and 623–656.