

Research Article

Exploring Mean Annual Precipitation Values (2003–2012) in a Specific Area (36°N–43°N, 113°E–120°E) Using Meteorological, Elevational, and the Nearest Distance to Coastline Variables

Fushen Zhang, Zhitao Yang, Shaobo Zhong, and Quanyi Huang

Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Shaobo Zhong; zhongshaobo@tsinghua.edu.cn

Received 5 September 2015; Accepted 26 October 2015

Academic Editor: Francesco Viola

Copyright © 2016 Fushen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gathering very accurate spatially explicit data related to the distribution of mean annual precipitation is required when laying the groundwork for the prevention and mitigation of water-related disasters. In this study, four Bayesian maximum entropy (BME) models were compared to estimate the spatial distribution of mean annual precipitation of the selected areas. Meteorological data from 48 meteorological stations were used, and spatial correlations between three meteorological factors and two topological factors were analyzed to improve the mapping results including annual precipitation, average temperature, average water vapor pressure, elevation, and distance to coastline. Some missing annual precipitation data were estimated based on their historical probability distribution and were assimilated as soft data in the BME method. Based on this, the univariate BME, multivariate BME, univariate BME with soft data, and multivariate BME with soft data analysis methods were compared. The estimation accuracy was assessed by cross-validation with the mean error (ME), mean absolute error (MAE), and root mean square error (RMSE). The results showed that multivariate BME with soft data outperformed the other methods, indicating that adding the spatial correlations between multivariate factors and soft data can help improve the estimation performance.

1. Introduction

The spatial distribution of precipitation over any region is an important parameter in the study of climate, flood disasters caused by heavy rain, and several other water-related issues. Although precipitation is typically gauged using ground-based meteorological observation stations, interpolating these records as a continuous map is useful for many environmental studies and risk assessments related to water-based issues [1, 2].

Although these sampled precipitation data sources collected through ground-based meteorological stations provide useful information that can be used to estimate the spatial distribution of precipitation, the main interest in the coupling between climate and topography has highlighted the need to accurately describe precipitation as a function of both elevation and position on the landscape [3–9]. Therefore, meteorologists and geographers still face several practical and methodological challenges in analyzing and modeling

sampled precipitation data, and they have concerns related to elevation and position data.

One challenge is analyzing how the spatial distribution of precipitation is influenced by elevation at each measuring location and by the nearest distance from that location to the coastline; this area still remains understudied. It is generally accepted that the above two aspects, elevation and distance to the coastline, affect the spatial distribution of precipitation to a certain degree [4]. Terrain uplifting, which occurs on windward flank, leads air masses to rise, expand, and cool adiabatically which results in increasing relative humidity of each air mass, creating clouds and precipitation. Generally, a universal law of the distribution of precipitation states that areas closer to the sea have a stronger marine influence; in contrast, areas farther from the sea receive less precipitation because water vapor from the ocean cannot easily reach far inland. In addition, the quality and representativeness of the sampling data have a strong influence on constructing the models of these relationships. Clearly, the

biased positioning of precipitation gauges cannot sufficiently express the complex distribution processes of precipitation in some area, particularly on slopes. As discussed below, a large body of literature contains assessments of these relationships. Through the use of geographically weighted regression, the relationship between annual precipitation and gauge elevation over Great Britain was reresearched [7]. Some researchers have developed computer-based methods of mapping precipitation based on precipitation-elevation relationships and topography [10], but the effects of topography are generally not accounted for explicitly. Although most researches indicate that precipitation trends to increase with increased elevation and decrease with increasing the nearest distance to the coastline, this relationship can vary considerably and may not be valid at all in some situations.

The second issue relates to the effects of the spatial correlations of meteorological factor to the distribution of precipitation. Accurate meteorological data can be available only at the sampling locations, that is, the meteorological stations; if researchers want to get the values at any other locations, they must be inferred from neighboring meteorological stations or from relationships with other related variables. Due to the fact that many other geographical and positional variables hinging on climate and related data are normally inaccessible, spatial autocorrelation of meteorological variables is of special interest. Therefore, integrating the spatial correlation information of these meteorological factors provides a better method of improving the accuracy of precipitation estimation. A common view in many earlier studies is that methods that make use of the relationship between estimated variable and auxiliary variables usually provide more accurate estimated results than approaches that are based on only one variable such as precipitation measurements [7]. These techniques can obtain satisfactory results from limited sampling data, based mainly on the geographical position of the sampling locations, on the topological relationships between these locations, and on the value of the variable to be measured.

Nevertheless, these related studies generally only take into account spatial relationships among sampling locations but do not consider the influences of the spatial correlations between meteorological factors. In addition, the results acquired by these researches in specific areas are still not very accurate.

Moreover, because variables cannot be measured in all locations of the studied region, researchers who deal with those variables have to use interpolation techniques in their studies. At present, ground-based meteorological observation systems cannot provide a continuous distribution map of meteorological factors; complex topography is one of the reasons, and the sparsity of meteorological stations in these regions is another reason [11]. In this way, spatial interpolation techniques are essential to the mapping of meteorological factors, and values at unmeasured locations can be estimated based on those observations. A number of interpolation techniques have been proposed for mapping precipitation at high resolution, such as inverse distance weighting, linear or nonlinear regression, geographically weighted regression, artificial neural networks, and the use

of Thiessen polygons, ordinary kriging, universal kriging, and regression kriging methods [11–13]. From these studies, because of their inherent constraints between several variables, the above spatial interpolation techniques do not always provide successful estimates. The Bayesian maximum entropy (BME) approach offers a framework for spatiotemporal analysis and mapping, and it can integrate various prior knowledge bases and assimilate uncertain observations (soft data). A burgeoning literature finds that BME approach is more accurate and physically meaningful than classical spatial interpolation techniques [12, 13], and these advantages facilitate the process of spatially mapping multiple meteorological factors accurately.

In view of the above considerations, in this study, we estimate mean annual precipitation data for our study area (36°N–43°N, 113°E–120°E) using annual precipitation data from 48 meteorological stations. In addition, spatial correlations between five variables were analyzed, including annual precipitation, average temperature, average water pressure, elevation, and distance to coastline. We used Bayesian maximum entropy (BME) by means of the BMElib 2.0 which is a powerful MATLAB numerical toolbox of Modern Spatiotemporal Geostatistics to implement BME theory. In addition, different BME methods were implemented with different datasets. The estimation accuracy was assessed by cross-validation with ME, MAE, and RMSE.

2. Materials and Methods

2.1. Area Description and Preprocessing of Original Data. The study area is located in the Haihe Plain (36°N–43°N, 113°E–120°E) in the northern part of the Huabei Plain. Taihang and Yanshan mountains are located in the western and northern parts of the region, respectively. The eastern and southern areas are adjacent to the Yellow River and Bohai Sea. Beijing, the capital of China, is located in the center of this region as are other large cities, that is, Tianjin and Shijiazhuang; more than 70 million people live in the region. Cultivated land covers about 8 million ha in this region. However, in this temperate monsoon climate, drought often strikes in the spring, and flooding in the summer. Figure 1 shows the study area and distribution of meteorological stations.

In this study, in addition to estimated variable (annual precipitation), some other variables were selected as ancillary variables that were used to find the distribution of precipitation. Table 1 lists a total of five variables that were used in the Bayesian maximum entropy method for spatial interpolation analyses. Figure 2 shows the nearest distance to the coastline of meteorological station.

The 48 meteorological stations in this region recorded annual precipitation, annual average temperature, and annual average water vapor pressure from 1959 to 2012. At 40 of these locations, so-called “true” values of the meteorological observed data were available. The mean values of these meteorological observed data of the period from 2003 to 2012 were calculated; these were considered exact data; that is, at these locations the sampling data were deemed to have no measurement errors and are used as hard data in the BME method. These locations were called “hard data locations.”

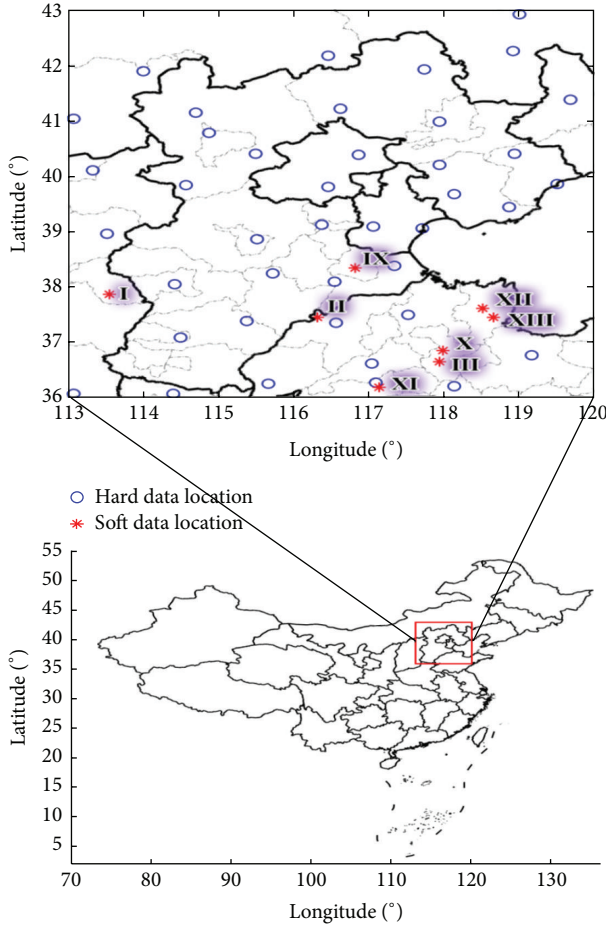


FIGURE 1: The study area and distribution of meteorological stations.

TABLE 1: Variables used in the Bayesian maximum entropy method.

Variable	Description
Annual precipitation	Annual precipitation at meteorological stations
Average temperature	Annual average temperature at meteorological stations
Average water vapor pressure	Annual average water vapor pressure at meteorological stations
Elevation	Height of region at the resolution of 1 radian second (about 30 m)
Distance to coastline	Distance to coastline at each $0.1^\circ \times 0.1^\circ$ grid point

At the remaining 8 locations, the available information consisted of historical statistical laws of annual precipitation, and these locations are called “soft data locations.” Annual precipitation data were missing for eight of those stations in some years during the period. To assimilate this part of the data into the BME method, the probability density functions (pdfs) of annual precipitation were estimated based on the historical values (1983 to 2012) from the corresponding stations allowing the missing values to be represented by the

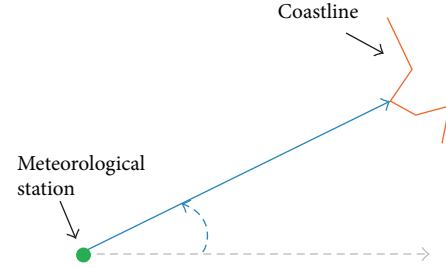


FIGURE 2: The nearest distance to the coastline.

pdfs. Then the ten-year mean values could be calculated in the probability distribution form, that is, probability soft data.

The region was divided into a $0.1^\circ \times 0.1^\circ$ grid; then, the distance to coastline was calculated for each grid point using ArcGIS 10.2.

Elevation data for the region were based on the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) version 2 [14]. The resolution of ASTER GDEM v2 data was 1 radian second, approximately 30 m. Figure 3 shows the detailed data preprocessing procedure.

2.2. Multivariate Statistical Analysis and Variograms. We used data from the 40 sets of “hard data” and eight sets of “soft data” from 48 meteorological stations as described above for multivariate analysis and to produce variograms. Statistical methods were used to relate the precipitation variable to other meteorological variables, the elevation variable, and the nearest distance to the coastline variable.

Density scaled histograms use regularly spaced bins between specified minimum and maximum values. The histogram has the meaning of a probability distribution function; that is, the sum of the displayed rectangular surfaces is equal to 1. Kernel density estimates were evaluated from -7000 to $35,000$ by steps of 1000 (annual precipitation), from -170 to 400 by steps of 5 (average temperature), and from -70 to 350 by steps of 1 (average water pressure).

The spatial distribution of an estimated variable (in our case, the mean annual precipitation) is represented by means of a spatial random field $X(p)$, where the vector p denotes spatial location. We want to estimate the mean annual precipitation values of arbitrary positions within the study region, $X(p)$, at a location, p_k ($k \neq i = 1, \dots, m$), given data at the adjacent meteorological stations in this region, p_i ($i = 1, \dots, m$).

The spatial relationships between precipitation and the other independent variables depended on the distance between these variables, which are represented by variograms and cross-variograms. They are a prerequisite for the analysis, and their use can enhance the accuracy of the estimated results. The variogram $\gamma(h)$ can be represented as

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [X(p_k + h) - X(p_k)]^2, \quad (1)$$

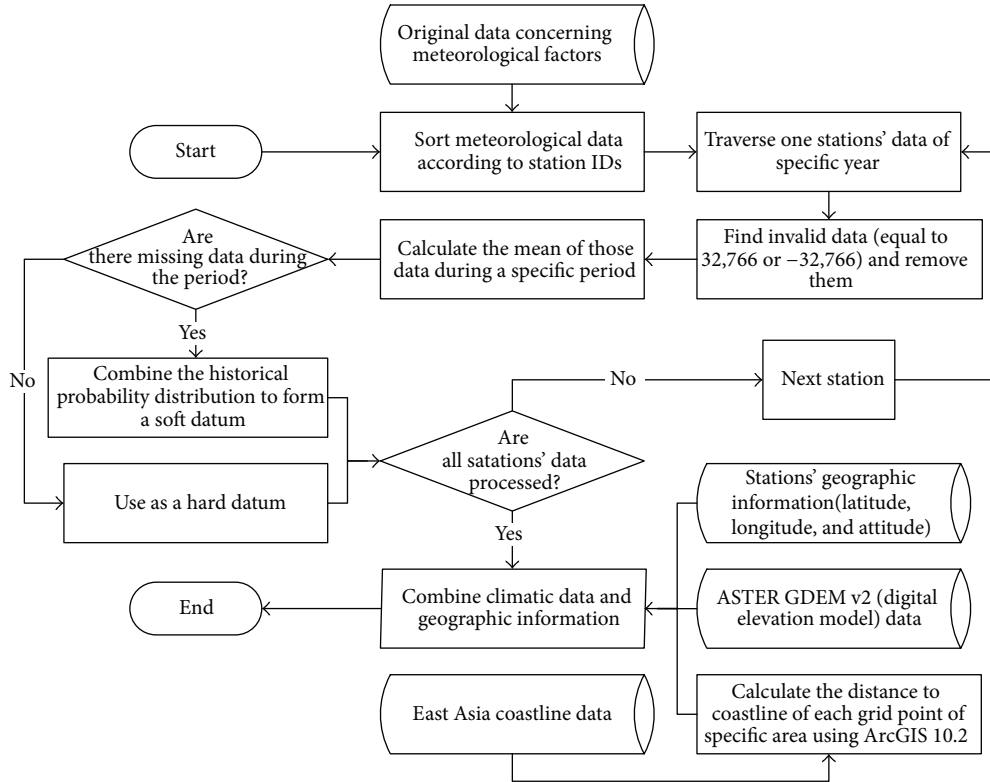


FIGURE 3: Flowchart of data preprocessing.

where h denotes the lag distance which separates pairs of meteorological stations; $X(p_k)$ is the value of annual precipitation at location p_k , and $X(p_k + h)$ denotes the value at location $(p_k + h)$ of annual precipitation; $n(h)$ represents the number of pairs separated by the distance h ; $\gamma(h)$ denotes the variogram for the lag-distance h .

Besides, cross-variograms between two used variables (ϕ and φ) in the estimation process can be represented as

$$\gamma_{\phi,\varphi}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [X_{\phi}(p_k + h) - X_{\phi}(p_k)] \cdot [X_{\varphi}(p_k + h) - X_{\varphi}(p_k)]. \quad (2)$$

The nested variogram can be modeled as linear combinations:

$$\gamma_{\phi,\varphi}(h) = \sum_{u=1}^{N_s} \gamma_{\phi,\varphi}^u(h) = \sum_{u=1}^{N_s} b_{\phi,\varphi}^u \cdot s_{\phi,\varphi}^u(h), \quad (3)$$

where N_s is the number of the spatial scale, $b_{\phi,\varphi}^u$ are coefficients, and $s_{\phi,\varphi}^u(h)$ are variogram functions.

Consequently, to characterize the spatial structure of the variables used here, we introduced a four-step procedure to construct the variogram model.

Step 1. We computed the directional variograms for the mean annual precipitation along the N-S and the W-E directions, using an angular tolerance of 90° .

Step 2. We computed the omnidirectional variograms for the mean annual precipitation.

Step 3. We plotted the nested variogram models. Nested models were defined as linear combinations of several basic models composed of one nugget effect and one Gaussian model (see the following equation):

$$r(h) = \begin{cases} 1.808 \times 10^6, & h = 0^\circ \\ 1.808 \times 10^6 + 7.095 \times 10^8 (1 - e^{-3h^2/158.7^2}), & h > 0^\circ. \end{cases} \quad (4)$$

Step 4. The entire set of variograms and cross-variograms were then fitted using the weighted least squares method and the same nested model above.

2.3. BME Methodology. The information used in BME includes two parts. Physical laws, statistics moments, and experts' experience form the general knowledge (G-KB). Sampled data (both hard data and soft data) consist of site-specific knowledge (S-KB). The BME approach is mainly based on the principle of maximum entropy and Bayes' theorem [13]. Under the constraint conditions from G-KB, the probability density function (pdf) that maximizes the information entropy is obtained, namely, prior pdf. And the posterior pdf is generated by applying prior pdf and S-KB to calculating conditional probability according to Bayes' theorem. The posterior pdf is the pdf we need to estimate

the probability distribution of meteorological factor. The formulations used in BME approach are as follows.

The joint cumulative distribution function (cdf) for the estimation location p_k and m hard data locations can be expressed as

$$F_x(\chi_1, \chi_2, \dots, \chi_m, \chi_k) = P(x_1 \leq \chi_1, x_2 \leq \chi_2, \dots, x_m \leq \chi_m, x_k \leq \chi_k) \quad (5)$$

and we also can obtain its corresponding pdf (i.e., prior pdf):

$$f_x(\chi_1, \dots, \chi_m, \chi_k) = \frac{\partial^{m+1} F_x(\chi_1, \dots, \chi_m, \chi_k)}{\partial \chi_1 \dots \partial \chi_m \partial \chi_k}. \quad (6)$$

Shannon's information entropy is simply expressed as follows:

$$\text{Info}(x_{\text{map}}) = -\log[f_G(x_{\text{map}})], \quad (7)$$

where $f_G(x_{\text{map}})$ has the same definition as (6), except that x is replaced by G . The mathematical expectation of $\text{Info}(x_{\text{map}})$ (i.e., Shannon entropy function) can be simply expressed as

$$E[\text{Info}(x_{\text{map}})] = -\int \log[f_G(x_{\text{map}})] f_G(x_{\text{map}}) dx_{\text{map}}. \quad (8)$$

The constraint conditions from G-KB are denoted as g_α . In the case of the present study, g_α functions include the variogram models. Using the Lagrange multipliers, the prior pdf which maximizes the information entropy is obtained:

$$f_G(x_{\text{map}}) = Z^{-1} \exp\left[\sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(x_{\text{map}})\right], \quad (9)$$

where $Z = \exp(-\mu_0)$ is a normalization constant; μ_α is Lagrange multiplier.

The stage that obtains prior pdf is called prior stage [13].

There are two kinds of sampled data χ_{data} according to their precision. Those data which are considered to be precise are so-called hard data. And those observations which are not precise as we need but are still valuable as supplementary information are named soft data. So that $\chi_{\text{data}} = [\chi_{\text{hard}}, \chi_{\text{soft}}]$. In this research, the hard data include annual precipitation at up to 40 meteorological stations. And the soft data include pdf fitted based on the analysis of long time sequences.

According to Bayes' theorem, the posterior pdf should be

$$f_K(\chi_k | \chi_{\text{data}}) = \frac{f_G(\chi_{\text{map}})}{f(\chi_{\text{data}})}, \quad (10)$$

where $f_K(\chi_k | \chi_{\text{data}})$ is the posterior pdf of annual precipitation.

Given the S-KB, which includes soft data, the posterior pdf should be

$$f_K(\chi_k) = \frac{\int_I f_G(\chi_{\text{map}}) f_S(\chi_{\text{soft}}) d\chi_{\text{soft}}}{\int_I f_G(\chi_{\text{data}}) f_S(\chi_{\text{soft}}) d\chi_{\text{soft}}}, \quad (11)$$

where $f_S(\chi_{\text{soft}})$ is the pdf of soft data and I is in the defined domain for $f_S(\chi_{\text{soft}})$.

2.4. Cross-Validation. The performances of four BME implementations, that is, univariate and multivariate BME with and without soft data, were evaluated and compared using cross-validation. The basic idea consists of reestimating mean annual precipitation at the locations of meteorological stations after removing it. The difference between the estimated value $X(p_k)$ and the corresponding measured value $X'(p_k)$ at the location p_k is the experimental error ε_k :

$$\varepsilon_k = X(p_k) - X'(p_k). \quad (12)$$

Thus, repeating this estimation for all hard locations with the size $n = 40$, the cross-validation statistics of mean error (ME) can be expressed as

$$\text{ME} = \frac{1}{n} \sum_{k=1}^n \varepsilon_k; \quad (13)$$

we also can obtain the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^n |\varepsilon_k|; \quad (14)$$

and the root mean square error (RMSE) can be calculated:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^n \varepsilon_k^2}. \quad (15)$$

The function has a determination coefficient (r) as a measure of the goodness of fit of the model:

$$r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}, \quad (16)$$

where $\text{Cov}(X, Y)$ represents the covariance between random variable X and Y , $\text{Var}(X)$ is the variance of variable X , and $\text{Var}(Y)$ is the variance of variable Y .

2.5. Flowchart of Methodology. The flowchart of the methodology applied in this study (Figure 4) is composed of preprocessing of original data, multivariate statistical analysis and variograms, soft pdfs used in the BME estimation, estimation of the spatial distribution of mean annual precipitation, and assessment of the performance of different BME method using cross-validation.

3. Results and Discussion

3.1. Numerical Results and Plots. Figure 5 shows density scaled histograms and kernel density estimates for annual precipitation, average temperature, and average water vapor pressure.

Figure 6 shows the cumulative distribution function of three meteorological factors.

Figure 7 shows the variograms for mean annual precipitation.

Figure 8 shows the entire set of variograms and cross-variograms for variables used in this study.

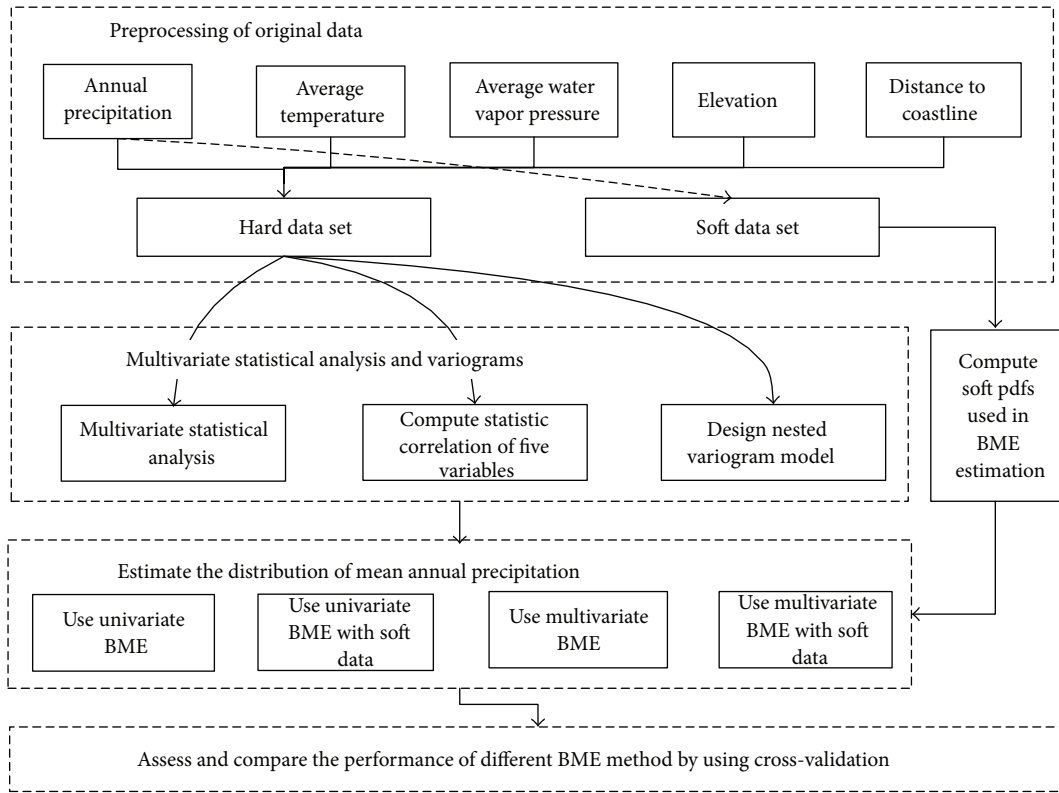


FIGURE 4: Flowchart of methodology.

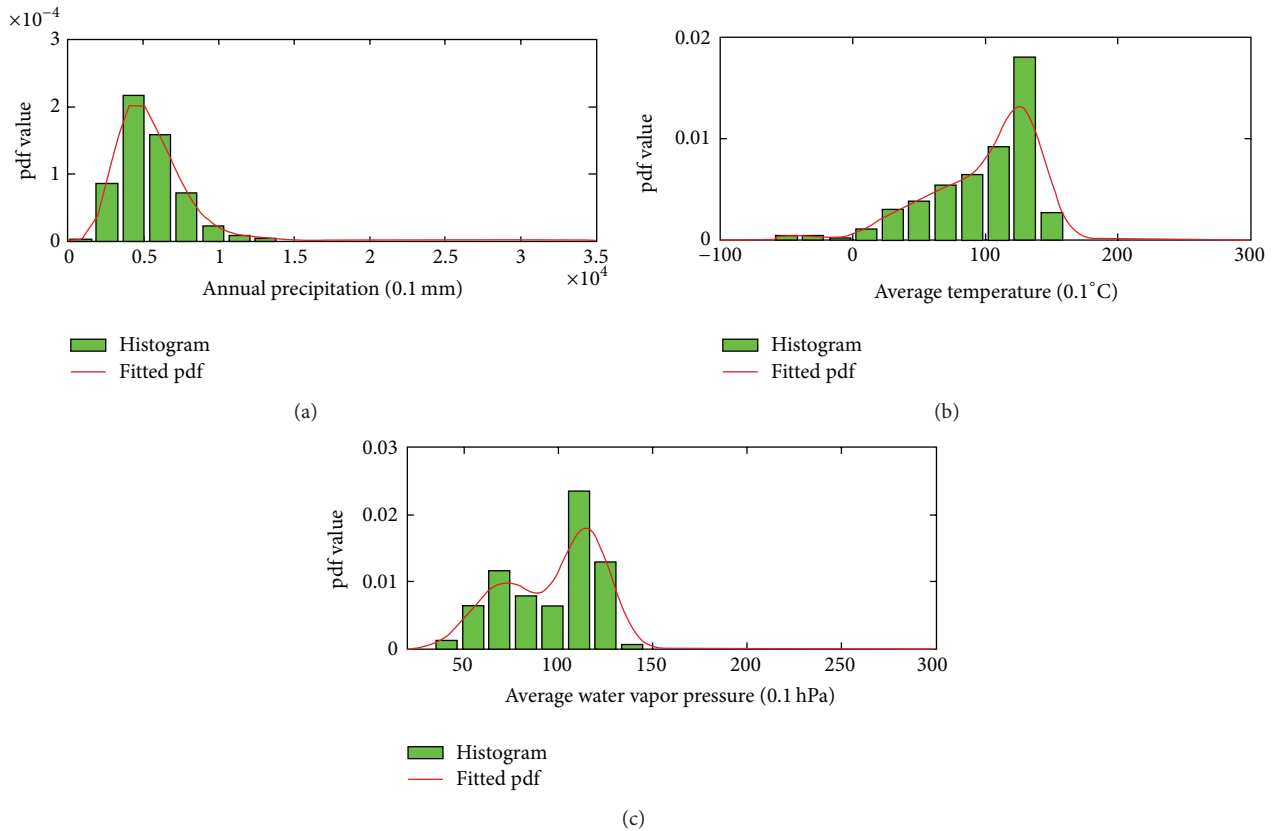


FIGURE 5: Density scaled histograms and kernel density estimates for three meteorological factors.

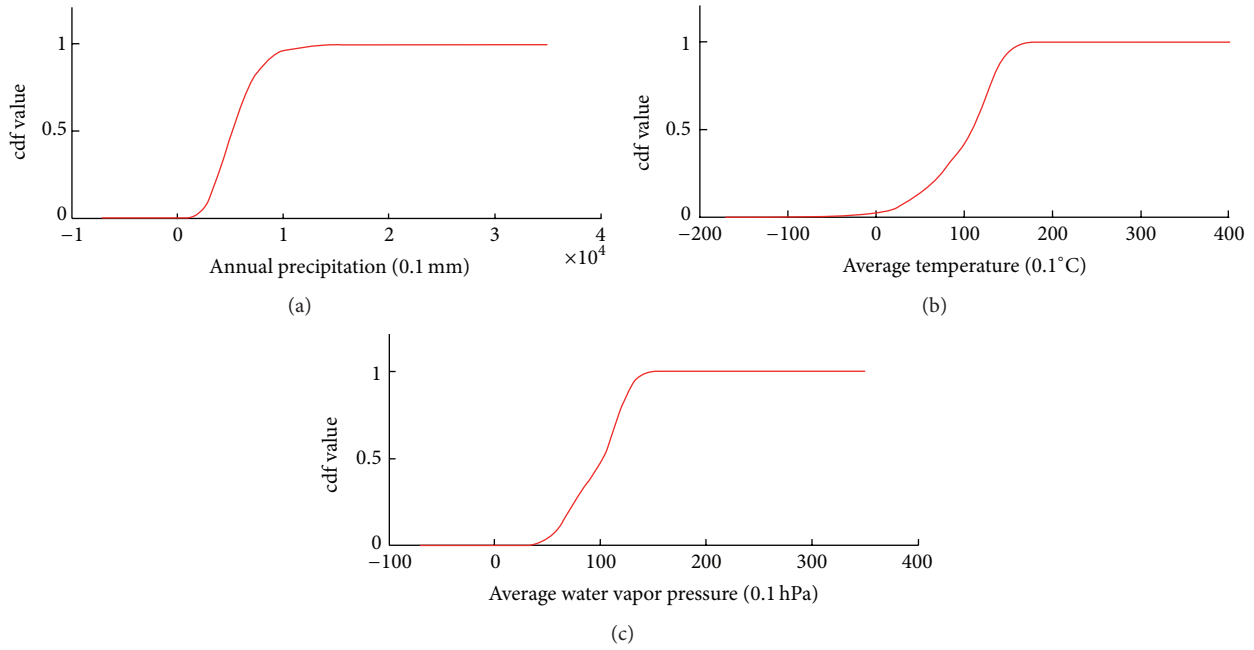


FIGURE 6: Cumulative distribution function of three meteorological factors.

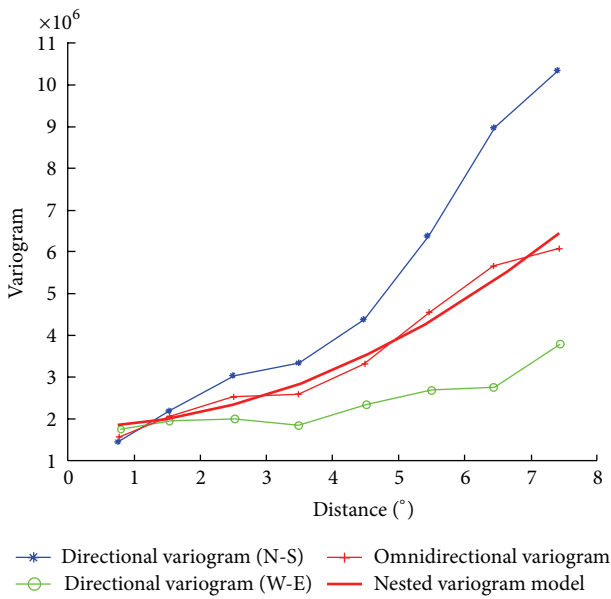


FIGURE 7: Variograms for mean annual precipitation.

Variables one through five represent annual precipitation, average temperature, average water vapor pressure, elevation, and distance to coastline, respectively.

Figure 9 shows the pdfs generated at these soft locations to have axisymmetric shapes. The shapes of soft pdfs were based on statistical analysis of 30-year data sequences (from 1983 to 2012) of meteorological factors and expert knowledge of the variability of the local distributions.

Figure 10 shows the values of correlation coefficients between each of the variables used here. The correlation plots

and correlation coefficients show strong correlations between each set of variables.

Variables one through five represent hard data related to annual precipitation, average temperature, average water vapor pressure, elevation, and distance to coastline, respectively, after zero mean and unit variance normalization.

Figure 11 shows the spatial distributions of mean annual precipitation in the studied area, obtained by univariate BME, univariate BME with soft data, multivariate BME, and multivariate BME with probabilistic soft data.

Spatial distributions of mean annual precipitation by using (a) univariate BME, (b) univariate BME with soft data, (c) multivariate BME, and (d) multivariate BME with probabilistic soft data.

Figure 12 presents the cross-validation results for the univariate BME, multivariate BME, univariate BME with soft data, and multivariate BME with probabilistic soft data.

Comparison of mean error (ME), mean absolute error (MAE), and root mean square error (RMSE) using these four methods for average annual precipitation.

Figure 13 presents the scatter plots of observed values versus estimated value for the average annual precipitation.

In order to assess the impact of removing an auxiliary variable on the estimated values of precipitation, we also perform cross-validation to test the performances of the multivariate BME with probabilistic soft data after removing it (Table 2).

3.2. Discussion. Researchers have studied the spatial distribution of precipitation in many ways, and a great deal of knowledge regarding the mechanisms related to the distribution has been reported in [1, 5, 7–9, 15–17], which are used in studies of climate, urban planning, flood disasters caused by

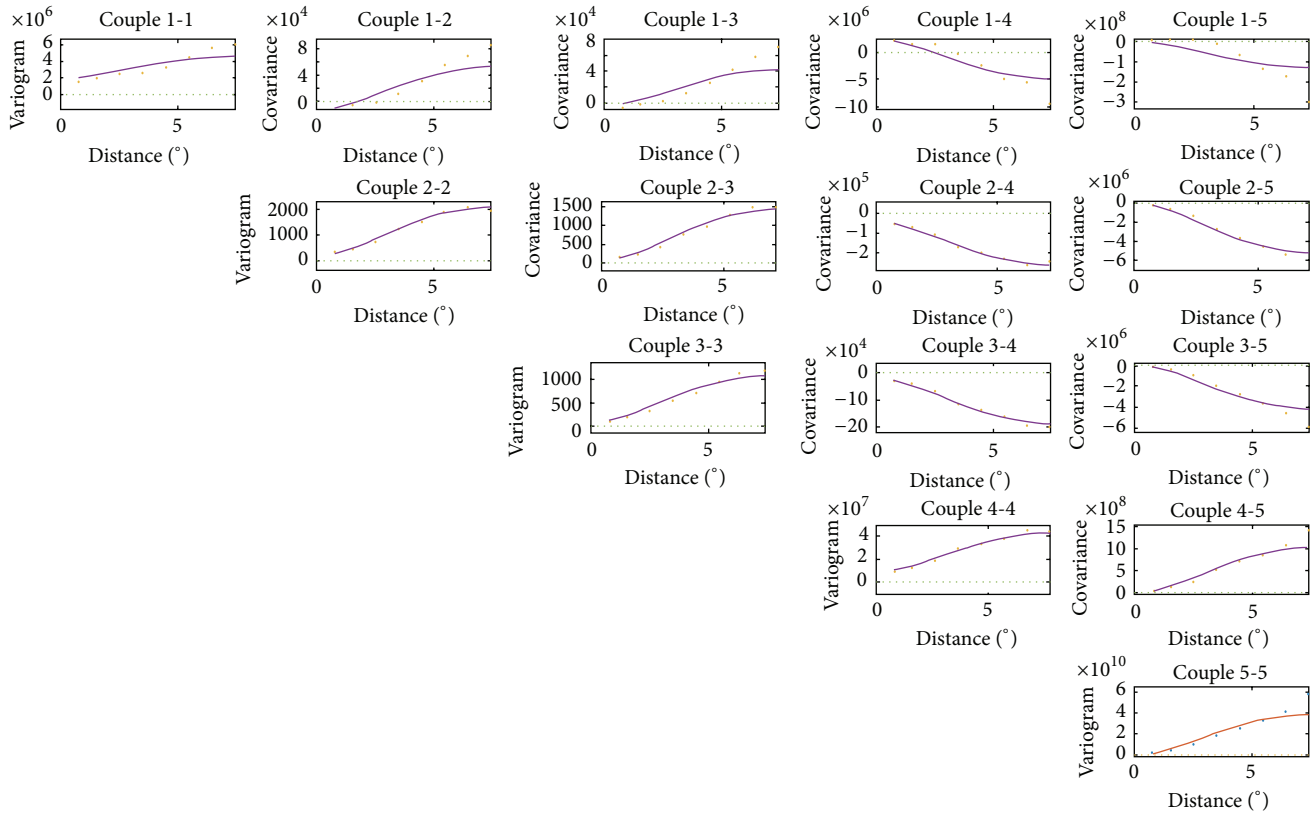


FIGURE 8: The entire set of variograms and cross variograms.

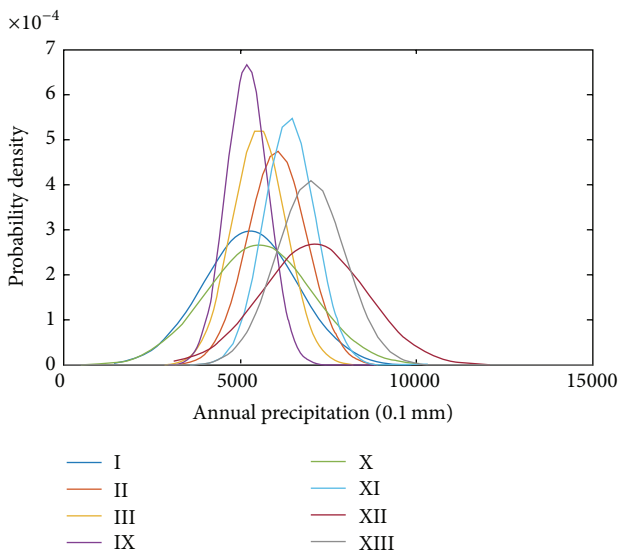


FIGURE 9: Soft pdfs used in Bayesian maximum entropy estimation.

heavy rain, and several other water-related issues. However, few scholars have studied the effects of the elevation of meteorological stations, the nearest distance to the coastline, and the statistical correlations between these two variables on the spatial distribution of precipitation in specific areas. Therefore, we investigated the distribution of mean annual

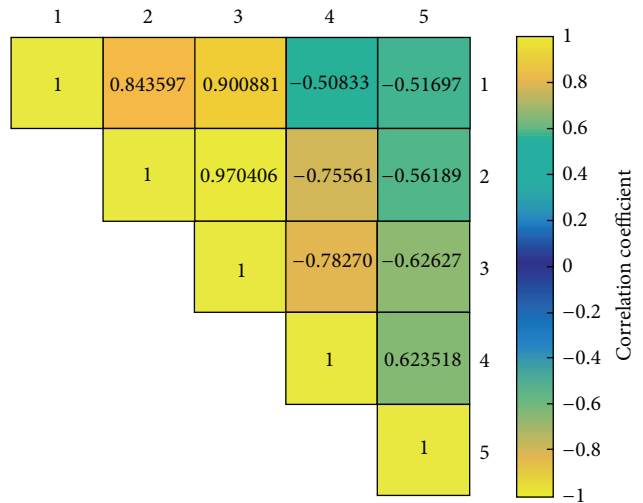


FIGURE 10: Correlation coefficients between the used variables.

precipitation over the study region (36°N–43°N, 113°E–120°E) with a focus on analyzing and spatial modeling of the sampled meteorological data and those two variables (elevation and distance to coastline) using the BME method.

Using four BME methods, that is, univariate BME, multivariate BME, univariate BME with soft data, and multivariate BME with probabilistic soft data, here we evaluated the spatial distribution of mean annual precipitation values over the

TABLE 2: Cross-validation results for the multivariate BME with probabilistic soft data after removing an auxiliary variable.

Results of cross-validation after removing a variable	Remove average water vapor pressure	Remove average temperature	Remove elevation	Remove distance to coastline
$ ME' $	30.0658	29.5423	28.2362	27.0232
MAE'	433.6511	426.5862	421.2383	419.9235
$RMSE'$	607.2357	603.2530	601.6712	599.7614

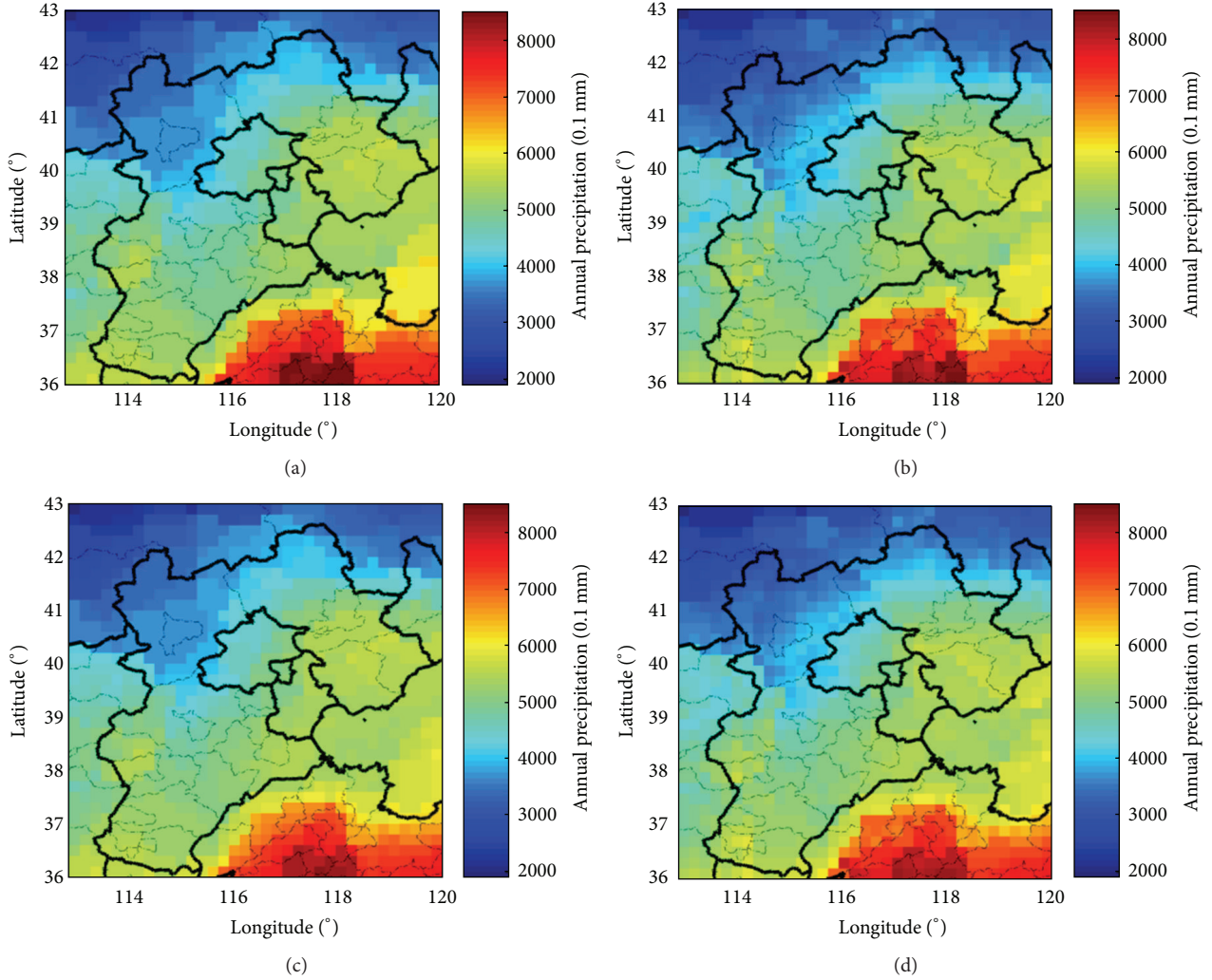


FIGURE 11: Spatial distribution of mean annual precipitation in the studied area.

study region defined above. The results showed that the southeastern region of the spatial distribution map, that is, Shandong Province, had the highest estimated values of mean annual precipitation, and the northwestern region, that is, northeast Shanxi Province, northern Hebei Province, and Inner Mongolia, had the lowest estimated values of mean annual precipitation (Figures 11(a)–11(d)). These results agree with the previous finding [18].

We found an inverse correlation between annual precipitation and the elevation of the meteorological stations. Locations with lower elevation had greater amounts of annual precipitation. In addition, with an increase in the distance

from the ocean, the correlation between mean annual precipitation and elevation decreased (Figure 8). In terms of statistic correlation and variogram, the precipitation-elevation relationship is similar to studies that employed other methods in analogous regions [5–8, 15]. However, the relatively small coefficient (-0.50833) suggests that the elevation variable in the studied region contributes relatively little to the spatial distribution of mean annual precipitation (Figure 10). The effect on the distribution of precipitation appears to be minor mainly because of the nature of the study region; specifically, plains dominate the study area with relatively low elevations and little change in topography and few areas

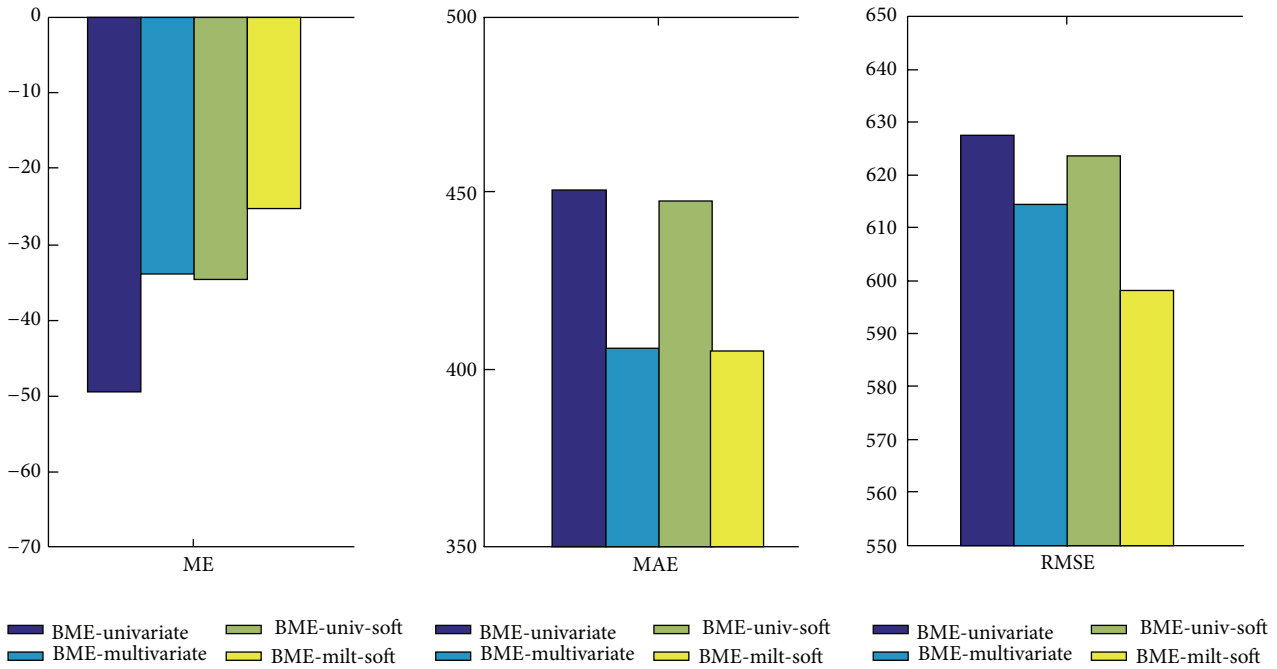


FIGURE 12: Results of cross-validation approach with the univariate Bayesian maximum entropy (BME), multivariate BME, univariate BME with soft data, and multivariate BME with probabilistic soft data.

of high mountains exist here. In addition, meteorological stations are generally established on the relatively flat plains of this region; few meteorological stations have been built on hilltops or hillsides.

An inverse correlation was observed between annual precipitation and the nearest distance to the coastline from each meteorological station. The maps showing the spatial distribution of precipitation acquired by the above four methods also verified the inverse relationship observed between mean annual precipitation and distance to the nearest coastline in the same latitude (Figure 11). From the perspective of the statistic correlation and variograms, the relationship between mean annual precipitation and the distance to the coastline is similar to findings of previous studies [1, 2, 9, 10, 16, 19]. Figure 11 also illustrates that variograms in the N-S direction and W-E direction have significant spatial heterogeneity, and the distributive differences of estimated precipitation in N-S direction are more outstanding than the differences in W-E direction.

Previous studies [18, 20] have shown that meteorological factors are not the only suitable regionalized variables that can be used to estimate annual precipitation rates, but they are correlated with other variables; if these other variables are not taken into consideration, more accurate estimates cannot be made. Among all studied meteorological variables, correlation coefficients between the variables used here (Figure 10) show that average water vapor pressure contributes more to the annual precipitation in the studied area compared to any other variables. Average water vapor pressure was closely correlated with annual precipitation, with a positive correlation between average water vapor pressure and annual precipitation. The second significant variable was average

temperature. The relatively smaller coefficient of average temperature when compared to average water vapor pressure suggests that average temperature in the area contributes relatively little to annual precipitation. Overall, for all variables in this study, the descending order of correlation coefficients to annual precipitation was average water vapor pressure > average temperature > elevation > distance to coastline. In addition, after removing the average water vapor pressure, the assessment results of $|ME'|$, MAE' , and $RMSE'$ were all greater than the case of removing other variables (see Table 2). In contrast, the assessment results of $|ME'|$, MAE' , and $RMSE'$ are all smaller than the case of removing other variables after removing the distance to coastline (also see Table 2). This provides further evidence that the relatively smaller coefficient of used variable in the studied area contributes relatively little to annual precipitation.

In this paper, we needed to integrate the correlation information of the above variables so as to gain more accurate estimates from the sites where fewer data were available from areas with sparse data condition sites. Based upon the BME framework, the univariate BME, multivariate BME, univariate BME with soft data, and multivariate BME with soft data analysis methods were compared, as summarized in the next paragraph.

The ME estimators determined by multivariate BME with probabilistic soft data were the closest to zero (Figure 12). The ME fluctuated because of the offset of the positive and negative errors. In addition, both MAE and RMSE estimators that had been determined by multivariate BME with probabilistic soft data provided the minimum values. The scatter plot of observed values and estimated values (Figure 13) shows the good performance of BME methods. And according to

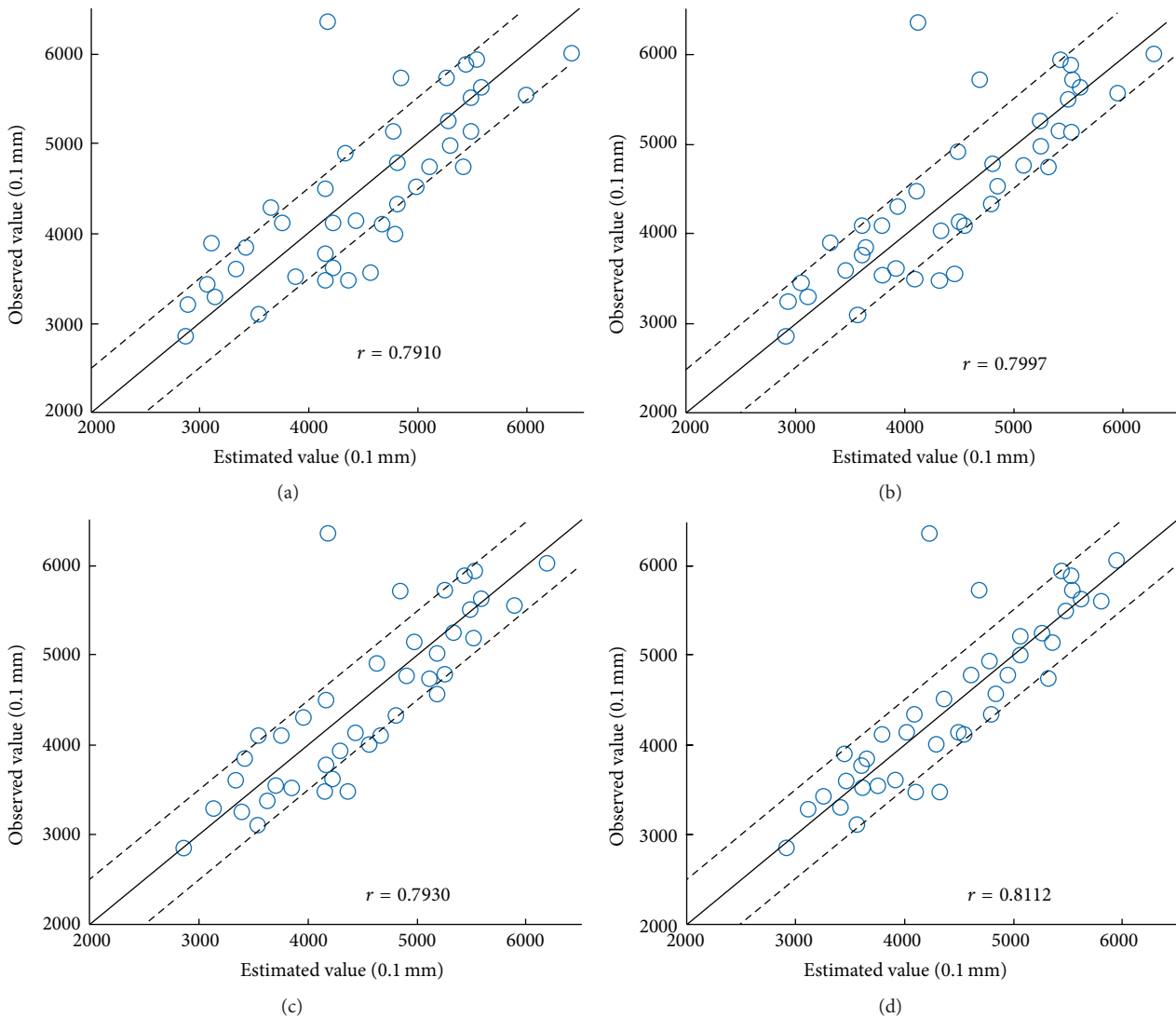


FIGURE 13: Scatter plots of observed values versus estimated value for the average annual precipitation. Scattergrams driven by using (a) univariate BME, (b) multivariate BME, (c) univariate BME with soft data, and (d) multivariate BME with probabilistic soft data.

the determination coefficients, multivariate BME with probabilistic soft data outdid the other methods. The above analysis indicates that the multivariate BME methods (multivariate BME and multivariate BME with probabilistic soft data) were more accurate than univariate BME methods (univariate BME and univariate BME with soft data), and the BME with probabilistic soft data methods were more accurate than BME without using soft data methods. In all the cases, the best results were produced by multivariate BME with probabilistic soft data, which produced better results than those produced using multivariate BME, univariate BME, and univariate BME with soft data.

Based on the above discussion, the main advantages of the multivariate BME with probabilistic soft data can be summarized as follows: (1) integration of spatial correlation information of the used variables into interpolation process, (2) assimilation of uncertain observations (soft data) into BME interpolation without hardening, and (3) lack of any

required assumption regarding the shape of any of the underlying probability distributions (non-Gaussian laws are automatically incorporated). All these advantages facilitate the process of spatially mapping precipitation accurately. However, approach must be improved and supplemented in more applications with verification and development.

Multivariate BME approach can assimilate spatial correlation information of the used variables into interpolation process but could not utilize uncertain observations. On the contrary, univariate BME with soft data can assimilate soft data, but spatial correlation information of the used variables cannot be used for improving the interpolation accuracy in this method. In addition, compared with other three BME methods, univariate BME neither assimilates uncertain observations nor integrates various prior knowledge bases. And it has been shown to be less accurate than the other three methods.

4. Conclusion

To explore the mean annual precipitation values over the current study region in northeast China including the Beijing area, the univariate BME, multivariate BME, univariate BME with soft data, and multivariate BME with probabilistic soft data were used and compared, covering five variables, that is, annual precipitation, average temperature, average water vapor pressure, elevation, and distance to coastline. The extendibility of BME allowed the amalgamation of prior knowledge and soft data, which enhanced the accuracy of the estimation process. Correlations between five variables provided in the multivariate BME method rendered the results even more accurate. Using cross-validation tests, multivariate BME with probabilistic soft data was found to be the best method for estimating the spatial distribution of mean annual precipitation in the studied area. The map showing the spatial distribution of precipitation showed that southeastern portion of the study area, that is, Shandong Province, had the highest estimated values of mean annual precipitation, while the lowest were found in the northwestern part of the study area, that is, northeast Shanxi Province, northern Hebei Province, and Inner Mongolia. These results demonstrate that BME is a promising tool that can be used for the estimation of the spatial distribution of average annual precipitation; it also can be used in other areas of interest or with other meteorological factors after further development of this approach to provide support for studies of climate, flood disasters caused by heavy rain, and several other water-related issues.

This study has limitations that should be addressed. First, the versatility of BME approach has yet to be verified. Some of the problems associated with BME may be overcome by extending the study to other research areas and studying additional types of meteorological factors. Second, we plan to obtain more field-based meteorological data using mobile weather stations to improve the accuracy of spatial estimation. Third, the spatiotemporal changes of the distribution of mean average precipitation were found to vary over time by extending the BME model from the current spatial dimensions to space-time dimensions and by analyzing the laws associated with this change.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

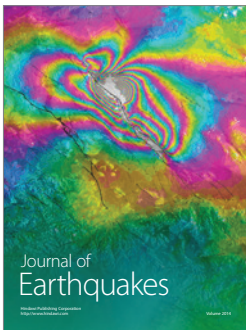
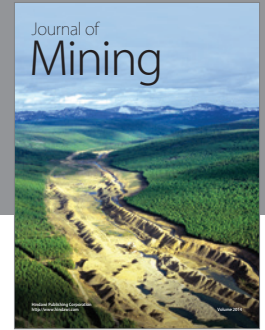
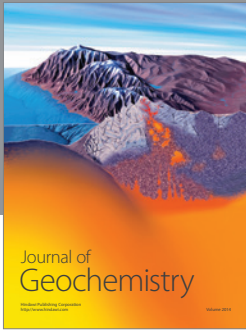
Acknowledgments

The authors thank National Natural Science Foundation of China for providing support (no. 91224004).

References

- [1] C. Daly, M. Halbleib, J. I. Smith et al., "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States," *International Journal of Climatology*, vol. 28, no. 15, pp. 2031–2064, 2008.
- [2] P. A. Bostan and Z. Akyürek, "Exploring the mean annual precipitation and temperature values over Turkey by using environmental variables," in *Proceedings of the ISPRS Joint Workshop of Visualization and Exploration of Geospatial Data*, University of Applied Sciences, Stuttgart, Germany, June 2007.
- [3] F. Liao, Y. Hong C, and G. Zheng, *Overview of Effects of Topography on Precipitation Study*, (Conference Paper), Chinese Meteorological Society, Guangzhou, China, 2007.
- [4] B. P. Fu, "Influence of topography and elevation on precipitation," *Journal of Geographical Sciences*, no. 4, pp. 302–314, 1992.
- [5] B. Johansson and D. Chen, "The influence of wind and topography on precipitation distribution in Sweden: statistical analysis and modelling," *International Journal of Climatology*, vol. 23, no. 12, pp. 1523–1535, 2003.
- [6] W. Ambach, W. Dansgaard, H. Eisner, and J. Møller, "The altitude effect on the isotopic composition of precipitation and glacier ice in the Alps," *Tellus*, vol. 20, no. 4, pp. 595–600, 1968.
- [7] M. A. Salles, P. O. Canziani, and R. H. Compagnucci, "Spatial variations in the average rainfall-altitude relationship in Great Britain: an approach using geographically weighted regression," *International Journal of Climatology*, vol. 21, no. 4, pp. 455–466, 2001.
- [8] H. W. Lull and L. Ellison, "Precipitation in relation to altitude in central Utah," *Ecology*, vol. 31, no. 3, pp. 479–484, 1950.
- [9] A. M. Makarieva, V. G. Gorshkov, and B.-L. Li, "Precipitation on land versus distance from the ocean: evidence for a forest pump of atmospheric moisture," *Ecological Complexity*, vol. 6, no. 3, pp. 302–307, 2009.
- [10] J. Marquinez, J. Lastra, and P. García, "Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis," *Journal of Hydrology*, vol. 270, no. 1-2, pp. 1–11, 2003.
- [11] H.-L. Yu, J.-C. Chen, G. Christakos, and M. Jerrett, "BME estimation of residential exposure to ambient PM10 and ozone at multiple time scales," *Environmental Health Perspectives*, vol. 117, no. 4, pp. 537–544, 2009.
- [12] A. Douaik, M. Van Meirvenne, and T. Tóth, "Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data," *Geoderma*, vol. 128, no. 3-4, pp. 234–248, 2005.
- [13] G. Christakos and X. Y. Li, "Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators?" *Mathematical Geology*, vol. 30, no. 4, pp. 435–462, 1998.
- [14] T. Tachikawa, M. Hato, M. Kaku, and A. Iwasaki, "Characteristics of ASTER GDEM version 2," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '11)*, pp. 3657–3660, IEEE, Vancouver, Canada, July 2011.
- [15] Z. Sokol and V. Bližňák, "Areal distribution and precipitation-altitude relationship of heavy short-term precipitation in the Czech Republic in the warm part of the year," *Atmospheric Research*, vol. 94, no. 4, pp. 652–662, 2009.
- [16] J. J. Attema and G. Lenderink, "The influence of the North Sea on coastal precipitation in the Netherlands in the present-day and future climate," *Climate Dynamics*, vol. 42, no. 1-2, pp. 505–519, 2014.
- [17] L. Gourcy, "Isotopic composition of precipitation in the Mediterranean Basin in relation to air circulation patterns and climate," IAEA-TECDOC 1453, 2005.
- [18] Z. H. Lin, X. G. Mo, H. X. Li et al., "Comparison of three spatial interpolation methods for climate variables in China," *Journal of Geographical Sciences*, vol. 57, no. 1, pp. 47–56, 2002.

- [19] C. Hellström and B. A. Malmgren, "Spatial analysis of extreme precipitation in Sweden 1961–2000," *Ambio*, vol. 33, no. 4-5, pp. 187–192, 2004.
- [20] K. E. Trenberth and J. M. Caron, "The southern oscillation revisited: sea level pressures, surface temperatures, and precipitation," *Journal of Climate*, vol. 13, no. 24, pp. 4358–4365, 2000.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

