

Research Article

Spatial Estimation of Losses Attributable to Meteorological Disasters in a Specific Area (105.0°E–115.0°E, 25°N–35°N) Using Bayesian Maximum Entropy and Partial Least Squares Regression

F. S. Zhang, S. B. Zhong, Z. T. Yang, C. Sun, C. L. Wang, and Q. Y. Huang

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

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

Received 30 October 2015; Accepted 9 February 2016

Academic Editor: Herminia García Mozo

Copyright © 2016 F. S. 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.

The spatial mapping of losses attributable to such disasters is now well established as a means of describing the spatial patterns of disaster risk, and it has been shown to be suitable for many types of major meteorological disasters. However, few studies have been carried out by developing a regression model to estimate the effects of the spatial distribution of meteorological factors on losses associated with meteorological disasters. In this study, the proposed approach is capable of the following: (a) estimating the spatial distributions of seven meteorological factors using Bayesian maximum entropy, (b) identifying the four mapping methods used in this research with the best performance based on the cross validation, and (c) establishing a fitted model between the PLS components and disaster losses information using partial least squares regression within a specific research area. The results showed the following: (a) best mapping results were produced by multivariate Bayesian maximum entropy with probabilistic soft data; (b) the regression model using three PLS components, extracted from seven meteorological factors by PLS method, was the most predictive by means of *PRESS/SS* test; (c) northern Hunan Province sustains the most damage, and southeastern Gansu Province and western Guizhou Province sustained the least.

1. Introduction

Meteorological disasters seriously threaten human life and property. Risk mapping methods in describing the spatial distribution of meteorological disaster risk have been applied to a range of major meteorological disasters, including floods and other water-based disasters [1, 2], drought [3, 4], wind [5], freezing rain and snow disaster [6], and tropical cyclones [7, 8]. A more reliable and efficient comprehensive form of risk assessment could improve emergency responses and allow people to evacuate and take protective measures before a disaster [9] and provide support for urban planning [10]. Meteorological factors are considered particularly important, because both meteorological disasters and secondary disasters are profoundly affected by meteorological factors: excessive precipitation is the main cause of flooding and other water-based disasters, abnormal temperatures contribute to

droughts and cold disasters, and strong winds contribute to gales. Losses attributable to meteorological disasters have a strong connection to meteorological factors, whose values can be estimated from the spatial and temporal distribution of those factors. In this way, research into the link between meteorological factors and losses associated with meteorological disasters is much needed.

Scholars determine the spatiotemporal distribution of meteorological factors using different techniques, such as linear or nonlinear regression [11–13], artificial neural networks [14, 15], and several kriging methods [16, 17]. At present, ground-based meteorological observation systems cannot provide explicit information, mainly because of the complex topography in the study areas, as well as the sparsity of meteorological stations [18]. In this way, spatial interpolation techniques are essential to the mapping of meteorological factors, and values at unmeasured locations can be estimated

based on existed observations. In addition, these studies have shown that meteorological factors are not only regionalized variables but also correlated with other variables, such as other meteorological factors, longitude, latitude, elevation, and slope [19, 20]. If these variables are not taken into consideration, accurate estimates cannot be made. Bayesian maximum entropy (BME) approach provides a numerical estimation framework for spatiotemporal analysis and mapping [21–23]. Some BME applications [24, 25] have been shown to be more accurate than previous spatial interpolation techniques, and all these advantages facilitate the process of spatially mapping multiple meteorological factors accurately.

In recent years, a large number of researchers have tried to utilize different theories to analyze meteorological disasters and their losses and have made great progress in the study of the prediction, prevention, and other matters related to disastrous meteorological events. Some works have used Grey systems theory to analyze the dynamic situation and future trends of disaster risk [8, 26–28]. Besides, 3S technology, that is, RS, GPS, and GIS, has been widely applied to disaster warning, monitoring, and integrated risk assessment [29–31]. Some regression methods have been used for building the relationship between hazard factors and disaster losses after disasters [32, 33].

These methods are often limited to analysis of the values of potential losses or impacts, including casualties, economic losses, resettled persons, and regression prediction models of disaster losses induced by single meteorological disaster. When using these methods, it is difficult for researchers to ensure the exact probability distribution of the disastrous meteorological events without sufficiently large numbers of sampling data; it is also difficult for them to evaluate whether or not the final assessment is reasonable.

Few scholars have evaluated the exact relationship between the distribution of multivariate meteorological factors and its historical disaster loss information for a specific area. However, traditional multivariate regression requires independent input variables, and the regression model obtained from this traditional method is unstable due to the strong correlation between meteorological factors [34]. Data dimension reduction techniques, such as partial least square (PLS), offer an approach to reduce multicorrelated variables to linearly independent components, and a reasonable regression model can be established using only a few components, which are able to explain most information in meteorological factors [35].

The following issues need to be resolved in this study: (1) The current sparse meteorological data from the observation station show that the accuracy of spatial distribution estimation of meteorological factors must be improved; (2) the relationship between the spatial distribution of mean annual values for multiple meteorological factors and the losses of meteorological disasters in the studied area needs to be made clear for the studied area; (3) for the sake of providing a scientific basis for meteorological disaster prevention and urban planning, the contributions of each meteorological factor to the disaster-induced losses and their specific causes need to be analyzed.

The purposes of this BME-PLS method are to (1) estimate the mean annual distribution (1951–2012) in the studied area of seven meteorological factors through the integration of prior knowledge and statistical correlation between meteorological factors and soft data using BME method, (2) compare the performance of four mapping methods (kriging, cokriging, and univariate and multivariate BME with probabilistic soft data) and identify the method with the best performance as indicated by the final distributions, (3) extract linearly independent PLS components from a sampling data set of meteorological factors using PLS regression, establish a fitted model between the PLS components and disaster loss information (including droughts, heavy-rain floods, high-temperature disasters, and frost disasters), and analyze the efficiency of the fitting, and (4) present a distribution map of regional integrated disaster loss. This methodology is useful in the evaluation of synthetic losses of any concerned meteorological disasters in any region of interest, and a loss risk map can be an important reference for disaster prevention and urban planning.

2. Material and Methods

2.1. Methodology. The main contents of the BME-PLS approach used in this research are composed of three processes, that is, the area of interest and preprocessing of original meteorological data, BME methodology, and partial least squares regression method (Figure 1). BME methodology was used to acquire the spatial distribution of seven meteorological factors, including average annual precipitation, average temperature, average relative humidity, average barometric pressure, average annual sunshine duration, average water vapor pressure, and average wind speed. A regression model was developed using PLS to estimate the effects of the spatial distribution of these factors on losses associated with meteorological disasters in a specific area to both illustrate the methodology and demonstrate disaster loss mapping.

2.2. Area of Interest and Preprocessing of Original Meteorological Data. The study area is located within 105.0°E–115.0°E, 25°N–35°N. This area has both temperate monsoon climate and subtropical monsoon climate, which are approximately divided by the Qinling Mountains-Huai River. More than 300,000,000 people live and work in the area, mainly within the urban agglomeration. There are also several villages with populations ranging from a hundred to several thousand inhabitants. The meteorological disasters that frequently take place in this area have had a negative impact on transportation, industrial production, and daily life. The four types of disasters that cause the most serious losses are droughts, heavy-rain floods, high-temperature disasters, and frost disasters [36]. The geographical location of the studied area and the locations of 121 meteorological stations in this area are shown in Figure 2.

Hard data locations are marked with triangles (Δ) and soft data locations are marked with circles (\circ).

Data set used in this research includes 121 sampling locations (i.e., meteorological stations), which consist of

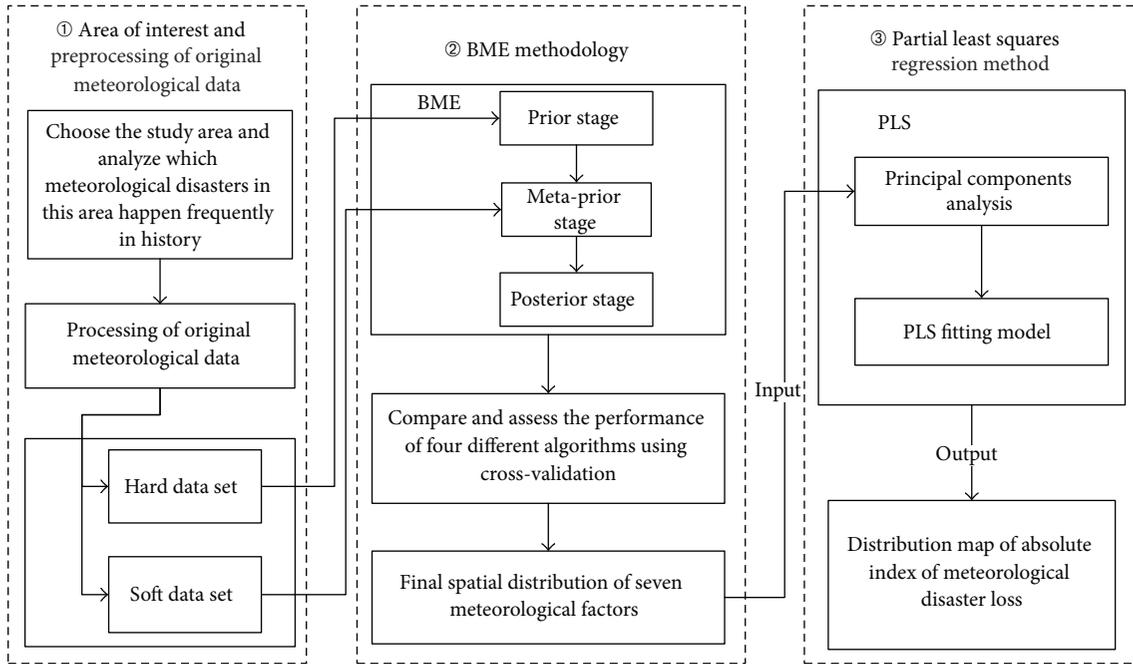


FIGURE 1: Flow chart of BME-PLS methodology.

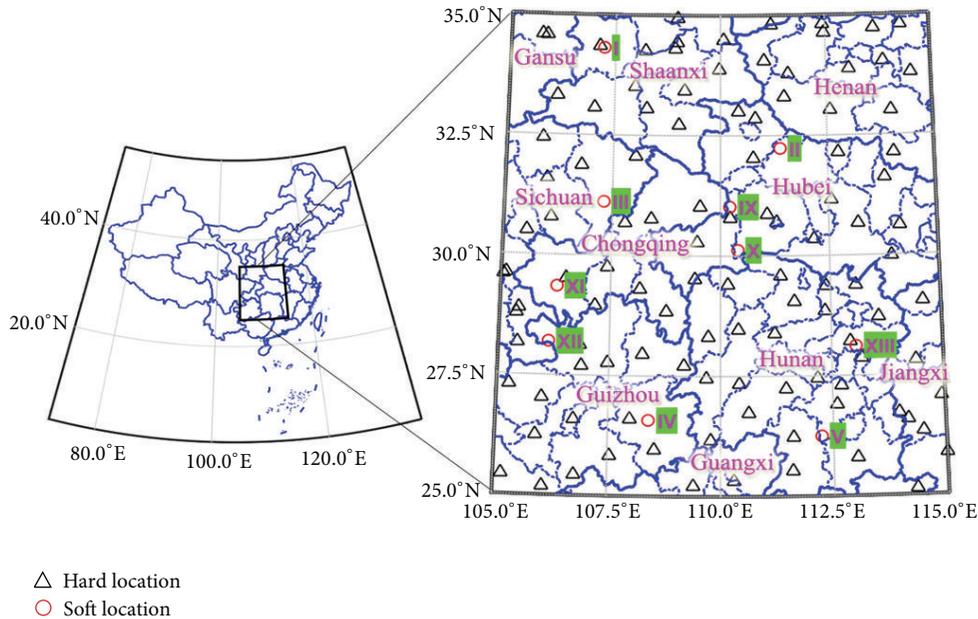


FIGURE 2: Geographic location of the studied area and locations of 121 meteorological stations in this area.

60-year-average values of each meteorological factor per location along with their coordinates. Figure 3 shows the flow chart of the processing of original meteorological data.

At 111 of these locations, so-called “true” values of meteorological observed data were available. These are considered exact data; that is, at these locations the sampling data were deemed to have no measurement errors. Therefore, we called these stations as “hard data stations.”

At the rest of 10 sampling locations, their available data consisted of the historical statistical laws of each meteorological factor, and these locations are called “soft data stations.” In this way, the soft data within each sampling location describes the uncertainty using probabilistic soft data. These data represent the historical statistical laws of each meteorological factor reflecting the uncertainty around the accurate sampling data.

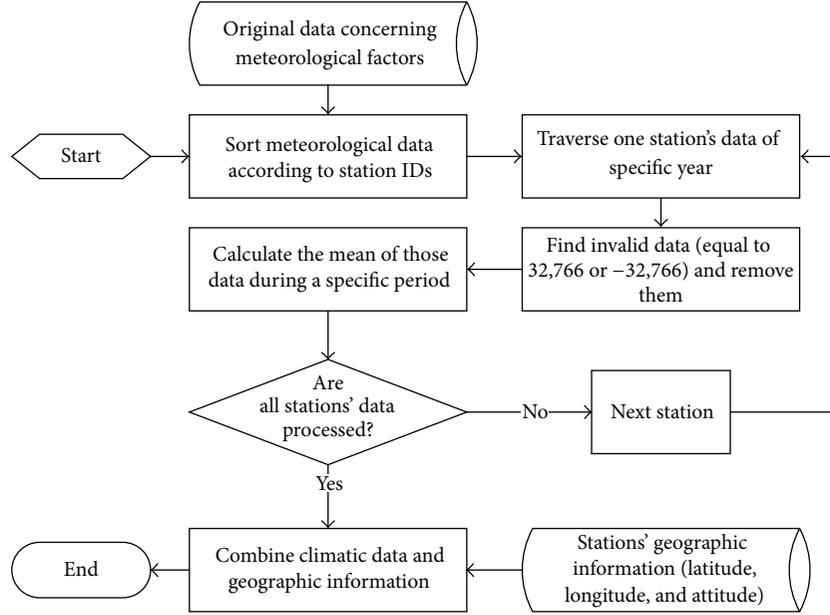


FIGURE 3: Flow chart of the processing of original meteorological data.

The soft data set (here, referred to as the soft probabilistic data) used in spatial mapping is based on the statistical analysis of long sequences of observed meteorological data available at the 10 locations mentioned above. These soft probabilistic data were rigorously incorporated into the BME mapping process presented in this research.

2.3. BME Methodology. The BME approach has been applied to a variety of studies with considerable success [18, 22, 25, 37–41]. We implement the BME estimation for the distribution of seven meteorological factors by using the BMElib suite of functions in Matlab [39].

The spatial distribution of average annual values of seven meteorological factors is represented as a spatial random field. The purpose of the present work was to estimate the values of the random field, $L(p_k)$, at a nonmeasuring location, p_k ($k \neq i = 1, \dots, m$), given data at hard locations p_i ($i = 1, \dots, m$). The knowledge used in this estimating process of BME approach can be divided into two aspects, that is, site-specific knowledge (S-KB) and general knowledge (G-KB) [40].

Present work primarily discusses the main steps of the BME methodology that are relevant to this study.

2.3.1. Prior Stage. The purpose of this step is to maximize information content using G-KB only [21, 25]. In this study, L_{map} is comprised of hard data vector L_{hard} , soft data vector L_{soft} , and the value to be estimated L_k : $L_{\text{map}} = [L_{\text{hard}}, L_{\text{soft}}, L_k]$.

Here, we use Shannon's information criterion [25, 42] as a measure of information:

$$\text{Info}(L_{\text{map}}) = -\log[f_G(L_{\text{map}})]. \quad (1)$$

Here, $f_G(L_{\text{map}})$ is the prior pdf acquired only by G-KB [25].

Shannon entropy function can be expressed as follows:

$$E[\text{Info}(L_{\text{map}})] = -\int \log[f_G(L_{\text{map}})] f_G(L_{\text{map}}) dL_{\text{map}}. \quad (2)$$

The following formula gives the physical constraints imposed by G-KB:

$$E[g_\alpha(L_{\text{map}})] = \int g_\alpha(L_{\text{map}}) f_G(L_{\text{map}}) dL_{\text{map}}, \quad (3)$$

$$\alpha = 0, \dots, N_c.$$

Here, $g_\alpha(L_{\text{map}})$ are functions chosen such that the G-KB is fully accounted for during the BME estimation process; and the expectations, $E[g_\alpha(L_{\text{map}})]$, provide the statistical moments of interest [25]. Equation (4) gives the prior multivariate pdf:

$$f_G(L_{\text{map}}) = A^{-1} \exp\left[\sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(L_{\text{map}})\right]. \quad (4)$$

Here, $A = \int \exp(\sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(L_{\text{map}})) dL_{\text{map}}$ is a normalization constant; μ_α are Lagrange multipliers [25].

2.3.2. Meta-Prior Stage. The probabilistic type of soft data can be represented as $S = \{L_{\text{soft}} : P_S[L_{\text{soft}} \leq \zeta] = F_S(\zeta) \text{ for } \zeta \in I\}$ [40]. In this study, assume that 108 exact meteorological stations are available at the locations $L_{\text{hard}} = (l_1, \dots, l_{107}, l_{108})$ and that 10 uncertain data sets are available at the locations $L_{\text{soft}} = (l_{109}, \dots, l_{117}, l_{118})$ (see Section 2.2 for details).

2.3.3. *Posterior Stage.* We can acquire the posterior pdfs of each meteorological factor based on the conditional probability law [25]:

$$\begin{aligned} f_K(L_k) &= f_G(L_k | L_{\text{hard}}, L_{\text{soft}}) \\ &= \frac{f_G(L_k, L_{\text{hard}}, L_{\text{soft}})}{f_G(L_{\text{hard}}, L_{\text{soft}})}. \end{aligned} \quad (5)$$

Here, $f_K(L_k)$ represents the posterior pdf; $f_G(L_k, L_{\text{hard}}, L_{\text{soft}})$ is the prior pdfs.

When soft data is expressed in the form of the pdfs, the posterior or BME pdf is produced:

$$f_K(L_k) = \frac{\int_I f_G(L_k, L_{\text{hard}}, L_{\text{soft}}) f_S(L_{\text{soft}}) dL_{\text{soft}}}{\int_I f_G(L_{\text{hard}}, L_{\text{soft}}) f_S(L_{\text{soft}}) dL_{\text{soft}}}. \quad (6)$$

Here, $f_S(L_{\text{soft}})$ is the pdf defined based on S-KB.

The estimated results of four different methods (i.e., kriging, cokriging, and univariate BME with probabilistic soft data and multivariate BME with probabilistic soft data) were evaluated and compared using cross-validation [18].

ε_i represents the experimental error between the estimated value Z and the corresponding sampling value Z_1 at location s_i , respectively:

$$\varepsilon_i = Z(s_i) - Z_1(s_i). \quad (7)$$

In this research, we repeat this estimation for the experimental data size $n = 10$.

The cross-validation statistics of mean error (ME) can be represented as follows:

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i. \quad (8)$$

The mean absolute error (MAE) is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\varepsilon_i|. \quad (9)$$

The root mean square error (RMSE) is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}. \quad (10)$$

2.4. *Partial Least Squares Regression Method.* This PLS method can be considered as a two-step regression technique, and its aim is to construct a linear relationship between the input variables and output variables [43]. Many previous works have described this regression method in detail [44–46]. The chief purpose of the following paragraphs is to introduce the main steps of this regression method used in this research.

In the first step, X (input variables) and Y (output variables) metrics are decomposed in some latent variables:

$$X = TP^t + E, \quad (11)$$

$$Y = UQ^t + F. \quad (12)$$

In (11), T is the score and P is the loading for the X matrix; similarly, U is the score and Q is the loading for the Y matrix in (12). The matrices E and F correspond to the residuals associated with the PLS modeling.

In the second step, we construct a linear inner relation linking between T and U :

$$U = TD + H, \quad (13)$$

where D represents the diagonal matrix; H is the residual matrix.

Finally, we can acquire the PLS regression by the following:

$$Y = XB + \varepsilon. \quad (14)$$

Here, B represents the matrix of regression coefficients:

$$B = X^t U (T^t X X^t U)^{-1} T^t Y. \quad (15)$$

ε is the residual matrix.

In this research, we used the Matlab (MathWorks) software suite to perform PLS regression. The input data set consisted of a 111×7 matrix of the sampling values of seven meteorological factors at 111 hard locations (7 variables). Let $X(111 \times 7)$ be the input matrix for PLS regression. The i th ($1 \leq i \leq 111$) row vector of X , denoted by x_i , consists of 7 variables which are the values of seven meteorological factors. Let Y be the (111×1) output matrix, which consisted of a 111×1 matrix of absolute index of meteorological disaster loss with 111 hard locations (1 variable). The j th ($1 \leq j \leq 111$) column vector y_j of Y is the output vector of which each variable corresponds to the attribute of stimulation, and j corresponds to the attribute of absolute index of meteorological disaster loss. Absolute index of meteorological disaster loss was used in this study to evaluate the regional degree of damage, calculated as follows:

$$C_i = \sum_{j=1}^n a_j \frac{X_{ij}}{\max\{X_{ij}\}_i}. \quad (16)$$

Here, i is the subarea of the studied area; j is the index of each disaster loss involved in this study, that is, disaster-affected population, death toll, evacuated population, area of farmland damaged by the disaster, area of crop failure, number of collapsed buildings, number of damaged buildings, and direct economic losses; C_i is the absolute index of meteorological disaster loss; X_{ij} is the index value of the j th disaster index in i th subarea; and a_j is the weights of these disaster indices. Here, higher values of absolute index of meteorological disaster loss represent greater losses due to meteorological disaster.

To evaluate the ability, it is justifiable to use the decrement of error as a criterion for choosing PLS components. This gives the researcher a measure of predictive power, the predictive residual error sum of squares (PRESS) [47]. When using k components to build the PLS regression model, PRESS_k is calculated as the sum of squares of the differences between the predicted result for each location (when it is left out of the PLS regression) and the sampling value of the

TABLE 1: Variable loadings of each PLS component.

PLS component	Meteorological factor						
	Annual precipitation	Average barometric pressure	Average wind speed	Average temperature	Average water vapor pressure	Average relative humidity	Annual sunshine duration
1	8.074	7.280	0.540	8.257	9.347	5.213	-2.254
2	-1.126	0.259	9.267	-4.785	-4.749	-4.795	8.002
3	0.593	2.181	-2.507	-2.146	-0.994	1.510	0.076

Meteorological factor values are normalized for zero mean and unit variance.

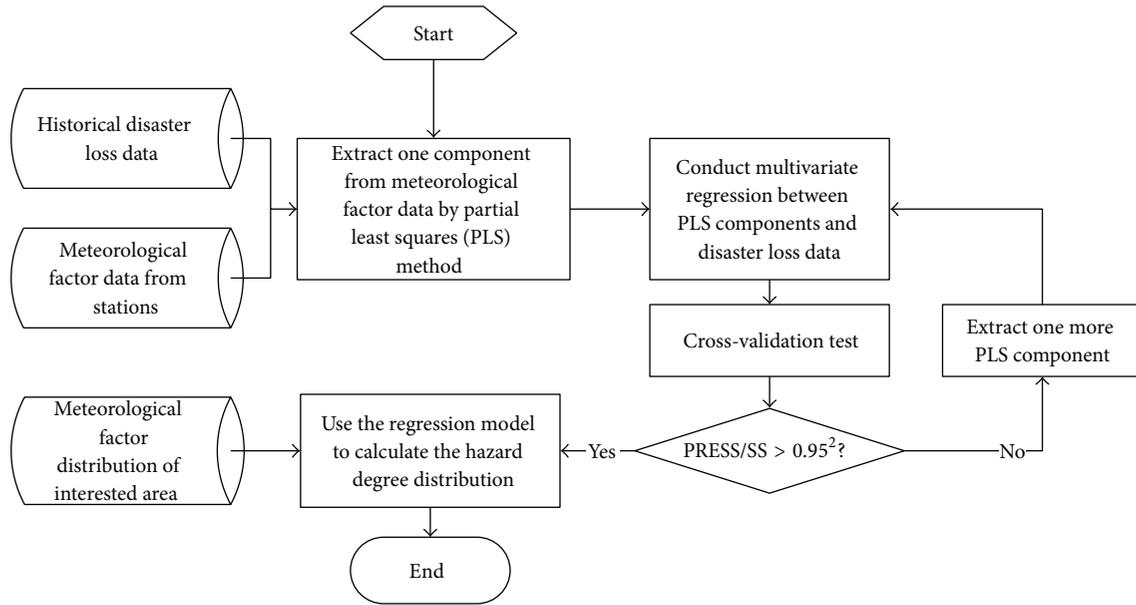


FIGURE 4: Flow chart of PLS.

dependent variables. Then, SS_{k-1} is calculated using $k - 1$ components and the whole data set to build the PLS regression model. SS_{k-1} denotes the error sum of squares (SS), similar to PRESS but leaving out no data during modeling. $PRESS_k/SS_{k-1}$ evaluates the predictive ability of the k th component in the model. If $PRESS_k/SS_{k-1} \leq 0.95^2$, the k th component makes a significant contribution to the predictive ability of the model [47], so the component will be used in building the model. Otherwise, no more components will be added to the model. Figure 4 shows a flow chart of PLS.

3. Numerical Results and Plots

Figure 5 shows the pdfs generated at these soft locations to have rectangle shapes. The shapes of soft pdfs were based on statistical analysis of long-time data sequences of meteorological factors.

Figure 6 presents the cross-validation results for the kriging, cokriging, and univariate BME with probabilistic soft data and multivariate BME with probabilistic soft data.

Comparison of ME, MAE, and RMSE uses these four methods for (a) annual precipitation, (b) average barometric pressure, (c) average wind speed, (d) average temperature, (e)

average water vapor pressure, (f) average relative humidity, and (g) annual sunshine duration.

Figure 7 shows the spatial distributions of seven meteorological factors in the studied area, obtained by multivariate BME with probabilistic soft data.

As presented in Figure 8(a), when using three PLS components in the regression model, the cross-validation result of $PRESS/SS$ was less than 0.95^2 , and those when using more components were greater than 0.95^2 .

Figure 8(b) shows that the first PLS components captured approximately 36.5% of the total variance, and the first and second PLS components captured approximately 62.6% of the total variance, and the first through third PLS components captured approximately 79.7% of total variance.

Table 1 shows the loadings of each meteorological factor in each PLS component. All these meteorological factors had similar loadings in the first PLS component except average wind speed and annual sunshine duration, which had small loadings. In the second PLS component, average wind speed and annual sunshine duration had the largest loadings, and average temperature, average wind speed, and average relative humidity had negative loadings. In the third PLS component, all the factors had relatively small loadings.

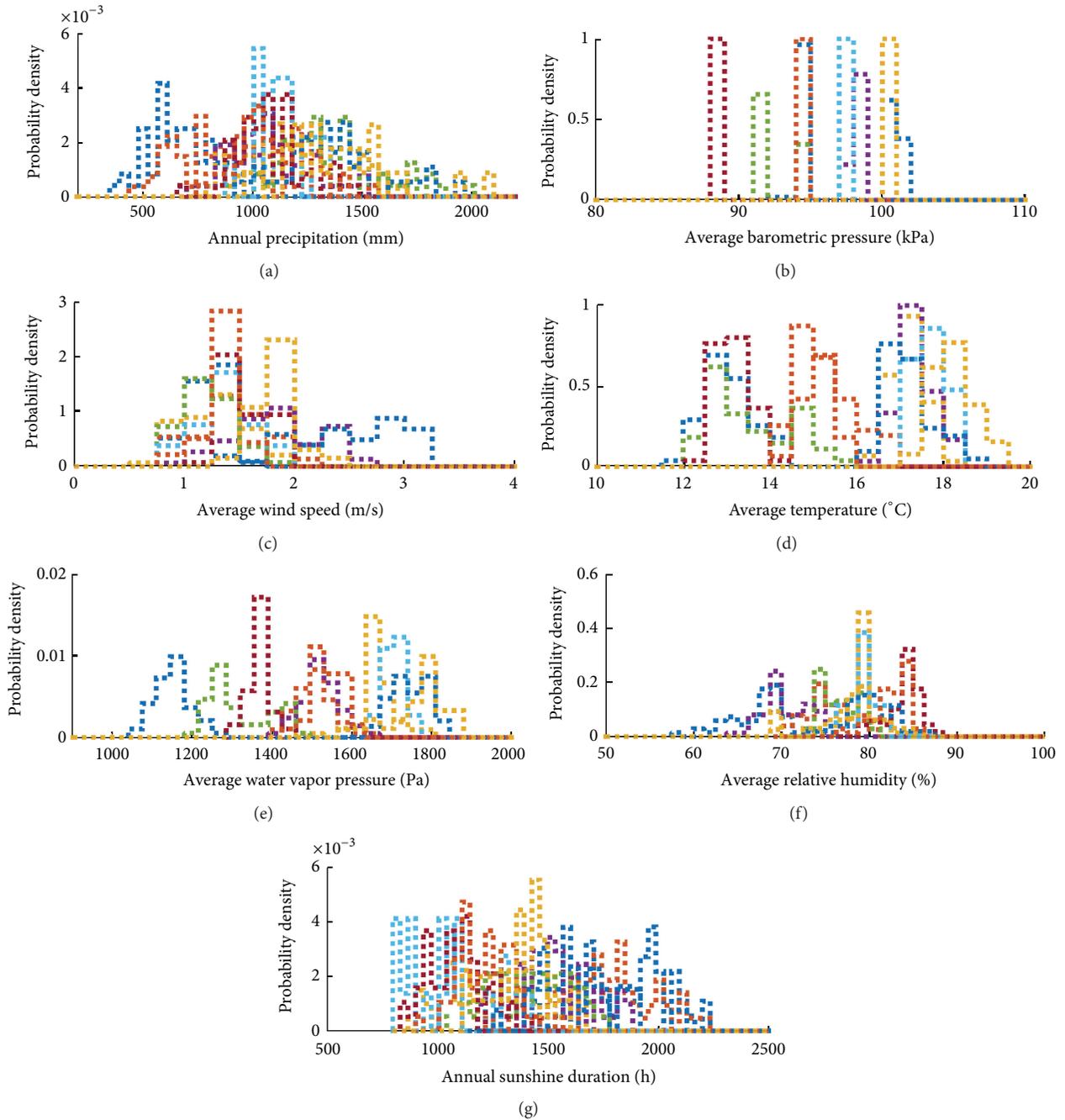


FIGURE 5: Soft pdfs used in BME estimation.

Figure 9 shows the correlation between each meteorological factor. The correlation plots (Figure 9(a)) and correlation coefficients (Figure 9(b)) show strong correlations between each set of meteorological factors except for the third, average wind speed.

Variables 1 through 7 represent hard data concerning annual precipitation, average barometric pressure, average wind speed, average temperature, average water vapor pressure, average relative humidity, and annual sunshine duration, respectively, after zero-mean-and-unit-variance

normalization. Figure 9(a) shows the probability distribution histogram and correlation plots of seven meteorological factors after zero-mean-and-unit-variance normalization. Figure 9(b) shows the correlation coefficients between seven meteorological factors.

Figure 10 shows the spatial distributions of the first three PLS components in the studied area. Figure 10(a) presents the distribution of these components, extracted from hard data of seven meteorological factors. With the spatial distributions of those factors, obtained with multivariate BME method,

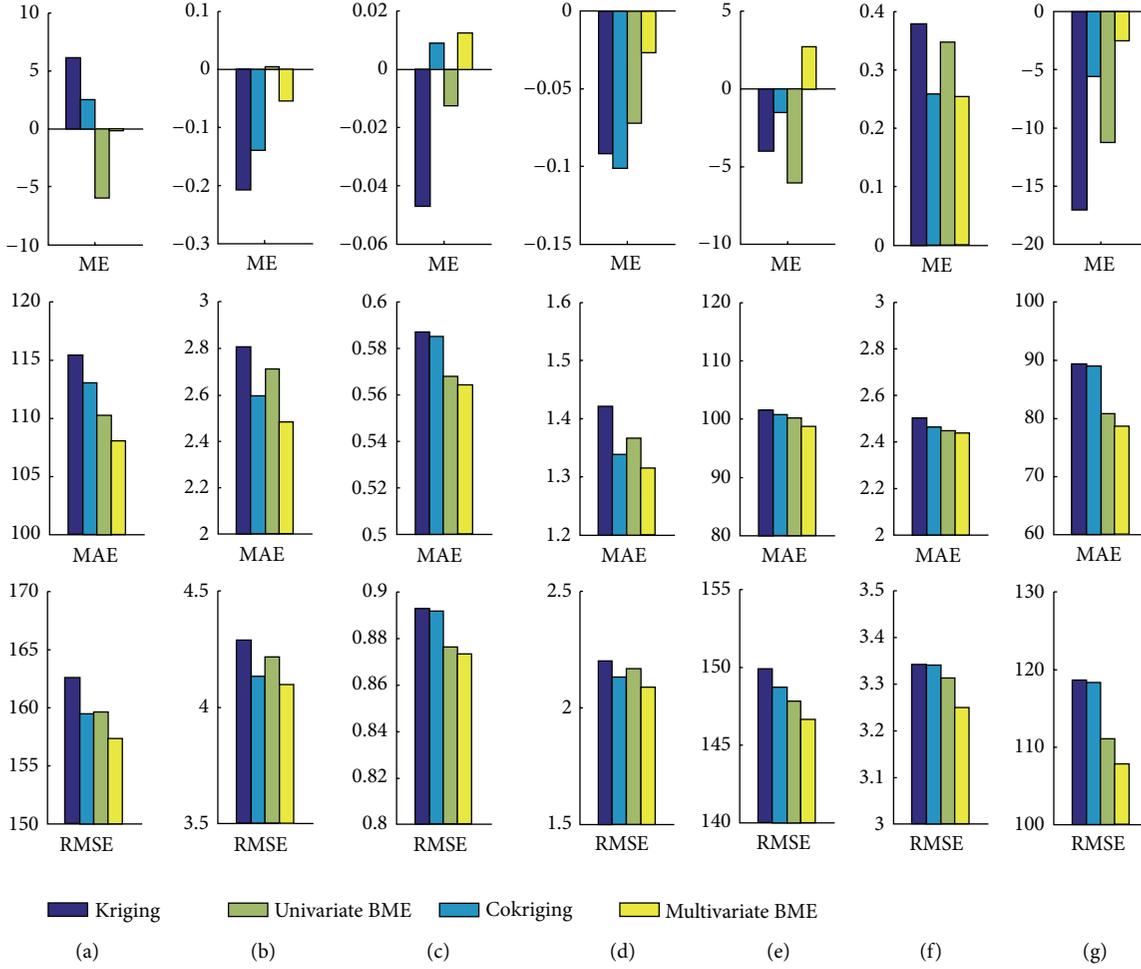


FIGURE 6: Results of cross-validation approach with kriging, cokriging, and univariate and multivariate BME with probabilistic soft data for seven meteorological factors.

and the loadings of each factor, the explicit spatial distributions of those components were obtained, as presented in Figure 10(b).

The regression model obtained by PLS regression is as follows:

$$\begin{aligned} \text{loss} &= 48.62 + 0.56\text{comp}_1 + 0.71\text{comp}_2 + 2.11\text{comp}_3 \\ &= 48.62 + 5.02f_1 + 8.90f_2 + 1.54f_3 - 3.24f_4 \\ &\quad - 0.17f_5 + 2.76f_6 + 4.53f_7. \end{aligned} \quad (17)$$

Here, comp_k ($k = 1, 2, 3$) is the k th PLS component and f_n ($n = 1, 2, \dots, 7$) is the zero-mean-and-unit-variance normalized value of the n th meteorological factor. And the 95% confidence intervals for each coefficient estimate are (46.91, 50.34), (3.85, 6.19), (6.16, 11.64), (0.71, 2.36), (-5.09, -1.40), (-0.85, 0.51), (1.59, 3.92), and (3.48, 5.57). Figure 11(a) presents the residuals of the predictions of PLS regression model, and Figure 11(b) presents the correlation plot of these predictions. The correlation coefficient r between the predicted output and measured output is 0.77.

Figure 12 shows the distribution of meteorological losses risk in studied area. In the middle of the area, which is in

northern Hunan Province, the risk of meteorological losses was greatest. It was lowest in the northwest and the southwest parts of the study area, which are in southeastern Gansu Province and western Guizhou Province, respectively.

4. Discussions

The spatial distributions of different meteorological factors and the risk of meteorological disasters have been studied in many ways, and a great deal of knowledge regarding the mechanisms underlying disasters has been reported, including ways of developing risk maps, which are used in disaster prevention, emergency resource scheduling, and urban planning. However, few scholars have studied the effects of the spatial distribution of multivariate meteorological factors on losses related to meteorological disasters in specific areas. In this study, the following issues were resolved: (1) the BME method in the BME-PLS model can be used to produce an accurate spatial distribution map of seven meteorological factors in the case of current sparse meteorological observation system and to provide inputs for PLS fitting model; (2) the PLS method in the BME-PLS model can be used to build a

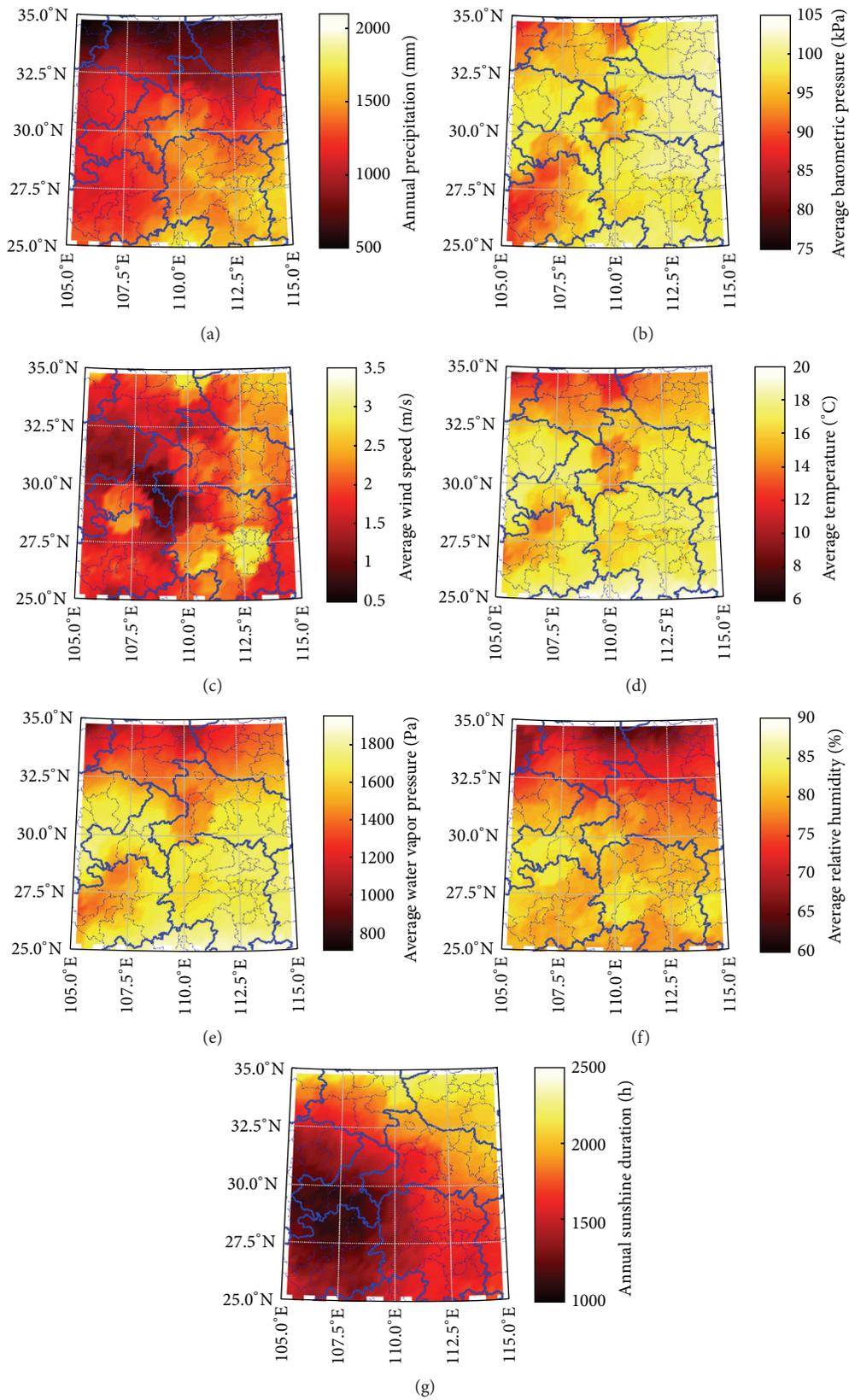


FIGURE 7: Spatial distribution of multivariate BME with soft probabilistic data estimates for seven meteorological factors.

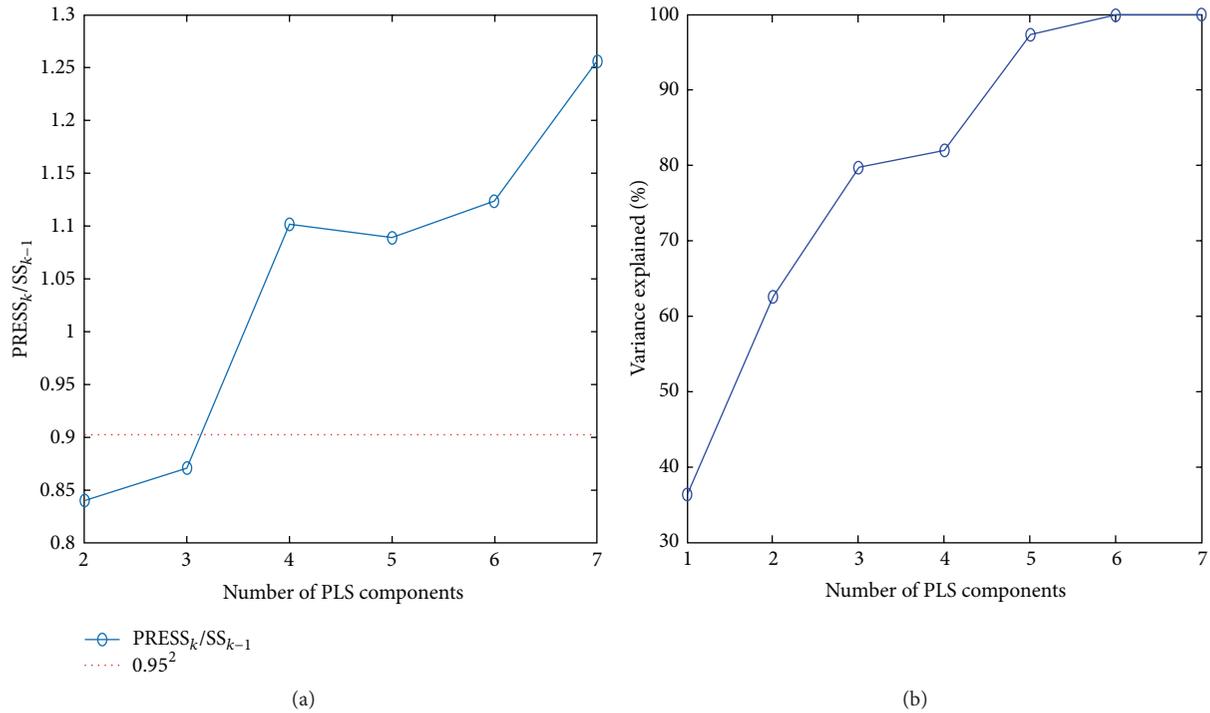


FIGURE 8: (a) PRESS/SS results of PLS components by cross-validation; (b) variance of meteorological factors explained by PLS components.

PLS fitting model to evaluate the regional degree of damage between the spatial distribution of mean annual values for seven meteorological factors and the losses originated by four meteorological disasters in the studied area; (3) the contribution of each meteorological factor to the disaster losses and its causes was analyzed, and these results may provide support for disaster prevention and urban planning.

Average annual spatial distributions of meteorological factors are important estimators, and the acquisition of accurate estimated values within the studied area is a prerequisite for building regression model of the relationship between the distribution of multivariate meteorological factors and historical disaster-related losses. Some studies have focused on finding the potential distribution regularity of meteorological factors in different regions or the laws governing the spread of infectious diseases by integrating hard data and soft data into a BME framework [22]. These studies show that meteorological factors not only are suitable regionalized variables but also are correlated with other variables and if these other variables are not taken into consideration, more accurate estimates cannot be made. In most cases, partial detailed and accurate meteorological data are available from present ground-based meteorological observation systems. Relatively few studies have evaluated the content of the distribution of various meteorological factors of a selected area, on the same time scale.

The BME approach showed good performance in mapping results after comparing the performance between BME and traditional spatial interpolation techniques (Figure 6). The MAE and RMSE of estimators by multivariate BME with probabilistic soft data were the closest to zero of any of the

meteorological factors for all these methods (Figures 6(a)–6(g)). The ME fluctuated due to the offset of the positive errors and minus errors, so Figures 6(a), 6(d), 6(f), and 6(g) show the minimum ME of estimators by multivariate BME, those associated with univariate BME are shown in Figure 6(b), and Figures 6(c) and 6(e) show the minimum ME of estimators by cokriging. The multivariate methods (cokriging and multivariate BME) were more accurate than univariate methods (kriging and univariate BME), and the BME methods (univariate BME and multivariate BME) were more accurate than kriging methods (kriging and cokriging). In all the cases, the best results were produced by multivariate BME with probabilistic soft data, better than those produced using kriging, cokriging, and univariate BME with probabilistic soft data. Generally speaking, performance in the estimated results was determined very well using BME approach, because it took soft information and spatial correlation information of these seven meteorological factors into account.

The estimation of the distribution of meteorological disaster losses in space is essential. Here, PLS regression was used to evaluate the effects of the spatial distribution of these seven factors on meteorological disaster losses in a specific area. Results showed that the reduction of the number of input variables by PLS removed their collinearity, and the 3-PLS-component regression model provided the best predictive ability (Figure 8). This reduction help offer an intuitive comprehension of the intricate relationships between meteorological factors and meteorological losses. The variable loadings are indicative of the importance of each meteorological factor used in the PLS method. The first PLS component is an integrated influence of all of these meteorological factors

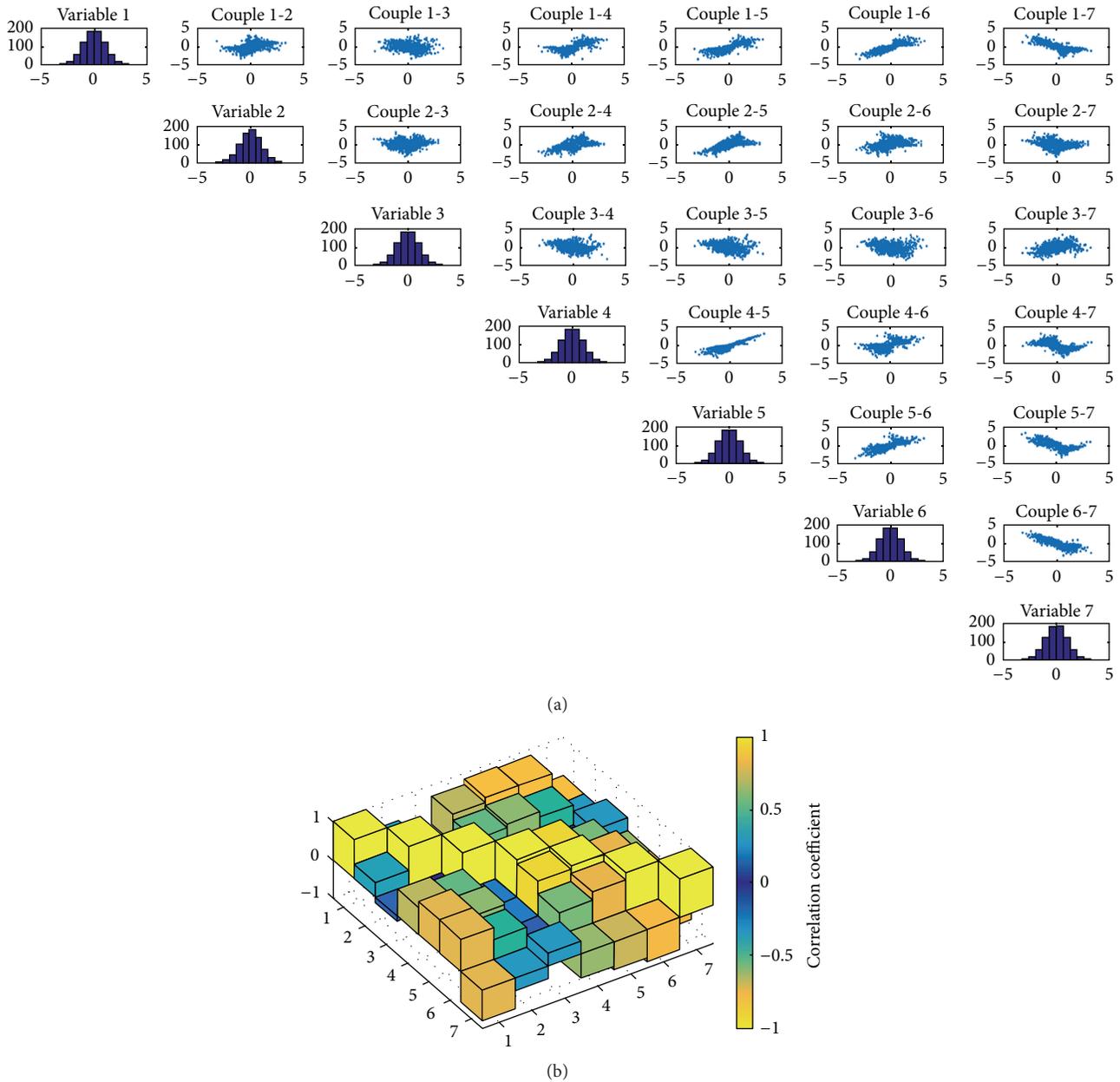


FIGURE 9: Correlations between seven meteorological factors.

except average wind speed and annual sunshine duration, which made few contributions to the first PLS component. The second PLS component was mainly affected by average wind speed and annual sunshine duration. Average temperature, average wind speed, and average relative humidity had negative influence on it. In the third PLS component, all the factors had relatively low loadings (Table 1).

Among all seven meteorological factors, the regression model (see (17)) shows that average barometric pressure contributes more to the meteorological disaster losses in the studied area than any other factor. Barometric pressure has a close correlation with altitude. The lower the altitude, the higher the barometric pressure. The low-altitude part in the

studied area is mainly Yangtze River Plain, Huabei Plain, and Sichuan Basin, where the population is large, and agriculture and commercial activity are prosperous, so meteorological disasters may cause large losses here. This was indirectly explained by the significance of barometric pressure that became visible in the model. Annual precipitation is the second significant factor to the disaster losses. It explains the losses caused by water-based disasters. The third significant factor is annual sunshine duration, which refers to times without precipitation and explains the severity of drought losses. The following factor, average temperature, also contributes to disaster losses, and its negative efficiency in the model indicates losses caused by low temperatures. The

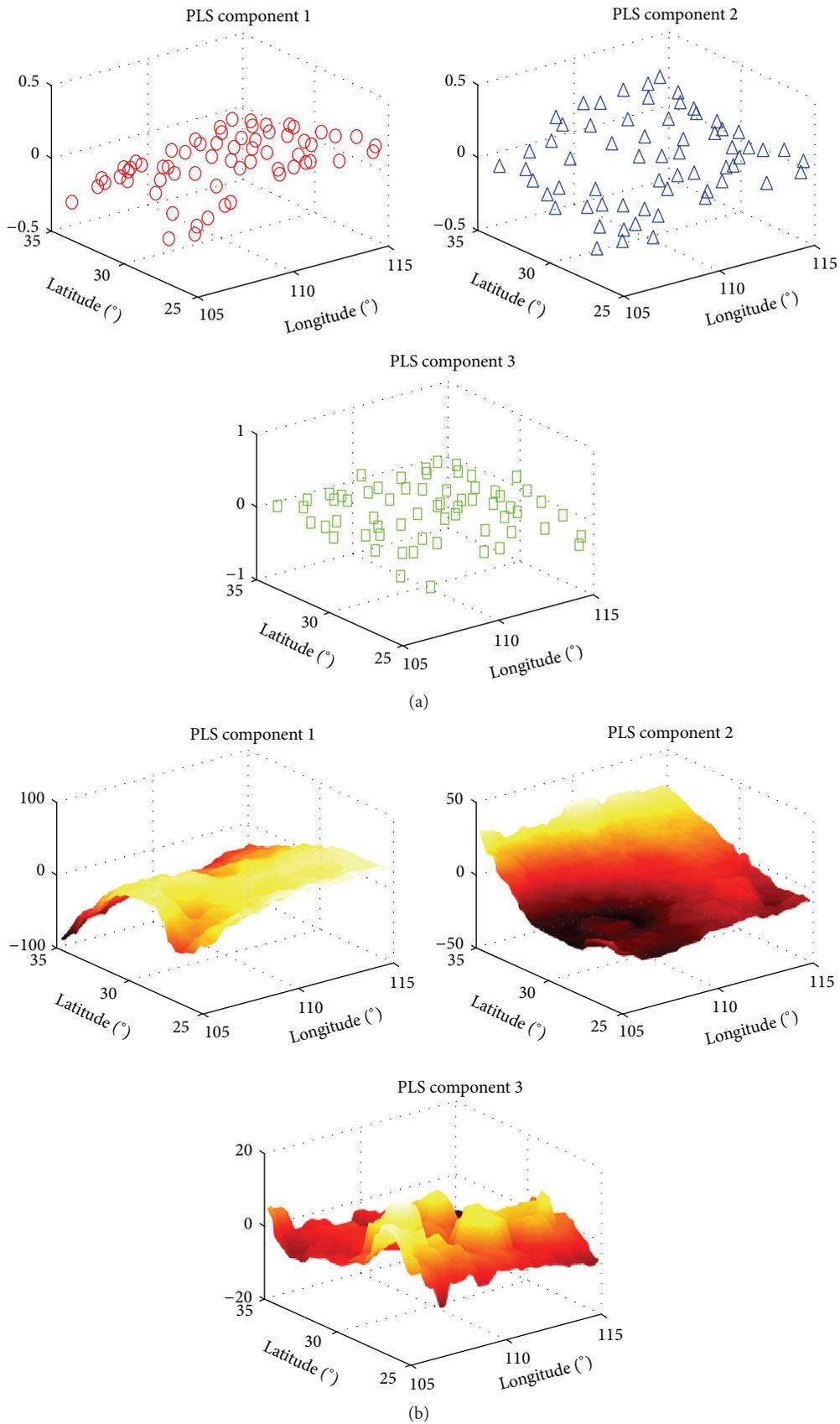


FIGURE 10: Spatial distributions of PLS components in the studied area. (a) Spatial distributions of the first three PLS components, extracted from hard data of seven meteorological factors. (b) Spatial distributions of the first three PLS components, calculated using spatial distribution of seven meteorological factors with multivariate BME with soft probabilistic data methods.

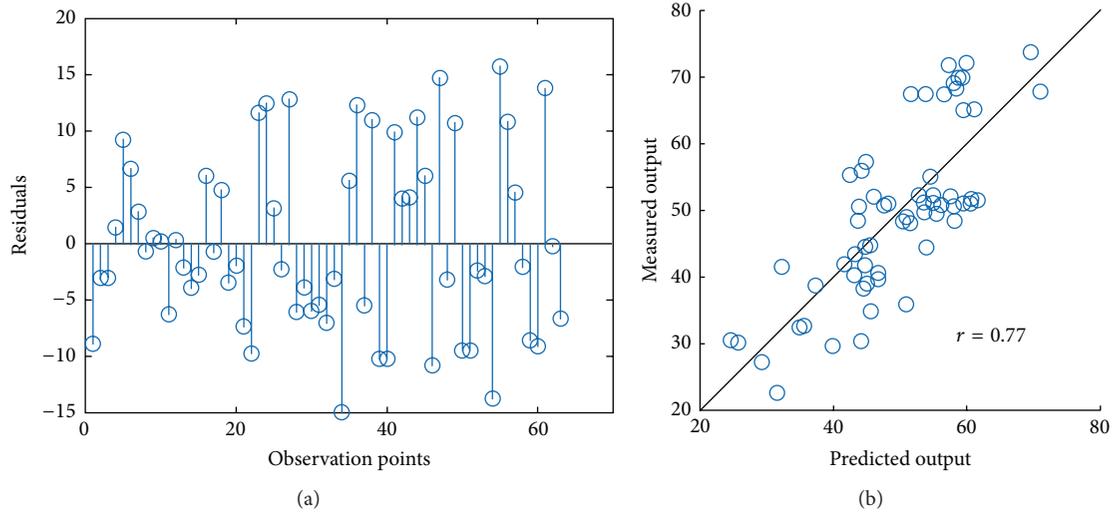


FIGURE 11: Residuals and correlation plot of PLS regression model. (a) Residuals of the predictions of PLS regression model. (b) Correlation plot between PLS regression model predicted output and measured output.

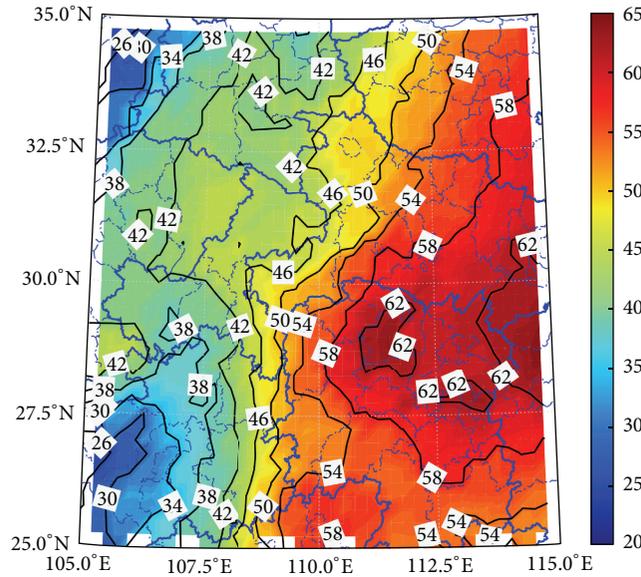


FIGURE 12: Distribution of meteorological losses risk in studied area.

fifth significant factor is average relative humidity, which is supplementary information regarding water-related disasters and drought. The penultimate factor is average wind speed. Its relatively small coefficient value suggests that the wind disaster in the area contributes relatively little to the meteorological disaster losses. Average water vapor pressure has the smallest coefficient value in the model, close to zero, which means this factor has little influence on those losses.

With the regression model and the spatial distributions of seven meteorological factors, the risk of losses due to meteorological disasters in the studied area was evaluated. Result showed northern Hunan Province to have the highest risk and southeastern Gansu Province and western Guizhou Province to have the lowest (Figure 12). The results of estimation were consistent with the traditional understanding of

regional disasters [36]. However, they are a relatively objective reflection of the spatial distribution of losses associated with four actual meteorological disasters. This indicates that the BME-PLS approach is suitable for studies of spatial distribution of losses associated with such disasters.

From what has been discussed above and from previous methods and results, the following can be found: (1) Through the comparison of the interpolation results using cross-validation, results showed that the BME method performs better than the traditional interpolation methods (kriging and cokriging) with respect to the precision of interpolated results, and the multivariate BME method was found to be the most precise of the four methods of estimation used here; (2) the PLS method in the BME-PLS model can be used to determine the relationship between meteorological factors; losses

associated with four kinds of meteorological disasters were constructed, the PLS components were extracted, the spatial correlations between meteorological factors were found, and the index of disaster losses was fitted. As shown in the residual plot and scatter plot, results showed that the model had a high fitting accuracy. However, other fitting methods cannot extract components or facilitate analysis of the contribution and relationship of meteorological factors.

This study suffers from limitations that should be addressed.

First, only moderately detailed descriptions of disaster-associated losses were available for the research area. PLS was used to reduce the dimensionality of the output variables, that is, by the aid of using the multidimensional disaster loss variables, to better facilitate PLS regression. Second, the spatial distribution of disaster losses at different times requires more investigation. The dynamic changes of the distribution of disaster losses were found to vary over time by extending the BME model from now spatial dimension to space-time dimension and analyzing the laws associated with this change.

5. Conclusions

To estimate losses associated with meteorological disasters in the studied area, kriging and BME methods and univariate cases and multivariate cases were compared, and a regression model was established using PLS, covering seven meteorological factors and meteorological disaster losses. The flexibility of BME allowed the assimilation of prior knowledge and soft data, which improved the accuracy of the estimation process. Correlations between meteorological factors provided in the multivariate cases 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 distributions of the seven meteorological factors in the studied area. The regression model using three PLS components extracted from seven meteorological factors using PLS was found to be most predictive by means of PRESS/SS test. With the regression model and the spatial distributions of seven meteorological factors, the risk of meteorological disaster losses in the studied area was evaluated. The results showed that northern Hunan Province had the highest risk and southeastern Gansu Province and western Guizhou Province had the lowest risk within the area. The distinct advantages of the BME-PLS method over previous methods can be summarized from the above discussion: (1) This method shows the more accurate spatial distribution of meteorological factors and regional disaster losses compared to previous methods. (2) It is not only able to identify the detailed distribution of regional disaster losses. Rather, it can also find the relationship between meteorological factors and determine the contribution of specific meteorological factor to disaster losses. It provides supports for meteorological disaster prevention and reduction, emergency resource allocation, and new urban development. (3) It is universal and extensible. This BME-PLS method was found to be predictive in the estimation of the risk of losses associated with meteorological disasters and

can be used in other areas of interest, covering more different types of meteorological disasters.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

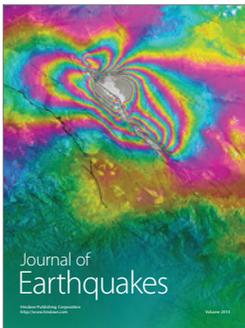
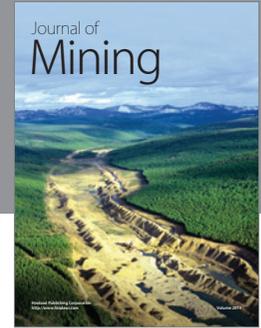
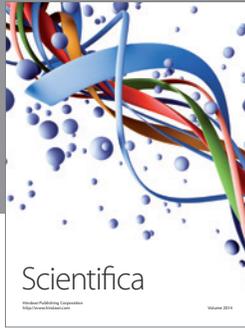
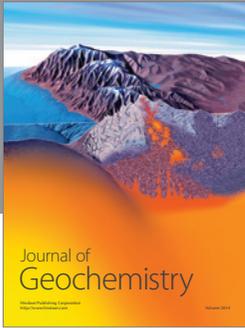
The authors would like to thank the Climatic Data Center, National Meteorological Information Center, China Meteorological Administration. Data attributed to spatial mapping of seven meteorological factors was downloaded from <http://www.escience.gov.cn/metdata/page/index.html>. They thank the supports provided by National Natural Science Foundation of China (no. 91224004) and Youth Talent Plan Program of Beijing City College (no. YETP0117).

References

- [1] Q. Zhang, J. Zhang, L. Jiang, X. Liu, and Z. Tong, "Flood disaster risk assessment of rural housings—a case study of Kouqian town in China," *International Journal of Environmental Research and Public Health*, vol. 11, no. 4, pp. 3787–3802, 2014.
- [2] P. Chen, J. Zhang, L. Zhang, and Y. Sun, "Evaluation of resident evacuations in urban rainstorm waterlogging disasters based on scenario simulation: Daoli district (Harbin, China) as an example," *International Journal of Environmental Research and Public Health*, vol. 11, no. 10, pp. 9964–9980, 2014.
- [3] J. Cheng and J.-P. Tao, "Fuzzy comprehensive evaluation of drought vulnerability based on the analytic hierarchy process—an empirical study from Xiaogan City in Hubei Province," *Agriculture and Agricultural Science Procedia*, vol. 1, pp. 126–135, 2010.
- [4] H. Jayanthi, G. J. Husak, C. Funk, T. Magadzire, A. Adoum, and J. P. Verdin, "A probabilistic approach to assess agricultural drought risk to maize in Southern Africa and millet in Western Sahel using satellite estimated rainfall," *International Journal of Disaster Risk Reduction*, vol. 10, pp. 490–502, 2014.
- [5] N. C. Gould, R. E. Vega, and S. H. Sheppard, "Extreme wind risk assessment of the Miami Marlins' New Ballpark in Miami, Florida," in *Proceedings of the ATC & SEI Conference on Advances in Hurricane Engineering*, C. P. Jones and L. G. Griffith, Eds., pp. 1194–1204, Miami, Fla, USA, October 2012.
- [6] L. Lu and P. Shi, "The comparison of function systems in large-scale disaster response between China and USA—cases of 2008 southern china freezing rain and snow storm disaster and 2005 hurricane katrina," *Journal of Catastrophology*, vol. 29, pp. 206–213, 2014.
- [7] K. Butler, C. Kafali, and V. Jain, *Cyclone Risk from Wind, Flood, and Storm Surge Perils in Australia: A Comprehensive Model*, American Society of Civil Engineers, 2013.
- [8] L. J. Zhang, H. Y. Zhu, and X. J. Sun, "China's tropical cyclone disaster risk source analysis based on the gray density clustering," *Natural Hazards*, vol. 71, no. 2, pp. 1053–1065, 2014.
- [9] A. S. Akintoye and M. J. MacLeod, "Risk analysis and management in construction," *International Journal of Project Management*, vol. 15, no. 1, pp. 31–38, 1997.
- [10] R. Kamat, "Planning and managing earthquake and flood prone towns," *Stochastic Environmental Research and Risk Assessment*, vol. 29, no. 2, pp. 527–545, 2014.

- [11] M. Stojković, A. Ilić, S. Prohaska, and J. Plavšić, "Multi-temporal analysis of mean annual and seasonal stream flow trends, including periodicity and multiple non-linear regression," *Water Resources Management*, vol. 28, no. 12, pp. 4319–4335, 2014.
- [12] L. Liu, J. Wang, Z. Qu, F. Gao, Q. Chai, and Y. Liang, "Research on non-linear regression model of meteorological factors for apple yield in Shaanxi Province," *Chinese Agricultural Science Bulletin*, vol. 28, pp. 248–251, 2012.
- [13] W. Xu, X. Pu, J. Tong, and J. Yang, "Non-linear regression model of solar radiation in some areas in China," *Yangzhou Daxue Xuebao Ziran Kexue Ban*, vol. 14, no. 1, pp. 27–31, 2011.
- [14] G. Tezel and M. Buyukyildiz, "Monthly evaporation forecasting using artificial neural networks and support vector machines," *Theoretical and Applied Climatology*, 2015.
- [15] T. Yang, X. Zhou, Z. Yu, V. Krysanova, and B. Wang, "Drought projection based on a hybrid drought index using artificial neural networks," *Hydrological Processes*, vol. 29, no. 11, pp. 2635–2648, 2015.
- [16] I. V. Sideris, M. Gabella, R. Erdin, and U. Germann, "Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland," *Quarterly Journal of the Royal Meteorological Society*, vol. 140, no. 680, pp. 1097–1111, 2014.
- [17] H. Y. Yeung, C. Man, S. T. Chan, and A. Seed, "Development of an operational rainfall data quality-control scheme based on radar-rain-gauge co-kriging analysis," *Hydrological Sciences Journal*, vol. 59, no. 7, pp. 1293–1307, 2014.
- [18] 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.
- [19] N. Diodato, "The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain," *International Journal of Climatology*, vol. 25, no. 3, pp. 351–363, 2005.
- [20] C. Prudhomme and D. W. Reed, "Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland," *International Journal of Climatology*, vol. 19, no. 12, pp. 1337–1356, 1999.
- [21] D. D'Or and P. Bogaert, "Continuous-valued map reconstruction with the Bayesian Maximum Entropy," *Geoderma*, vol. 112, no. 3–4, pp. 169–178, 2003.
- [22] J. Angulo, H.-L. Yu, A. Langousis et al., "Spatiotemporal infectious disease modeling: a BME-SIR approach," *PLoS ONE*, vol. 8, no. 9, Article ID e72168, 2013.
- [23] M. L. Serre and G. Christakos, "Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge—the Equus Beds study," *Stochastic Environmental Research and Risk Assessment*, vol. 13, no. 1–2, pp. 1–26, 1999.
- [24] 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.
- [25] G. Christakos and X. Li, "Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators?" *Mathematical Geology*, vol. 30, no. 4, pp. 435–462, 1998.
- [26] Z. Gong and J. Y.-L. Forrest, "Special issue on meteorological disaster risk analysis and assessment: on basis of grey systems theory," *Natural Hazards*, vol. 71, no. 2, pp. 995–1000, 2014.
- [27] N. Xie, J. Xin, and S. Liu, "China's regional meteorological disaster loss analysis and evaluation based on grey cluster model," *Natural Hazards*, vol. 71, no. 2, pp. 1067–1089, 2014.
- [28] M. Shao, Z. Gong, and X. Xu, "Risk assessment of rainstorm and flood disasters in China between 2004 and 2009 based on gray fixed weight cluster analysis," *Natural Hazards*, vol. 71, no. 2, pp. 1025–1052, 2014.
- [29] T. Zhijun, Z. Jiquan, and L. Xingpeng, "GIS-based risk assessment of grassland fire disaster in western Jilin province, China," *Stochastic Environmental Research and Risk Assessment*, vol. 23, no. 4, pp. 463–471, 2009.
- [30] P. A. Burrough, "GIS and geostatistics: essential partners for spatial analysis," *Environmental and Ecological Statistics*, vol. 8, no. 4, pp. 361–377, 2001.
- [31] B. Ramlal and S. M. J. Baban, "Developing a GIS based integrated approach to flood management in Trinidad, West Indies," *Journal of Environmental Management*, vol. 88, no. 4, pp. 1131–1140, 2008.
- [32] G. Demir, M. Aytekin, and A. Akgun, "Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar-Resadiye (Tokat, Turkey)," *Arabian Journal of Geosciences*, vol. 8, no. 3, pp. 1801–1812, 2015.
- [33] T.-C. Lei, Y.-M. Huang, B.-J. Lee, M.-H. Hsieh, and K.-T. Lin, "Development of an empirical model for rainfall-induced hillside vulnerability assessment: a case study on Chen-Yu-Lan watershed, Nantou, Taiwan," *Natural Hazards*, vol. 74, no. 2, pp. 341–373, 2014.
- [34] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn III, "The collinearity problem in linear regression. The Partial Least Squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 3, pp. 735–743, 1984.
- [35] Y. Liu and W. Rayens, "PLS and dimension reduction for classification," *Computational Statistics*, vol. 22, no. 2, pp. 189–208, 2007.
- [36] Y. Ding, *China Meteorological Disaster Authority (Comprehensive Volume)*, China Meteorological Press, Beijing, China, 2008.
- [37] K. Modis, K. I. Vatalis, and C. Sachanidis, "Spatiotemporal risk assessment of soil pollution in a lignite mining region using a Bayesian maximum entropy (BME) approach," *International Journal of Coal Geology*, vol. 112, pp. 173–179, 2013.
- [38] D. D'Or, P. Bogaert, and G. Christakos, "Application of the BME approach to soil texture mapping," *Stochastic Environmental Research and Risk Assessment*, vol. 15, no. 1, pp. 87–100, 2001.
- [39] G. Christakos, P. Bogaert, and M. L. Serre, *Temporal GIS: Advanced Functions for Field-Based Applications*, Springer Science & Business Media, Berlin, Germany, 2001.
- [40] G. Christakos, "On the assimilation of uncertain physical knowledge bases: Bayesian and non-Bayesian techniques," *Advances in Water Resources*, vol. 25, no. 8–12, pp. 1257–1274, 2002.
- [41] G. Christakos, "A Bayesian/maximum-entropy view to the spatial estimation problem," *Mathematical Geology*, vol. 22, no. 7, pp. 763–777, 1990.
- [42] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [43] Y.-K. Tu, G. D. Smith, and M. S. Gilthorpe, "A new approach to age-period-cohort analysis using partial least squares regression: the trend in blood pressure in the Glasgow alumni cohort," *PLoS ONE*, vol. 6, no. 4, Article ID e19401, 2011.
- [44] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, Springer, New York, NY, USA, 1998.

- [45] E. P. Green, *Mathematical Tools for Applied Multivariate Analysis*, Academic Press, San Diego, Calif, USA, 1976.
- [46] A. Basilevsky, *Applied Matrix Algebra in the Statistical Sciences*, North-Holland Publishing, New York, NY, USA, 1983.
- [47] G. Wang and H. Ye, *Principal Component Analysis and Partial Least Square*, Tsinghua University Press, Beijing, China, 2012.



Hindawi

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

