# Comparative Analysis of Single and Mixed Spatial Interpolation Methods for Variability Prediction of Temperature Prediction

Yunhua Gu, Lingmei Sun and Jin Wang

*Jiangsu Engineering Center of Network Monitoring, School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China, yhgu6655@163.com, slm.happy@163.com, wangjin@nuist.edu.cn*

***Abstract***

*This paper analyzes the validity of temperature maps obtained by means of single and mixed interpolation methods. In this context, several interpolation methods are used for the temperature mapping: inverse distance square (IDS), ordinary kriging (OK) and co-kriging (CK) and mixed methods (combined global, local and geostatistical methods). And the validity of the maps is checked through the cross-validation error statistics presented in terms of Mean absolute error (MAE) and root-mean-square error (RMSIE). The results show that the spatial temperature distribution formed by IDS, OK and CK show similarity in general while the CK method better reflects the temperature distribution. Moreover, the best result for temperature mapping is obtained using the mixed method, which suggests that the correction of regression models using residuals improves the interpolation accuracy. Reliable calculation for agricultural management and improvement of climatic models at local scales can be obtained with increased efficiency.*

***Keywords:*** *spatial interpolation, single interpolation method, mixed interpolation method, ArcGIS, exploratory data analysis*

## 1. Introduction

Spatial interpolation is an essential tool in processing the data of natural and social science. It has been widely used especially in the discipline of hydrological, meteorological climate, ecology, environment, geology etc. [1, 2]. The essence of spatial interpolation is to estimate the values of unobserved points based on known sample data. Due to non-availability of abundant measurement points, reliable estimation of temperature distribution poses a great challenge. At present, many domestic and foreign scholars and institutions study the spatial interpolation of meteorological elements. The spatial interpolation prediction techniques (like spline, inverse distance weighting and kriging) provide better estimation of temperature than conventional methods [3]. There is no single preferred method for data interpolation which can meet with the selection criteria of required level of accuracy, the time and/or computer resources etc. The common approach to select the optimal spatial interpolation method has become the focus.

Domestic studies tend to compare the different effects of single interpolation methods, few studies are involved in the comparative analysis of single and mixed spatial interpolation methods [4]. The main purpose of this study is to provide the basis for the selection of an appropriate interpolation method and map the spatial variability. Take the spatial interpolation of Henan province as an example, this paper mainly focuses on the following specific objectives: (1) to make the preprocessing data analysis; (2) to utilize the single interpolation methods (inverse distance square, ordinary kriging, co-kriging); (3) to conduct the multiple regression analysis and utilize the mixed method; (4) to determine the validity of interpolated temperature maps through statistical criteria and subjective comments. According to cross-validation error statistics, the spatial temperature distribution formed by the three interpolation methods show similarity in general while the CK method better reflects the temperature distribution. Mixed interpolation method shows better accuracy measurement than other three methods, the correction of regression models using residuals improves the interpolation accuracy. The result indicates that the proposed method and procedures are highly effective to predict temperature.

## 2. Exploratory Data Analysis

The data we get often contains lots of noise or inconsistent data due to a variety of errors from human or system. To meet actual interpolation needs, data preprocessing is always indispensable. In this paper, the original data of 32 climatic stations don't follow normal distribution according to the Kolmogorov–Smirnov test. To meet needs of normal distribution, logarithmic transformation and Box-Cox transformation are carried out for the long-term mean monthly temperature. And the skewness, kurtosis and P after logarithmic transformation best meet the normal distribution.

The statistical fitting parameters and the variogram model fitting are accomplished through use of ArcGIS and GS+ software. The analysis in this paper indicates that all nugget/sill ratio values of three semivariograms are in the range of 25%-75% [5, 6]. They all have moderate nugget/sill ratio. Figure 1 shows the spherical, exponential and Gaussian model fitting curve for annul temperature.

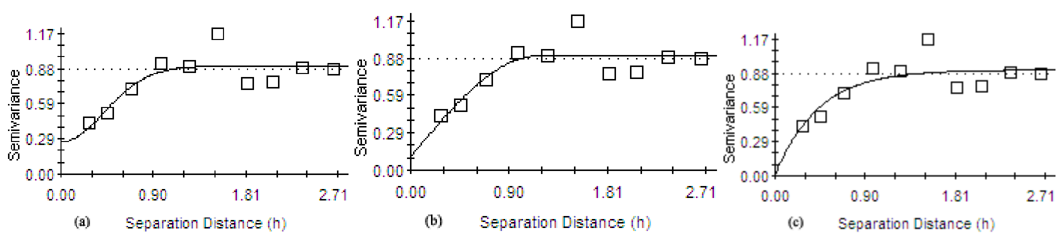



**Figure 1. Semivariogram Models for Temperature: (a) Sepherical Model; (b) Exponential Model; (c) Gaussian Model**

The small rectangle in the figure indicates the separation distance for the average distance limited by the step value of all points. The vertical axis indicates semi-variance of the points. Smaller the overall deviation from the fitting curve, better the curve fits. In this study, the Gaussian model fit better than the other two models.

## 3. Single Interpolation Methods

The single deterministic function methods (global interpolators, local interpolators and geostatistical methods) are called as single interpolation methods in this article. The single interpolation methods assessed in this study include inverse distance square (IDS), ordinary kriging (OK) and co-kriging (CK).

### 3.1. Inverse Distance Square (IDS)

The IDS method is based on the first law of geography, which indicates nearby values contribute more to the interpolated values than distant observations. When the weight value is given by the inverse of the square of the distance, we call it the inverse distance square. This supposes that predictions are obtained from the nearest sampling points. The formula is shown as (1).

$$Z = \sum_{i=1}^{n} \frac{Z_i}{d_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{d_i^2} \tag{1}$$

Where $Z$ refers to the predicted value; $Z_i$ is the observed value of a neighboring weather station $i$; $d_i$ is the distance between $Z$ and $Z_i$; and $n$ refers to the number of sample points. The interpolation accuracy of the method usually depends on the non-uniformity of the sample points. IDS is sensitive to outliers, while unevenly distributed data clusters results in introduced errors.

### 3.2. Ordinary Kriging (OK)

Kriging is a statistically-based interpolation method. Based on spatial autocorrelation, it uses the structural properties of the raw data and the variogram. It's an interpolation method that gets the partial valuation of regionalized variables of unknown sample points [5]. OK assumes that sampling points do not have the potentially global trends, and that it can well estimate the unknown values by using only local factors. The climatic value at point $Z(x_0)$ is expressed by formula (2) .

$$Z(x_0) = \sum \lambda_i z(x_i) \tag{2}$$

Where $\lambda_i$ refers to the weight of the regionalized variable $z(x_i)$, which is used to represent the contribution of the observation $z(x_i)$. $z(x_i)$ is the estimated value of $z(x_0)$. The weight coefficient $\lambda_i$ is obtained by formula (3) .

$$\begin{cases} \sum_{i=1}^{n} \lambda_i = 1 \\ \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \varphi = \gamma(x_i, x_0) \forall_j \end{cases} \tag{3}$$

Where $\gamma$ (xi, xj) refers to the semi-variance of $Z$ between the sample points $x_i$ and $x_j$; $\gamma$ ($x_i$, $x_0$) refers to the semi-variance of $Z$ between the sample point $x_i$ and estimates point $x_0$; $\varphi$ is the Lagrange multiplier in the minimal treatment. These variables are obtained through the variogram and its formula is shown as (4) .

$$\gamma(h) = \frac{1}{2n} \sum_{1}^{n} [Z(x_i + h)]^2 \tag{4}$$

As in formula (4), $h$ refers to the step; $\gamma$ (h) refers to semi-variance of $Z$ whose step is $h$; $n$ refers to the number of experimental data split by $h$. The most commonly used variogram model in OK are ring model, exponential model, spherical model, gaussian model, etc [6].

### 3.3. Co-Kriging (CK)

There are different types of kriging, and several papers describe them in detail [7] (Goovaerts 1997). Kriging method to estimate the spatial variables with the spatial correlation can be classified as CK method. With this method, you can make spatial estimation of one variable or multiple variables using the correlation between several space variables, so that the accuracy and reasonableness of the estimation can be improved. The key to use CK is calculating the covariation function between variables. If $v_1(x)$, $v_2(x)$ ... $v_m(x)$ denote the values of the variables $v_1$, $v_2$ ,... , $v_m$ in the position $x$. If $x_1$, $x_2$ ,... , $x_n$ is the position of the sample points and the measured values are $Vx_1$, $Vx_2$ ... $Vx_n$, then the Co-kriging method can be expressed as formula (5) .

$$V_\varepsilon(x) = \sum_{i=1}^{n} V(x_i)\Gamma_i \tag{5}$$

As in formula (5), $\Gamma_i$ refers the weight vector given by the covariation function, which is calculated as formula (6) .

$$Y_{12}(h) = E\{[v_1(x+h) - v_1(x)][v_2(x+h) - v_2(x)]\} \tag{6}$$

CK depends on spatial and statistical relationships to calculate the surface. There are some advantages of this method, such as the incorporation of variable interdependence, the available error surface output, etc. A disadvantage is that it requires substantially more computing and modeling time [8].

## 4. Mixed Methods

### 4.1. Overview of Mixed Methods

Global methods are inexact interpolators, since there is a known error between the final predicted data and the original observed data (residual). Mixed interpolation method uses a correction method (interpolation of residuals) to further improve the interpolation accuracy. It’s an effective interpolation means, which combines the advantages of global interpolation and local interpolation methods [9].

$$\text{residual} = \text{observed data} - \text{predicted data} \tag{7}$$

Then, interpolation of residuals must be conducted to obtain correction maps. Many interpolation methods can be used to interpolate the residuals of climatic maps, such as splines, inverse distance weighting, kriging etc. The sum of residual interpolation and predicted regression maps modifies the initial results. The real values can be obtained as:

$$\text{observed data} = \text{predicted data} + \text{residual interpolation} \tag{8}$$

### 4.2. Universal Processes

Mixed interpolation method is a combination of multiple regression and geostatistical methods. To establish the regression equation between the dependent variables and independent variables, the independent variables must be confirmed in advance. After a thoroughly analysis of correlation between temperature and geographic features (latitude and longitude altitude and terrain, etc), features with strong correlation are taken as independent variables. The multiple regression equation can be obtained as formula (9).

$$T = a_0 + a_1 X + a_2 Y + a_3 Z + \cdots \tag{9}$$

Where T refers to temperature, the independent variables X, Y, Z are the geographic features such as latitude, longitude, elevation, slope, etc. And a0 ~ a3 are the constant and coefficient of the independent variables. Then the residuals between observed data and the predicted data can be calculated, and the interpolation for residuals can be carried on. A diversity of methods can be used for residual interpolation, include splines, inverse distance weighting, series of kriging etc. After interpolation for residuals, the corrected temperature map can be obtained by adding residual interpolation results and the predicted regression maps.

### 4.3. An Instance

The mixed interpolation method in this paper is a combination of multiple regression and kriging interpolation method. It's based on the assumption that the residual retains the inherent spatial structure of the original data. Since the collinearity statistic has a great impact on the result of stepwise regression analysis, collinearity analysis is essential. And the tolerance and variance inflation factor are the main parameters of collinearity analysis. If the tolerance is less than 0.2, it means the existence of collinearity among factors. And the tolerance less than 0.1 indicate an obvious collinearity. The longitude, latitude and elevation are taken as variables of stepwise regression analysis. Table 1 shows that 3 tolerance values are larger than 0.2, so the latitude, longitude, and elevation can be taken as independent variables of stepwise regression.

**Table 1. Multicollinearity Diagnosis Indexes for Variables**

| Independent Variable | Tolerance | variance inflation factor |
|---|---|---|
| longitude | 0.449 | 2.228 |
| latitude | 0.994 | 1.006 |
| elevation | 0.448 | 2.231 |

The stepwise multiple regression analysis is performed with temperature as dependent variables and geographic variables as predictors. So, the predicted temperature (T) can be created by the regression equation. It is shown as formula (10).

$$T = 33.867 - 0.057 * X - 0.378 * Y - 0.001 * Z \tag{10}$$

Where X refers to the longitude, Y is the latitude and Z represents the elevation. To obtain correction maps generated by residuals interpolation, CK is used to interpolate the residuals of the temperature map created by means of the regression-based method. The sum of residual interpolation and predicted regression maps comes to the final temperature map.

## 5. Interpolation Effect Assessment

Spatial interpolation is widely used for creating continuous data from data collected at discrete locations. When an interpolated surface is used as part of larger research project [8] both the method and accuracy of the interpolation technique are important. To create a surface best represent the empirical reality, the method selected must be assessed for accuracy.

To test the accuracy of the interpolation results of meteorological stations in the climate indicator, actual validation and cross-validation are commonly used [10].For the purpose of maximize the use of observations, cross-validation is chosen to comparatively analyze the interpolation results. Temperature maps must be assessed by statistics that indicate the degree of concordance between models and reality. Mean absolute error (MAE) and root-mean-square error (RMSIE) are among the 'best' overall measures of model performance, as they summarize the mean difference in the units of observed and predicted values. The MAE provides a measure of bias; the RMSIE provides a measure of accuracy [11]. Measurements of MAE and RMSIE are shown as formula (11), (12), Where $T_{oi}$ is the measured value of Z at location i, $T_{ei}$ is the predicted value at the same location.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} ABS(T_{oi} - T_{ei}) \tag{11}$$

$$RMSIE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{oi} - T_{ei})^2} \tag{12}$$

Nevertheless, not only statistical criteria are used to determine the validity of interpolated temperature maps. Subjective comments, based on the knowledge of spatial distribution of temperature, are also added.

## 6. Experiment and Verification

As an essential meteorological element of biological survival, temperature has a very important impact on Earth. Despite abundant temperature data have been probed through long-term observation, but these data only represent the meteorological stations within the limited confines of the location of the temperature conditions during the regional / global ecological system.

### 6.1. Data

The location of the study area is indicated in Figure 2. The total area of the study area is 167,000 km2 with only thirty-two meteorological stations to measure temperature all over the Henan province. It is located between latitude 31.23 °~36.22 °, east longitude 110.21 °~116.39 °.

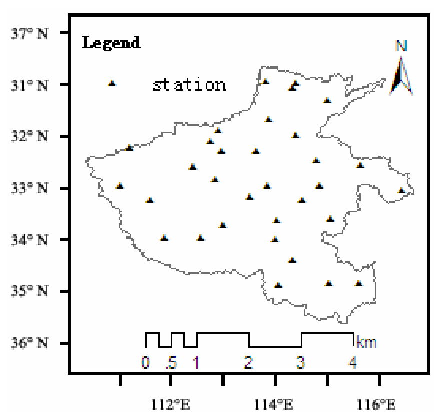



**Figure 2. Monthly Mean Temperature Distribution of Henan Province**

The climate of Henan province is warm temperate-subtropical monsoon with mean annual temperatures ranging between 12 ℃ and 16 ℃. Based on available historic records, a number of studies have been carried out to support crop production systems planning, meteorological disaster warning and so on. In this context, it is essential to select an appropriate interpolation method to map the spatial variability for the whole study area.

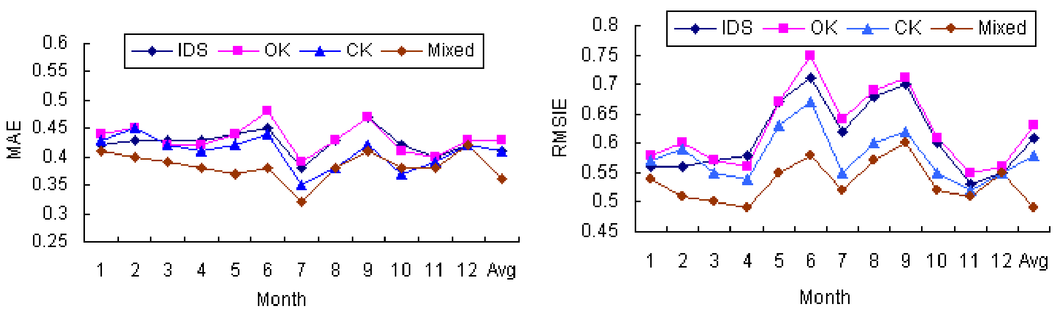
### 6.2. Statistical Results

The results of the accuracy measurements in the different temperature models are shown in Table 2. It indicates that the monthly average MAE values of interpolation IDS, OK, CK and mixed method are 0.41, 0.43, 0.41 and 0.36. Respectively, the average RMSIE values are 0.61, 0.63, 0.58 and 0.49. OK can be considered the worst model in statistical terms, as it gives the higher values of MAE and RMSIE.

**Table 2. Accuracy Measurements for Temperature Models: IDS, OK, CK and Mixed Methods are Expressed in ℃**

| Month | IDS | | OK | | CK | | Mixed | |
|---|---|---|---|---|---|---|---|---|
| | MAE | RMSIE | MAE | RMSIE | MAE | RMSIE | MAE | RMSIE |
| 1 | 0.42 | 0.56 | 0.44 | 0.58 | 0.43 | 0.57 | 0.41 | 0.54 |
| 2 | 0.43 | 0.56 | 0.45 | 0.60 | 0.45 | 0.59 | 0.40 | 0.51 |
| 3 | 0.43 | 0.57 | 0.42 | 0.57 | 0.42 | 0.55 | 0.39 | 0.50 |
| 4 | 0.43 | 0.58 | 0.42 | 0.56 | 0.41 | 0.54 | 0.38 | 0.49 |
| 5 | 0.44 | 0.67 | 0.44 | 0.67 | 0.42 | 0.63 | 0.37 | 0.55 |
| 6 | 0.45 | 0.71 | 0.48 | 0.75 | 0.44 | 0.67 | 0.38 | 0.58 |
| 7 | 0.38 | 0.62 | 0.39 | 0.64 | 0.35 | 0.55 | 0.32 | 0.52 |
| 8 | 0.43 | 0.68 | 0.43 | 0.69 | 0.38 | 0.60 | 0.38 | 0.57 |
| 9 | 0.47 | 0.70 | 0.47 | 0.71 | 0.42 | 0.62 | 0.41 | 0.60 |
| 10 | 0.42 | 0.60 | 0.41 | 0.61 | 0.37 | 0.55 | 0.38 | 0.52 |
| 11 | 0.40 | 0.53 | 0.40 | 0.55 | 0.39 | 0.52 | 0.38 | 0.51 |
| 12 | 0.42 | 0.55 | 0.43 | 0.56 | 0.42 | 0.55 | 0.42 | 0.55 |
| Annual | 0.41 | 0.61 | 0.43 | 0.63 | 0.41 | 0.58 | 0.36 | 0.49 |

Then, we try to evaluate the performance of IDS, OK, CK and the mixed methods by comparing the statistical errors generated by cross-validation. The MAE and RMSIE of each interpolation method are respectively shown in Figure 3(a) and Figure 3(b).




(a) (b)

**Figure 3. MAE and RMSIE of IDS, OK, CK and Mixed Methods: (a) MAE; (b) RMSIE**

It can be found that the MAE and RMSIE of IDS, OK and CK interpolation methods show similarity in general, with good general results in all cases. Nevertheless, CK produces lower error values than the other two single interpolation methods. In addition, the mixed methods have lowest MAE and RMSIE. It's proved to be an effective approach for the identification of best spatial interpolation techniques for further analysis.

### 6.3. Temperature Maps

Nevertheless, not only statistical criteria are used to determine the validity of the interpolated climatic maps. Direct empirical climatic knowledge can help determine the models that reflect reality best, as long as their statistical values are reasonable. Therefore, subjective comments, based on the knowledge of spatial distribution of temperature in Henan province, are also added in this paper. The annual mean temperature maps are shown in Figure 4.

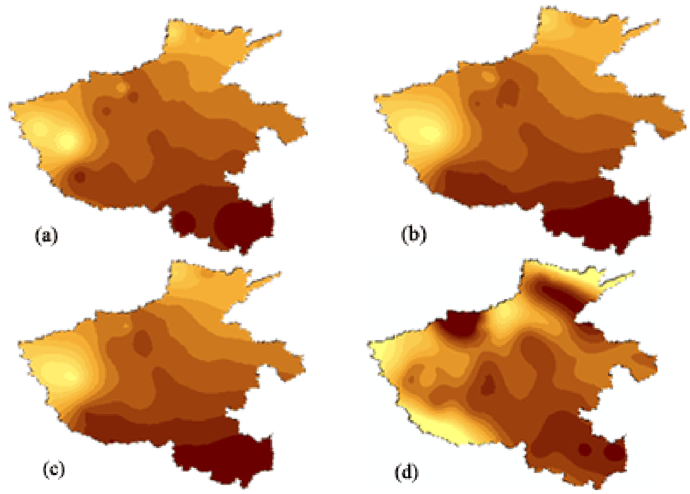



**Figure 4. Results of Interpolation Methods for Annual Mean Temperature (℃): (a) IDS Interpolation; (b) OK Interpolation; (c) CK Interpolation; (d) Mixed Interpolation**

It can be found that north-south gradient is detected by the four interpolation methods, while the general spatial patterns have substantial differences in specific areas. Large spatial differences between the diverse interpolated maps are observed. The single interpolation methods show similar results, although CK produces the better map than IDS and OK. The model obtained by IDS method show has concentric circles. Mixed interpolation method shows better accuracy measurement than other three methods, the correction of regression models using residuals improves the interpolation accuracy.

## 7. Discussion

Most of the interpolation models just take the factor of temperature into account. And the interpolation results without considering the impact of topography and other factors do have higher errors in prediction [12]. How to take the factors of elevation, slope and so on into interpolation models has been a major problem of spatial interpolation. In this paper, the use of different topographic and geographic factors that determine the spatial distribution of climatic variables makes it possible to improve the interpolation accuracy. Moreover, mixed method is used to map the climatic variables, which generates an adequate temperature map.

At present, researches from home and abroad show that data of one station always represent for data of nearby region. Less distribution of stations means larger area the observed value of one single station will represent for. Usually, characteristics of a single meteorological station can't accurately represent the regional climate, especially in the region with complex climate or large regional difference. Since there is no appropriate method to directly interpolate for temperature of a large scale region, the temperature information of the large region can only be obtained from the analysis of the very limited space. Downscaling interpolation is to reduce a large region into several small regions, means a large region is divided into a few small regions. To improve the interpolation accuracy, then the operation of interpolation for each small region separately is made on the basis of region division. The annual temperature maps developed by downscaling can be very useful for hydrological and agricultural management in the region. These maps must be the first step in the creation of a more complete spatial climatic database, which includes the diversity offered by seasonal or monthly temporal scales. The latter, which are very useful for planning, expand our understanding of spatial climatic complexity and its variation over time in the study area.

## 8. Conclusions

Creation of digital grid maps makes it possible to obtain climatic information at any point, whether there is a weather station or not. Multiple factors condition the difficulty of map creation, such as the location of the site samples, spatial density, spatial variability etc. The use of different topographic and geographic factors that determine the spatial distribution of climatic variables makes it possible to map more local climatic features. Since the same analysis can give the different results in different areas, prior tests need to be established to determine the most adequate method. Before geographic spatial interpolation, many procedures must be taken into account such as data preprocessing, analysis of spatial variability and correlation, the variation of key parameters and theoretical model. Based the characteristics and inner laws, the parameter settings of interpolation methods and the choice of variogram model are discussed in this study. And the mixed interpolation method is very useful in the mapping of climatic variables, because it adapts to almost any space and usually generates adequate maps. The proposed method and procedures in this paper can not only reduce the workload of the spatial interpolation model comparing and screening, but also improve the accuracy of spatial interpolation. It’s proved to be an effective approach for the identification of best spatial interpolation techniques for further analysis. The main conclusion to be drawn from this paper is that, although the topography of Henan province is not very complex, spatial climate diversity is significant. The annual temperature map developed can be useful for agricultural management in the region. Reliable calculation for agricultural management and improvement of climatic models at local scales can be obtained with increased efficiency.

## Acknowledgements

The work has been supported by the public service sectors (WMO) research project, project NO.GYHY201106024. This work was also supported by the Natural Science Foundation of Jiangsu Province (No. BK2012461).

## References

[1] X. Li, Y. Y. Xiao, W. C. Liu and J. Y. Tang, Meteorological Science and Technology, vol. 1, **(2002)**, pp. 41.
[2] D. Birkes and Y. Dodge, “Alternative Methods of Regression”, Wiley Online Library, Hoboken, **(2009)**.
[3] S. B. Liao and Z. H. Li, Meteorological Science and Technology, vol. 5, **(2004)**, pp. 252.
[4] A. Sarangi, C. A. Cox and C. A. Madramootoo, “ Geostatistical methods for prediction of spatial variability of rainfall in a mountainous region”, Transactions of the ASAE, vol. 48, no. 3, **(2005)**, pp. 943-954.
[5] C. A. Cambardella, T. B. Moorman, J. M. Novak, T. B. Parkin , D. L. Karlen, R. F. Turco and A. E. Konopka, “Field-Scale Variability of Soil Properties in Central IOWA Soils”, Soil Science Society of America, vol. 58, **(1994)**, pp. 1501-1511.

[6] H. V. Jansen, N. R. Tas and J. W. Berenschot, in Encyclopedia of Nanoscience and Nanotechnology, Edited H. S. Nalwa, American Scientific Publishers, Los Angeles, vol. 5, **(2004)**, pp. 163-275.
[7] P. Goovaerts, "Geostatics for Natural Resources Evaluation", Oxford University Press Publishers, New York **(1997)**.
[8] A. Bhowmik and P. Cabral, "Statistical Evaluation of Spatial Interpolation Methods for Small-Sampled Region: A Case Study of Temperature Change Phenomenon in Bangladesh", in B. Murgante, O. Gervasi, A. Iglesias, D. Taniar & B. Apduhan (Eds.), Computational Science and Its Applications - ICCSA 2011, Springer Berlin Publishers, Heidelberg, vol. 6782, **(2011)**, pp. 44-59.
[9] C. J. Willmott, "Some comments on the evaluation of model performance", Bulletin of the American Meteorological Society, vol. 63, **(1982)**, pp. 1309–1313.
[10] H. Yan, Geography and Geo-Information Science, vol. 5, **(2003)**, pp. 27.
[11] H.D. Zeng, Geo-information Science, vol. 7, **(2005)**, pp. 82.
[12] D. T. Price, D. W. Mckenney, L. A. Nalder, M. F. Hutchinson and J. L. Kesteven, "A Comparison of Two Statistical Methods for Spatial Interpolation of Canadian Monthly Climate Data", Agricultural and Forest Meteorology, vol. 101, **(2000)**, pp. 81-94.

## Authors

**Yunhua Gu**

Professor Yunhua Gu was born 1965. She received the B.E. degree in the Computer Science and Technology from Shanghai Jiaotong University in 1987, and M.E. degree in the Computer Science and Technology from Southeast University in 2000. Now, she is a professor in the Computer and Software Institute, Nanjing University of Information Science and technology. Her main research interests are computer network, database, GIS systems. She is a member of ACM and a senior member of CCF.

**Lingmei Sun**

Lingmei Sun was born in 1988. She received the B.E. degree in the Computer Science and Technology from Nanjing University of Information Science & Technology (NUIST). Now, she is a master degree candidate of NUIST，major in computer application technology. Her main research areas are database, GIS systems.

**Jin Wang**

Dr. Jin Wang received the B.S. and M.S. degree in the Electronical Engineering from Nanjing University of Posts & Telecomm., China in 2002 and 2005, respectively. He received Ph.D. degree in Ubiquitous Computing laboratory from the Computer Engineering Department of Kyung Hee University Korea in 2010. Now, he is a professor in the Computer and Software Institute, Nanjing University of Information Science and technology. His research interests mainly include routing protocol and algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a member of the IEEE and ACM.