# Modeling Spatial Pattern of Precipitation with GIS and Multivariate Geostatistical Methods in Chongqing Tobacco Planting Region, China

Xuan Wang, Jiake Lv, Chaofu Wei, Deti Xie

# Modeling Spatial Pattern of Precipitation with GIS and Multivariate Geostatistical Methods in Chongqing Tobacco Planting Region, China

Xuan Wang[1,2], Jiake Lv[1,2*], Chaofu Wei[2], Deti Xie[2]

[1] College of computer and information science, Southwest University, Chongqing, 400715, P.R.China
[2] Chongqing key laboratory of digital agriculture, Chongqing, 400716,P.R.China
{wangxuan, lvjk, weicf, xdt}@swu.edu.cn

**Abstract.** Precipitation is important factor affecting vegetation and controlling key ecological processes. In order to quantify spatial patterns of precipitation in Chongqing tobacco planting region, China, under ArcGIS platform, three multivariate geostatistical methods including cokriging, small grid and regression kriging, coupled with auxiliary topographic factors extracted from a 1:100000 DEM   were applied to predict spatial distribution of precipitation for January (the least month), June (the richest month) and the whole year. The results showed that cokriging was the best for prediction precipitation of January, which could explain 58% of the total variation. Small grid simulation with IDW interpolation exhibited higher accuracy for both June precipitation and annual precipitation, which explained 72% and 61% of the total variation respectively. Generally, multivariate geostatistics accounted for most of the spatial variability in mean precipitation and especially could exhibit great improvement for estimating precipitation in areas where topography has a major influence on the precipitation.

**Keywords:** Modeling spatial prediction, GIS, Cokriging, Small grid simulation, Regression Kriging

## 1    Introduction

Precipitation is a very important climatic characteristic used in determining site suitability for agricultural and forest crops[1, 2], and it is also important in parameterizing the habitat of plant species and determining the patterns of vegetation zonatio [3].However, accurate precipitation data only exist for point locations, the meteorological stations, as a result of which values at any other point in the terrain must be inferred from neighboring stations or from relationships with other variables[4]. Therefore, spatial modeling of precipitation is of importance for understanding and supporting agricultural sustainability development.

---

* Corresponding author: Tel:+86-23-13983377663; Fax:+86-23-68367743
  *Email address*: lvjk@swu.edu.cn (J.Lv)

Over the past decades, a number of methods have been proposed for modeling the spatial distribution of precipitation. The simplest approach is Thiessen polygon method which amounts at drawing around each gage a polygon of influence with the boundaries at a distance halfway between gage pairs [5]. Although the Thiessen polygon method is essentially used for estimation of areal rainfall, it has been applied to the interpolation of point measurements [6]. Meanwhile inverse distance interpolation that makes unknown rainfall depth as a weighted average of surrounding values and the weights being reciprocal to the square distance from the unsampled location is also widely applied in many regions [6, 7]. However, for both Thiessen polygon method and inverse distance interpolation weighting method, they do not consider topographic variables such as elevation, altitude or latitude influences on precipitation, which rather is believed to be important factors especially in mountain regions with complex terrain. In order to overcome this deficiency, some researchers develop the relationships between precipitation and a range of topographic variables and used regression analysis and GIS techniques to model rainfall spatial pattern [1, 8-10].But the method heavily depends on the accuracy of regression model and rarely consider spatial relationships among sample points.

Multivariate Geostatistics which is based on the theory of regionalized variables [11], is recently increasingly preferred because it allows one to capitalize on the spatial correlation between neighboring observations to predict attributes values at unsampled locations, and also can be complemented by ancillary attributes [7, 12, 13]. Several authors have shown that the multivariate geostatistical prediction provides better estimates of rainfall than conventional methods or single ordinary kriging.For example Goovaerts[8] used cokriging for incorporating elevation to predict mean monthly rainfall spatial distribution of Algarve, Portugal and compared that with Thiessen polygon, inverse square distance as well as ordinary kriging ,found that cokriging could achieve better estimation results. Lapen[3] used regression kriging to spatially model the total precipitation normals in the Great Lakers. Hussain[14] interpolated precipitation during monsoon periods in Pakistan. Ouyang [15]incorporated altitude, latitude, elevation, slope, aspect components and used small grid interpolation to calculate annual precipitation distribution of Beijing mountain area, china and the result showed the error tolerance was within 40mm.

Chongqing tobacco planting region, which locates in the southwest China and mostly is covered by hilly and mountains. Understanding the spatial pattern of precipitation especially increasing the accuracy of spatial prediction would be very important for local tobacco planting precise plan and management. In this study, we present an application of cokriging, small grid interpolation, regression kriging respectively for modeling precipitation in January (least), June (most) and the whole year of this region. Topographic variables including elevation, altitude, latitude, slope and aspect were used as auxiliary data. Mean squared error (MSE) and root mean squared standardized effect (RMMSE) as performance evaluation indicators were investigated to determine which approach is appropriate for predicting regional precipitation spatial distribution within different periods.

This paper is organized as follows. Following the introduction, some theory and methods used in this study are described in Section 2 .The study area and data resources are depicted in Section 3. In section 4, results and discussion are illustrated. Conclusions are finally made in Section 5.

# 2  Theory and Method

## 2.1  Cokriging

Multivariate geostatistics comprise a set of techniques and estimators which use the spatial variability, multivariate variables and correlation of a continuous spaced-distributed phenomenon to predict unsampled locations. In this section, we briefly introduce three methods used in this paper.

$$\gamma_{uu}(h) = \frac{1}{2}\left[\{z_u(x) - z_u(x+h)\}^2\right] \tag{1}$$

Where $h$ is a vector, the lag. For $v$ also, the expected difference is zero and its autovariogram is $\gamma_{vv}(h)$. The two variables have a cross-variogram $\gamma_{uv}(h)$, defined as

$$\gamma_{uv}(h) = \frac{1}{2}E\left[\{Z_u(x) - Z_u(x+h)\}\{z_v(x) - z_v(x+h)\}\right] \tag{2}$$

Which describes the way in which $u$ is spatially related to $v$.

To compute the usual cross-variogram, there must be sites where both $u$ and $v$ have been measured. The experimental cross-variogram $\overset{\wedge}{r_{uv}}(h)$ can be estimated by

$$\overset{\wedge}{\gamma_{uv}}(h) = \frac{1}{2m(h)}\sum_{i=1}^{m(h)}\{z_u(x_i) - z_u(x_i+h)\}\{z_v(x_i) - z_v(x_i+h)\} \tag{3}$$

Where $z_u$ and $z_v$ have been measured at sites $x_i$ and $x_i+h$, and $m(h)$ is the number of pairs of data points separated by the particular lag vector $h$.

The cross-variogram can be modeled in the same way as the autovariogram but there is an added condition. Any linear combination of the variables itself should be a regionalized variable, and its variance must be positive or zero. This is ensured if we adopt the linear model of coregionalization. For any pair of variables $u$ and $v$, the variogram is

$$\gamma_{uv}(h) = \sum_{k=1}^{k}b_{uv}^{k}g_k(h) \tag{4}$$

Where the $b_{uv}^k$ represents the variances, for example the nugget or sill variances. $g_k(h)$ is the spatial autocorrelation function which must be the same for both analyzed variables.

The ordinary punctual cokriging prediction of the primary variable $\hat{Z}_u$ is obtained from the linear sum

$$\hat{Z}_u(x_0) = \sum_{i=1}^{V} \sum_{i=1}^{n_l} \lambda_{il} z_l^{ok}(x_i) \tag{5}$$

Where there are $V$ variables, $l = 1,2,...,V$ of which $u$ is the one to be predicted and the subscript $i$ refers to the sites of which there is $n_l$ in the search neighborhood where the variable $l$ has been measured. The $\lambda_{il}$ represents the weights, which in the case of 'classical' cokriging (Goovaerts, 1997) satisfy

$$\sum_{i=1}^{n_l} \lambda_{il} = \begin{cases} 1, & l = u \\ 0, & l \neq u \end{cases} \tag{6}$$

There are non-bias conditions and the weights $\lambda_{il}$ which minimize the estimation variance of $\hat{z}_u$ for a point $x_0$ are found by solving the kriging system for all $v = 1,2,...,V$ and all $j = 1,2,...n_v$. The weights $\lambda_i$ are inserted into Eq. (5) to estimate $\hat{z}_u(x_0)$.

## 2.2 Small grid simulation

Small grid simulation is based on a linear regression between a target variable such as certain climatic property ($Z$) and a secondary or third variable such as elevation or slope ($Y_i$). The regression model so obtained is used to predict $Z$ to the locations of the prediction grid at which $Y_i$ is known. The residuals from the regression $\varepsilon$ are interpolated deterministically to the prediction grid. Since there is still no any deterministic interpolation proved to fit all variables. In this paper, we used IDW and Spline respectively and determined the proper one by calibration and comparison. The predicted values $\hat{z}_R$ and the interpolated values of the residuals $\hat{\varepsilon}$ are summed to give the predicted values of the target variable $\hat{z}$.

$$\hat{Z}(x) = \hat{Z}_R(x) + \hat{\varepsilon}(x) \qquad (7)$$

## 2.3 Regression Kriging

Odeh et al [16]described three types of regression kriging: model A, B and C which are the developments of the general theme. In this study, we used model C. Like small grid simulation, the method is also based on a linear regression between a target variable and a secondary or third variable. But the residuals from the regression $\varepsilon$ are kriged stochastically to the prediction grid using the variogram computed from the residuals. The predicted values $\hat{Z}_R$ and the kriged values of the residuals $\hat{\varepsilon}_{ok}$ are then summed to give the predicted values of the target variable $\hat{Z}_{RK}$ . in the author index.

$$\hat{Z}_{RK}(x) = \hat{Z}_R(x) + \hat{\varepsilon}_{ok}(x) \qquad (8)$$

## 2.4 Performance evaluation indicators

In order to evaluate the performance of different multivariate geostatistical methods, mean squared error (MSE) and root mean squared standardized effect (RMSSE) in this paper, are used as performance measure indicators.
  The mean squared error (MSE) is expressed as

$$MSE = \frac{1}{l} \sum_{j=1}^{l} [z_1(x_i) - z_2(x_i)] \qquad (9)$$

The root mean squared standardized effect (RMSSE) is expressed as

$$RMSSE = \sqrt{\frac{1}{1} \sum_{i=1}^{l} [z_1(x_i) - z_2(x_i)]^2} \qquad (10)$$

Where $z_1(x_i)$ is standardized site true value and $z_2(x_i)$ represents standardized site prediction value, $l$ is the number of validation sites. When MSE is more close to 0 and RMSSE more approaches 1, the accuracy of prediction is regarded better.

# 3    The Study Area and Data Source

## 3.1    The study area

The study area is located in the east part of Chongqing between north latitude 28°09′and 32°12′and east longitudes 106°23′and 110°11′(Fig. 1).The landform of this area is dominated by hills, low mountains and medium mountains. Generally, it is undulating with slopes ranging from 0 to 84 °and altitudes within the range from 100 to 2750.92 m. Due to weak transportation and poor economy development, tobacco planting has been one of local dominant crops productions.
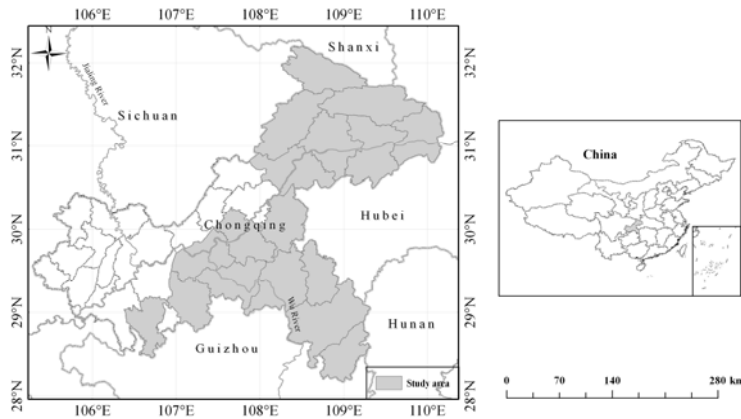


**Fig. 1**. Location of the study area

## 3.2    Data sources

Precipitation data were obtained from 34 meteorology stations of the Chongqing Meteorological Institute monitored monthly from 1977 to 2006, and 13 additional stations belonging to other organizations, irregularly dispersed throughout the whole region. From the observations, 10 sites randomly selected, were kept aside for validation. As precipitation of this region is strongly seasonal (Fig. 2.), and in this study we just model mean monthly precipitation for the least rainfall (January), the richest rainfall (July), as well as the whole year.

A DEM was constructed using ArcMAP Version 9.2 (ESRI).50×50m DEM data were derived from automated image matching of scanned panchromatic aerial photograph based on AUNDEM. Primary topology attributes including elevation, slope and aspect. Latitude and longitude were available by Binary ACCII interpolation.
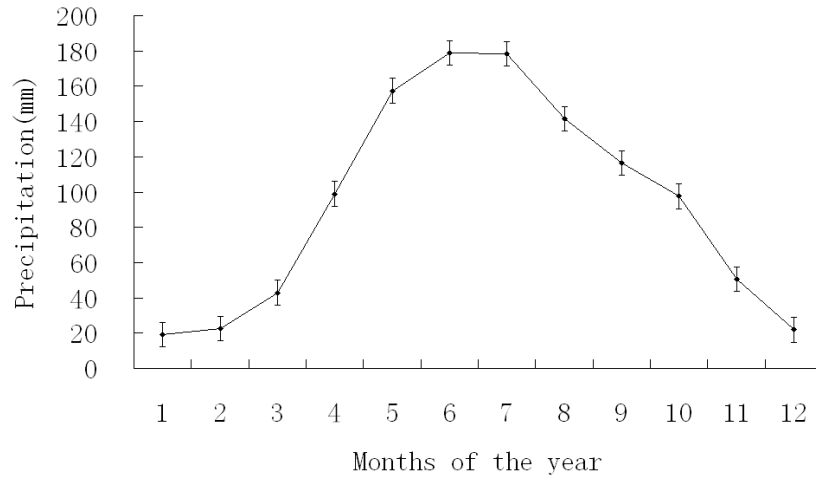
**Fig. 2. Mean monthly precipitations of the stations over past thirty years in study area**

## 4    Results and Discussion

### 4.1    Exploratory data analysis

The Pearson correlation coefficients were calculated between precipitations and the independent topographic variables to determine the influence of different terrain elements on rainfalls. It was shown in Table 1.For precipitations within three different periods, they were all related to elevation and aspect. And precipitation generally increased with elevation, but aspect otherwise inversely. For longitude, only annual precipitation showed strong positive correlation and both precipitation in January and that in July rather showed negative correlation whereas annual precipitation is likely to be indirect. Except precipitation of July, the precipitations of other periods had no obvious links with slope.

**Table 1.** Bivariate correlations between the independent topographic variables and precipitation data within different periods

|  | January | July | Annual |
|---|---|---|---|
| Longitude | -0.086 | 0.233 | 0.461(**) |
| Altitude | -.693** | -.401* | -0.212 |
| Elevation | 0.521** | 0.466* | 0.415* |
| Slope | -0.03 | -0.381* | -0.021 |
| Aspect | -0.471* | -0.416* | -0.423* |

## 4.2　Cokriging

Experimental cross- and auto-variograms were obtained by applying Eq. (4) using the longitude, altitude, elevation, slope and aspect at the observation sites for precipitations. It was shown as in Table 2.

**Table 2.** Variogram model parameters with longitude, altitude, elevation, slope and aspect of different precipitations

| Item | Variable | Model | Nugget | Sill | Range(km) |
|---|---|---|---|---|---|
| | Longitude | Linear | 1.2 | - | - |
| | Altitude | Exponential | 0.7 | 0.9 | 141.16 |
| January | Elevation | Gaussian | 1.2 | 1.2 | 320.60 |
| | Slope | Linear | 1.2 | - | - |
| | Aspect | Spherical | 0.7 | 0.8 | 118.19 |
| | Longitude | Linear | 1.4 | - | - |
| | Altitude | Exponential | 0.7 | 1.1 | 320.67 |
| June | Elevation | Exponential | 0.6 | 0.8 | 117.37 |
| | Slope | Spherical | 1.2 | 1.3 | 340.70 |
| | Aspect | Exponential | 0.6 | 0.8 | 117.37 |
| | Longitude | Gaussian | 0.7 | 0.9 | 116.19 |
| | Altitude | Linear | 0.9 | - | - |
| Annual | Elevation | Exponential | 0.6 | 1.0 | 121.04 |
| | Slope | Linear | 0.7 | - | - |
| | Aspect | Spherical | 1.1 | 1.4 | 326.70 |

It was indicated that the two sets variograms of elevation, aspect for precipitation within different periods were all bounded and were fitted by Gaussian, Spherical and Exponential, Exponential, Exponential as well as Spherical respectively. The variograms of longitude for precipitation of January and June were both unbounded, whereas it was bounded for annual precipitation with the range set to 116.19 km. The variogram of slope was bounded for precipitation of July whereas it was unbounded for both June and Annual. Under the ArcGIS platform, altitude, elevation and aspect were associated as covariances for precipitation of January and altitude, elevation, slope as well as aspect were associated as covariances for precipitation of July. Longitude, elevation and aspect acted as covariances for the spatial prediction of annual precipitation. The whole prediction maps were shown in Fig. 3.
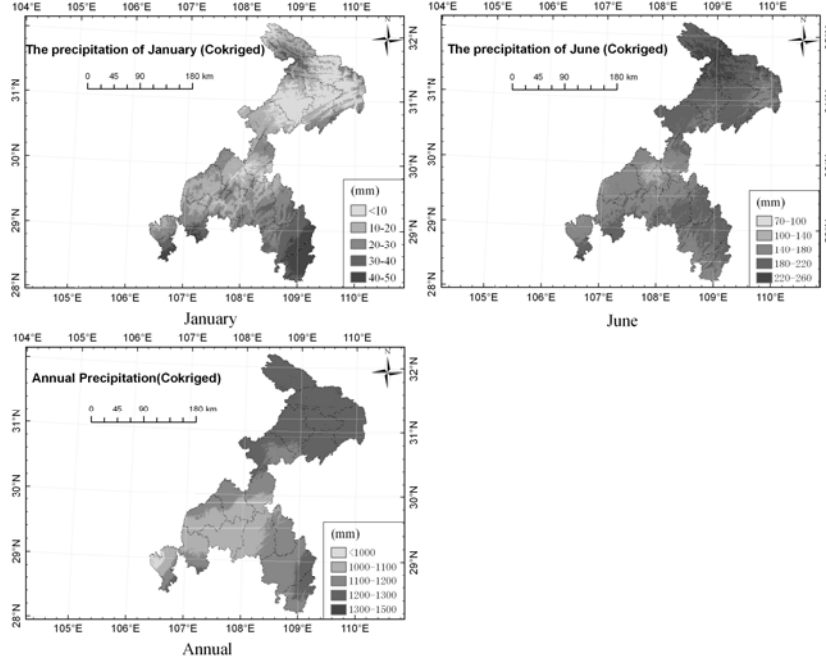
**Fig. 3.** Spatial predictions for precipitations within different periods with cokriging

### 4.3　Small grid interpolation

Linear regressions of precipitations of January, July and Annual were done. The equations were depicted as follows:

$$y_1 = 1.124\alpha - 4.888\beta + 0.008\lambda - 0.167\phi + 0.005\varphi + 41.881 \quad (r^2 = 0.790)$$

$$y_2 = 8.246\alpha - 14.542\beta + 0.009\lambda - 0.339\phi - 0.004\varphi - 272.081 \quad (r^2 = 0.638)$$

$$y_3 = 45.520\alpha - 30.698\beta + 0.118\lambda - 2.938\phi + 0.059\varphi - 287.142 \quad (r^3 = 0.573)$$

Where $y_1$, $y_2$ and $y_3$ represents precipitation of January, July and Annual respectively, $\alpha, \beta, \lambda, \phi, \varphi$ represents longitude, altitude, elevation, slope and aspect.

The residuals of regression models calculations were interpolated by IDW and Spline respectively, and then summarized with regression predicted values. The whole spatial prediction maps were displayed in Fig. 4.
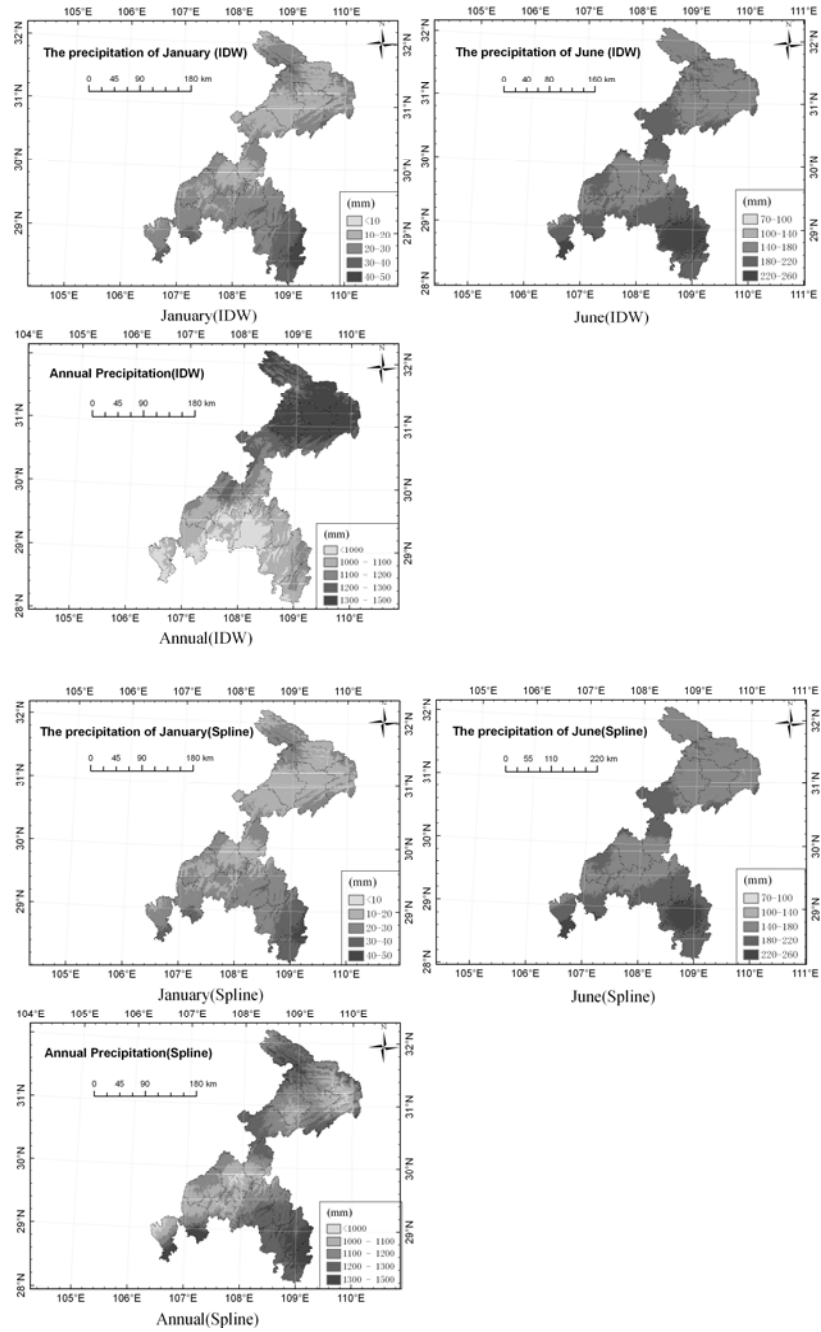
**Fig. 4.** Spatial predictions for precipitations within different periods with small grid simulation

## 4.4    Regression Kriging

Regression modeling is the same with small grid interpolation, and an experimental variogram was computed on the residuals of precipitations within different periods from the regressions at each site. It was shown in Table 3.

**Table 3.** Parameters for the fittest residuals theoretical models of precipitation within different periods

| Item | Model | Nugget | Sill | Range(km) | $R^2$ |
|------|-------|--------|------|-----------|-------|
| January | Spherical | 0 | 0.093 | 411.1 | 0.631 |
|  | Exponential | 0 | 0.079 | 411.1 | 0.652 |
|  | Gaussian | 0.011 | 0.107 | 411.1 | 0.724 |
| June | Spherical | 0.002 | 0.014 | 388.9 | 0.713 |
|  | Exponential | 0.001 | 0.014 | 411.1 | 0.762 |
|  | Gaussian | 0.004 | 0.016 | 358.3 | 0.714 |
| Annual | Spherical | 0.004 | 0.009 | 354.8 | 0.841 |
|  | Exponential | 0.001 | 0.009 | 411.1 | 0.784 |
|  | Gaussian | 0.003 | 0.010 | 301.6 | 0.843 |

It was found that the highest determined coefficient ($R^2$) existed in Gaussian, Exponential and Gaussian model for precipitation of January, June, Annual respectively. The whole spatial prediction maps were displayed in Fig. 5.
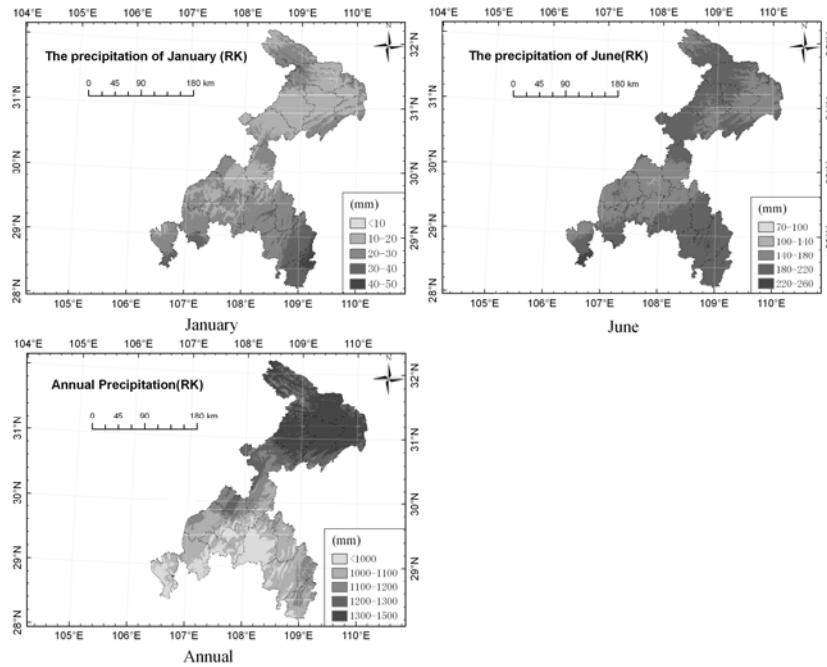
**Fig. 5.** Spatial predictions for precipitations within different periods with regression kriging

## 4.5 Discussion

The prediction residuals and MSEs as well as RMMSEs for the precipitations of January, June and Annual were given in the Table 4 for each method of spatial prediction

In terms of prediction residuals, for precipitation of January, cokriging was the smallest, followed by the small grid (Spline interpolation), small grid (IDW interpolation), and last was the regression kriging. And cokriging could explain 58% of total variation. The MSEs Variation for different methods showed same trend. The RMSSE of being close to 1 was ranked by cokriging, small grid (Spline interpolation), small grid (IDW interpolation) and regression kriging. Fig.3, Fig.4 and Fig.5 have showed the spatial prediction maps with different methods. Generally, the major patterns of variation were evident but the detail was different.

Table 4 showed that small grid (IDW interpolation) was the most accurate method of prediction for precipitation of June, which could explain the 72% of total variation, followed by small grid (Spline interpolation), cokriging. And the regression kriging was the worst in this spatial variability of January, June and Annual mean case. The less accurate predictions from regression kriging probably reflected regression model less good determined coefficient

For annual precipitation, the prediction accuracy rank showed similar pattern with June but the least accuracy method was cokriging and the explained percentage of total variation by small grid (IDW interpolation) simulation was 61%.

**Table 4.** Prediction accuracy at validation points

| Item | | Cokriging | Small Grid(Spline) | Small Grid( IDW) | Regression Kriging |
|------|---|-----------|--------------------|--------------------|--------------------|
| January | TSS[a] | 10947 | 10947 | 10947 | 10947 |
| | SSR[b] | 4598 | 5255 | 5692 | 6568 |
| | MSE | 0.074 | 0.087 | 0.092 | 0.093 |
| | RMSSE | 0.891 | 0.851 | 0.871 | 0.832 |
| June | TSS[a] | 961037 | 961037 | 961037 | 961037 |
| | SSR[b] | 451687 | 374804 | 269090 | 470908 |
| | MSE | 0.034 | 0.031 | 0.027 | 0.072 |
| | RMSSE | 0.911 | 0.952 | 0.973 | 0.897 |
| Annual | TSS[a] | 32921134 | 32921134 | 32921134 | 32921134 |
| | SSR[b] | 17448201 | 15143722 | 12839242 | 16131356 |
| | MSE | 0.065 | 0.051 | 0.042 | 0.057 |
| | RMSSE | 0.872 | 0.897 | 0.925 | 0.887 |

[a] TSS represents the total sum of residuals
[b] SSR represents the sum of squares of residuals

# 5    Conclusion

Understanding the spatial variation of climatic factors is very important for crops planting planning and management. Increasing the accuracy of the spatial predictions of precipitation with the aid of available ancillary data is quiet environmental typically in complex terrain mountain areas, where topographic factors usually have influence on precipitation spatial pattern. This study predicted the spatial distributions of precipitations within different periods with three multivariate geostatistical methods including cokriging, small grid and regression kriging in Chongqing tobacco planting area. And the methods showed good advantage which enabled us to describe 58-72% of the precipitation. However, it should be noted that although ancillary data have been proved to contribute to improving prediction results, there is still no single optimal method for all regions. The coregionalization and the relations between the deterministic components of the variation still should be carefully examined before deciding on the most appropriate methods of prediction.

# References

1. Marquinez, J., Lastra, J., and Garcia, P.: Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. J Hydrol. 270, (2003)
2. Sanchez-Palomares, Rubio, O., and Blanco, A., Elena, A., and Gomez, R.: Parametric Autoecology of beech tree forest in Castilla y Leon. Investigation Agraria: Sistemas de Recursos Forestales. 12, 87--110(2003)
3. Lapen, D.R., and Hayhoe, H.N.: Spatial analysis of seasonal and annual temperature and precipitations normals in Southern Ontario, Canada.Journal of Great Lakers Research.29, 529--544(2003)
4. Michaelides, S., Levizzani, V., and Anagnostou, E., Bauer, P., Kasparis, T., and Lane, J.E.: Precipitation: Measurement, remote sensing, climatology and modeling. Atmos Res. 94, 512--533(2009)
5. Thiessen, A.H.: Precipitation averages for large areas, Monthly Weather Rev. 39,1082--1084(1911)
6. Dirks, K.N., Hay, J.E., and Stow, C.D., and Harris, D.: High-resolution studies of rainfall on Norfolk Island Part II: interpolation of rainfall data. Journal of Hydrol. 208, 187--193(1998)
7. Carrear-Hernandez, J.J., and GASKIN, S.J.: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. Journal of Hydrol.336, 231--249(2007)
8. Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J Hydrol. 228, 113--129(2000)

9. Ninyerola, M., Pone, X., and Roure, J.M.: A methodological approach of climatological modeling of air temperature and precipitation thourgh GIS techniques. Int.J.Climatol. 20, 1823--1841(2000)

10. Lloyd, C.D.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, J Hydrol. 308, 128--150(2005)

11. Wackernagel, H.: Multivariate Geostatistics 2nd edition. Springer, Heidelberg(1998)

12. Els, V., Van Lancker Vera, and Van Meirvenne Marc: Multivariate geostatistics for the predictive modeling of the surficial sand distribution in shelf seas. Continental Shelf Research. 6, 2454--2468(2000)

13. Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, J Hydrol.32, 144--157(2007)

14. Hussain, I., Spöck, G., and Pilz, J., and Yu, H.: Spatio-temporal interpolation of precipitation during monsoon periods in Pakistan.Adv Water Resour.33, 880--886(2010)

15. Ouyang, Z., Xingping, Z., and Youzhong, Z., and Lianqiang, Z.: A small grid method for local climate calculation in mountain area. Transactions of the Chinese Society of Agricultural Engineering, 12,144--148(1996)

16. Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J.: Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging.Geoderma. 67, 215--226(1996)