ISSN 1009－3095 *Journal of Zhejiang University SCIENCE* V. 3，No. 5，P. 584－590，Nov. －Dec.，2002
http://www.periodicals.com.cn; http://www.zju.edu.cn/English
http://www.zjupress.com; http://lib.zju.edu.cn/eindex.htm; jzu_s@mail.hz.zj.cn

Comparison of semivariogram models for Kriging monthly rainfall in eastern China

TANG Yan-bing(汤燕冰)

(*Department of Earth Science, Zhejiang University, Hangzhou 310028, China*)

E-mail: y_tang@css.zju.edu.cn

Received Nov. 28, 2001; revision accepted Apr. 28, 2002

Abstract: An exploratory spatial data analysis method (ESDA) was designed Apr. 28, 2002 for kriging monthly rainfall. Samples were monthly rainfall observed at 61 weather stations in eastern China over the period 1961－1998. Comparison of five semivariogram models (Spherical, Exponential, Linear, Gaussian and Rational Quadratic) indicated that kriging fulfills the objective of finding better ways to estimate interpolation weights and can provide error information for monthly rainfall interpolation. ESDA yielded the three most common forms of experimental semivariogram for monthly rainfall in the area. All five models were appropriate for monthly rainfall interpolation but under different circumstances. Spherical, Exponential and Linear models perform as smoothing interpolator of the data, whereas Gaussian and Rational Quadratic models serve as an exact interpolator. Spherical, Exponential and Linear models tend to underestimate the values. On the contrary, Gaussian and Rational Quadratic models tend to overestimate the values. Since the suitable model for a specific month usually is not unique and each model does not show any bias toward one or more specific months, an ESDA is recommended for a better interpolation result.

Key words: Kriging, Semivariogram model, Monthly rainfall, Eastern China
Document code: A **CLC number:** P468.0

INTRODUCTION

As one of the spatial interpolation methods, Kriging has proved useful and popular in many fields (Burrough, 1987; Li et al., 2000; Oliver, 1990; Robeson, 1997). In climate research it had been used, for example, to interpolate observed climate fields (Ali et al., 2000; Antonic et al., 2001; Biau et al., 1999; Goovaerts, 2000; Jarvis et al., 2001; Merino et al., 2001).

Previous studies (Biau et al., 1999; Goovaerts, 2000) mainly focused on using Kriging to interpolate the spatial distribution of annual rainfall. Since the variation of rainfall distribution, temporally and to an even greater degree spatially, depends strongly on the physical geographic settings such as latitude and the distribution of land and ocean, the models of semivariogram, which are proper for the interpolation of rainfall, are quite different. What kinds of semivariogram models are proper for interpolating monthly rainfall fields? Are there any significant differences in the performance of the models that are suitable for interpolating monthly rainfall distribution? These are the questions this study will hopefully provide the answer as helpful reference for proper use of kriging to interpolate monthly rainfall.

METHODS AND DATA

Different from other mathematical interpolating methods, kriging is essentially a method of estimation by local weighted averaging. The optimal interpolation weights are determined by the semivariogram model that fits the data well (Burrough, 1987; Zhang et al., 1995). Owing to the complexity of spatial variations of different properties, there are several different kinds of theoretical semivariogram models (Li et al., 2000) which mathematically specify the spatial variability of some data sets. Therefore, to choose an appropriate model, the form of the semivariogram is clearly very important for accurate interpolation. In order to obtain the proper models with adapted parameters for interpolating monthly rainfall, an exploratory spatial data analysis (ESDA) method was designed and performed*. The ESDA includes the following steps:

1. Computing the sample (or experimental) semivariogram

Using Eq. 1 to compute experimental semivariogram from the data under study is the only certain way to describe how semivariance changes with distance, determine which semivariogram model should be used. By changing h, both in distance and direction, a set of the sample (or experimental) semivariograms for the data is obtained (Burrough, 1987).

$$\hat{r}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{z(x_i) - z(x_i + h)\}^2 \quad (1)$$

For examining the anisotropy, the sample semivariograms of monthly rainfall observations under study are computed in six different directions, beginning at 0° (along the positive X axis) to 180° (along the negative X axis) counterclockwise angular steps of 30°.

2. Fitting the sample semivariogram model with different theoretical models

Five types of theoretical models, Linear, Spherical, Exponential, Gaussian and Rational Quadratic, are selected to fit the sample semivariogram obtained in the last step. With one exception, Linear model, they are all bounded models (Oliver et al., 1990). Model fitting is carried out using one of the most applied methods, least squares approximation. The initial values for the parameters of the selected models are set to one, the starting point. Through the fitting process, a better set of the parameter values for the selected model is determined.

3. Interpolating the sample field

After determining the suitable values of the parameters, the monthly rainfall fields can be interpolated. The interpolation result is presented as a grid with evenly spaced rows and columns. As a direct by-product, an estimation standard deviation grid with the same size as the interpolated grid is generated at the same time as the kriging processing.

4. Evaluating the interpolation

To evaluate the interpolation, the following three approaches are recommended:

a) Compute the residuals, the differences between the observations and the estimations at sampled points, thus giving a quantitative measurement of how well the estimations agree with the original observations;

b) Plot the isohyet map based on the estimation grid; then compare it with the isohyet map drawn manually using the original observations to see whether the interpolation is reasonable;

c) Plot the map of the estimation standard deviations (ESD) to show how the errors associated with the interpolated values distribute over the study area.

The data used to perform the analysis are monthly rainfall measured from 61 weather stations in eastern China over a 38 year period, 1961–1998. There are a total 456 samples (12 monthly samples for 38 years). The dots in Fig. 2b show the distribution of the 61 weather stations. The maximum, minimum and average distances to the nearest neighboring stations are approximately 1.49° (about 154 km), 0.43° (about 44 km) and 0.86° (about 89 km) respectively.

RESULTS

ESDA was applied to all 456 samples in order to determine the spatial variation of each monthly rainfall observation under study, so that the most common semivariogram forms of the data in the study area are revealed. Furthermore, the performances of the five models are compared with each other.

Semivariogram forms of monthly rainfall fields

According to the ESDA, the most common patterns of the experimental semivariograms for the data are: convex (Type I), linear (Type II) and concave (Type III), with Type III being the most popular and Type II being the least popular. Among the total of 456 samples, more than half of them fall into Type III, and only 65 are representative of Type II.

As shown in Fig. 1, Type I shows that the

* All computing and mapping tasks involved in the ESDA are performed using Surfer 7.0, a Microsoft Windows application from Golden Software Inc., USA.

semivariance, $\hat{\gamma}(h)$, of the monthly rainfall rises with the increase of the lag h first, then starts to level off or even to fall slightly after maximizing. The semivariogram has a clear range and sill (Burrough, 1987). The shape of Type I implies that the correlation between the sampling points decreases with increasing distance between them. After the distance increases to a certain level, the correlation no longer exists. Type II shows that the experimental curve rises almost linearly with increase of h. The situation follows well the basic hypothesis of the regionalized variable theory (Burrough, 1987). Type III shows contrast to Type I. The values either increase very slowly or even decrease slightly at small lags, which makes the experimental semivariogram concave upward close to the origin, then increase rapidly as h increases. The upward concaving pattern is the usual trend (Oliver et al., 1990). Statistically, Type I and III still obey the basic hypothesis although there are some minor deviations.

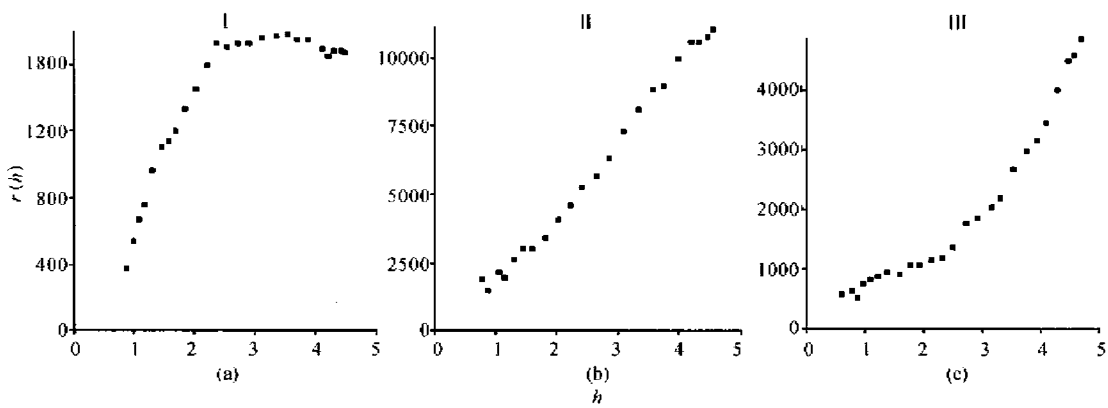

Fig. 1 Three common types of experimental semivariogram forms for the monthly rainfall
(The unit for *h* is degree, and 1 degree is approximately 103km)
(a) convex; (b) linear; (c) concave

Besides the three common types, a few rare semivariogarm types, for instance, 'Λ' shape, 'W' shape and sawtooth shape are observed through the ESDA.

The anisotropy of monthly rainfall fields

The forms of the semivariograms computed and plotted in the six directions having some variation imply that the spatial variation of monthly rainfall in study area is not isotropic. Furthermore, use of AutoFit (a tool provided by Surfer 7.0) yields the "best" fitting anisotropy parameters for the selected models are presented. The auto fitting results reveal that none of the anisotropy ratio values for the data are over 2 (the value of 1 means isotropic), which is considered very mild; and that the anisotropy angles are either 0° – 30° (E-ENE) or 150° – 180° (WNW-W), approximately along X-axis. Since the orientation of the main rain-band in the study area is more zonal than meridional, it makes sense that there is a slightly higher continuity between the sampling points in the east-west direction than in the north - south direction.

Interpolation analysis and model comparison

The five selected models are all suitable for interpolating the data, but under different circumstances. It is worth noting that their performances do not vary for different months in the study area during the period 1961 – 1998. In other words, each selected model tends not to favor one or more specific month.

All selected models but the Linear model work well for Type I, for which August 1994 is a good example. All of the four models fit well the semivariogram. The parameters and kriging estimations for each model are listed in Table 1.

Table 1 shows that all the (other than Gaussian) models are consistent with each other in the ratio of anisotropy. The favorable direction for higher correlation between the data is WNW. Although among the four models, Gaussian model has the lowest residuals and estimation variances; the estimations are the worst, both in estimation range and the spatial distribution pattern displayed in the isohyet map. Therefore, the Gaussian model is not really suitable for the data; and the lowest residuals and errors mean that fitting is only in a mathematical sense. It should be stressed that a good fit does not always guarantee the best prediction.

Table 1 Results of model fitting and data interpolation for August 1994

		Spherical	Exponential	Gaussian	R. Quadratic
Anisotropy	Ratio	2	2	1.2	2
	Angle	166.0	174.5	0	166.5
Range of estimation (mm)		32.2 - 466.2	34.3 - 449.0	-18.7 - 517.8	26.7 - 496.9
Range of residuals (mm)		-11.8 - 28.3	-12.8 - 41.0	-2.0 - 2.3	-3.0 - 5.1
Range of ESD (mm)		13.7 - 102.1	16.9 - 107.5	0.5 - 103.0	2.5 - 103.1

On the other hand, Spherical, Exponential and Rational Quadratic models are suitable for the data. Comparing the results according to the three approaches mentioned earlier, the Rational Quadratic model is relatively the best among the three suitable models although they all provide reasonable predictions. The Rational Quadratic model gives reasonable estimations with minimum residuals and estimation standard deviations (ESD). Fig. 2 shows the isarithmic maps derived from Rational Quadratic estimations and ESD. The distribution pattern of the data displayed by the isohyet map is reasonable. Heavy rainfall occurred on the southern coast and in the northern part of the study area. Areas along the lower reaches of the Changjiang River receive relatively light rainfall. This is quite common for summer in the area. The map of ESD accurately reflects well the distribution pattern of weather stations and shows clearly how estimation errors increase with increased distance from the stations. The areas with smaller ESD values are those around the stations. Severe deviations appear along the edges of the study area and at locations far from the stations. It is obvious that ESD value is proportional to the distance from the stations. Moreover, it should be mentioned that suitable model for a specific month has more than one possibility, just like this example.

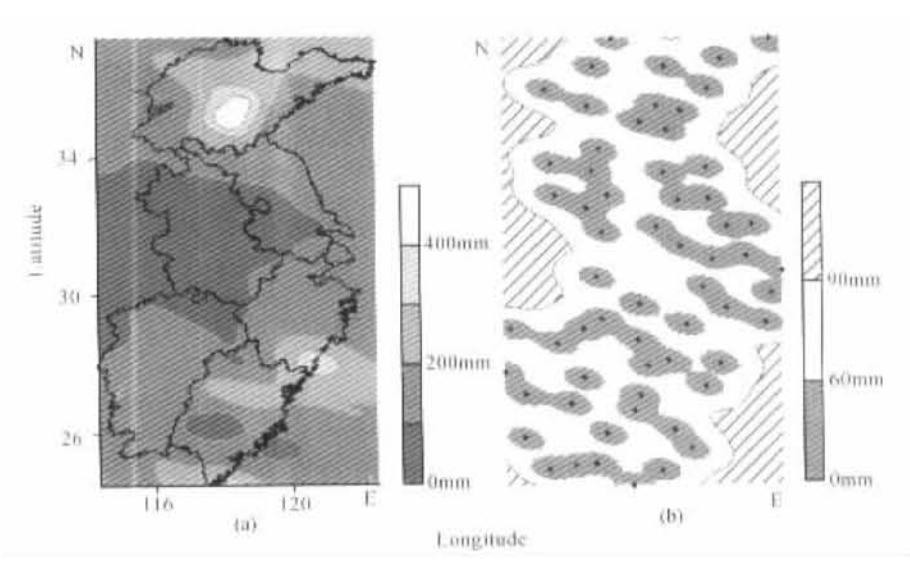

Fig. 2 Isohyet map (a) and the corresponding ESD map (b) for August 1994

Although the Gaussian model does not really work for August 1994, it does work well for some other samples, for example November 1987, which is also an example of Type I but in different season. In summary, all four models are appropriate for Type I. Generally, Spherical and Exponential models tend to underestimate the values and work as smoothing interpolators of the data. Thus, they normally do not perform as well as the other two models for the fields with a few unusually high or low values. On the contrary, Gaussian and Rational Quadratic models usually are exact interpolators of the original data. Although they always produce lower residuals and ESD values than Spherical and Exponential models, they are suitable for a purely mathematical fitting, which occasionally is the case with the other two models. This is probably why good fitting does not always guarantee reasonable estimation. A purely mathematical fitting makes no sense to kriging no matter how precise it is because kriging is based on the regionalized variable theory. So ESDA is an excellent method, if not the only reasonable means, for avoiding a purely mathematical fitting by using kriging properly. It is highly recommended!

For Type II, the Linear model works well. April 1977 is a good example. The Linear model is the only model with the best fitting for the experimental curve. The parameters and estimations are recorded in Table 2 and the maps are shown in Fig. 3.

Table 2 Results of model fitting and data interpolation for April 1977

		Linear values
Anisotropy	Ratio	2
	Angle	177.5
Range of estimation (mm)		18.0 – 385.7
Range of residuals (mm)		−10.3 – 7.9
Range of ESD (mm)		9.0 – 117.9

The pattern revealed in Fig. 3 (a) is quite similar to that drawn manually and shows that the main rain band locates south of 30°N; and that rainfall decreases sequentially northward. The ESD distribution suggests that the estimated values of rainfall at and around the stations are more reliable than those far from the stations.

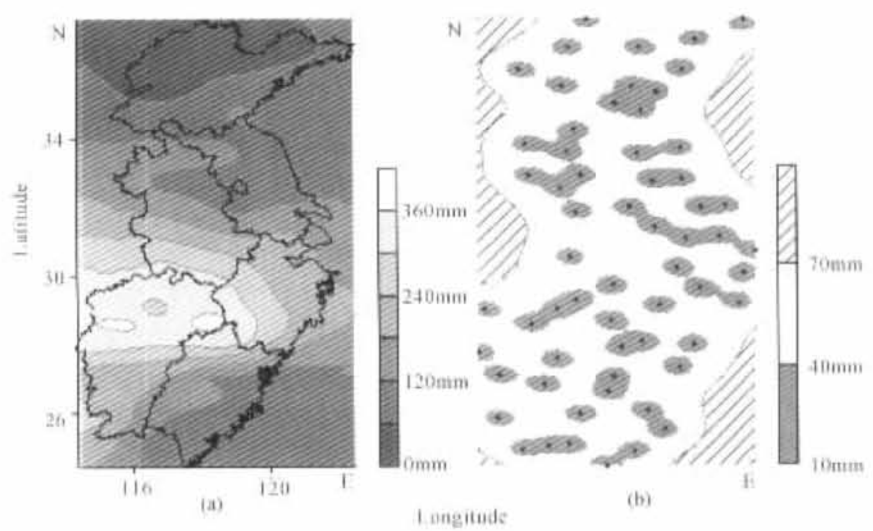

Fig. 3 Isohyet map (a) and the corresponding ESD map (b) for April 1977

Similar to Spherical and Exponential models, the Linear model also tends to underestimate the values, thus producing more smooth estimations. Moreover, it seldom provides a purely mathematical fitting to the data. Normally, a Linear model is used in situations either where the semivariogram does not appear to have a sill, like the last example (April 1977), or when the magnitude of the range far exceeds the distance one wishes to interpolate. Under the latter circumstance, Spherical, Exponential, Gaussian and Rational Quadratic models would work well too, as proved in this study. For example, July 1975 is fitted well using Spherical model.

As stated earlier, Type III is the most popular form of experimental semivariogram for the data used in this study. The concave upward shape of the semivariogram implies proof of local trend in the spatial distribution of the data. The five selected models, especially the Gaussian model, can provide an approximation fitting to the semivariogram derived from the original observations. In order to get better estimations for the data, it is helpful to remove the trend before creating an experimental semivariogram. The detrending options, provided by Surfer, carry out a simple polynomial least squares regression of the data and compute the semivariogram for the resulting residuals only. After detrending, the interpolation can be done either by ordinary kriging (the method used in this study) or universal kriging (Burrough, 1987). For the spatial interpolation of monthly rainfall in the study area, ordinary kriging performs better than universal kriging according to the comparison of the results derived from them respectively.

Fig. 1 (c) is the experimental semivariogram of rainfall in May 1981, an excellent example of Type III. For the raw semivariogram, both Gaussian and Rational Quadratic models seem to provide good fittings. But the results (the column 3 & 4 in Table 3) prove that such fittings are no more than a pure mathematical fitting. Following linear detrending, an Exponential model provides a good fitting to the data. The parameters and estimations are recorded in the second column of Table 3 and both the maps are shown in Fig. 4. The results are reasonable compared with the map drawn manually.

As exemplified above, five selected models are appropriate for kriging monthly rainfall in the area under different conditions. Besides anisotropy, another common point of all models under different conditions is that the nugget

Table 3 Results of model fitting and data interpolation for May 1981

		Exponential	Gaussian	R. Quadratic
Anisotropy	Ratio	2	2	2
	Angle	160.6	171.2	171.2
Range of estimation (mm)		1.3 – 282.3	−909.5 – 796.3	−216.9 – 387.0
Range of residuals (mm)		−7.7 – 6.2	−2.0 – 2.5	−1.8 – 1.1
Range of ESD (mm)		4.6 – 38.3	0.0 – 46.7	0.0 – 52.2

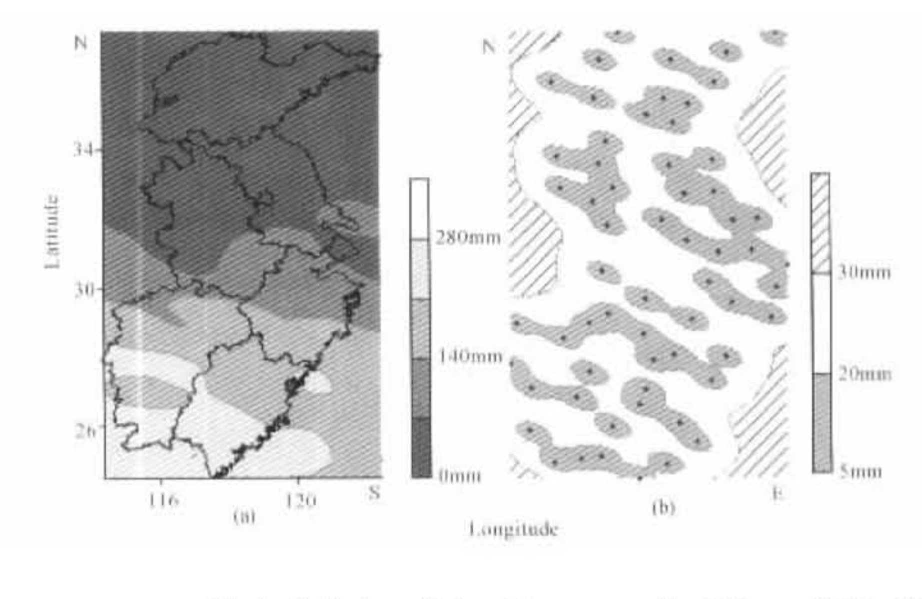

Fig. 4 Isohyet map (a) and the corresponding ESD map (b) for May 1981

effect is insignificant. Comparing the results derived from the same model but with different nugget variances, the best or most reasonable result is always the one without nugget variance. The nugget effect tends to over-smooth the estimations. Therefore, the nugget effect is not applicable to monthly rainfall data in the study area no matter what model is used.

Regarding other rare semivariogram types of monthly rainfall in the study area mentioned before, none of the five selected models performs well, and need to combine or nest different models. This is beyond the scope of this study, and will be investigated later.

CONCLUSIONS

Clearly, kriging fulfills the aim of finding better ways to estimate interpolation weights and provides information on inherent errors for monthly rainfall interpolation in eastern China. Thus, it is profitable to analyze the data using kriging. After application of ESDA to the data and comparing the results, the author reached the following conclusions:

1. Spherical, Exponential, Linear, Gaussian and Rational Quadratic models are appropriate for kriging monthly rainfall in eastern China under certain circumstances.

2. Spherical, Exponential and Linear models perform satisfactorily as smoothing interpolators for the data, but tend to underestimate the values.

3. Gaussian and Rational Quadratic models act as exact interpolators for the data, but tend to overestimate the data. In addition, they must be used with care because they are prone to provide just a purely mathematical fitting to the data.

4. Each model does not show any bias to one or more specific month.

5. For a specific month, the suitable model usually is not unique. So an ESDA is highly recommended for a better result.

6. The anisotropy of monthly rainfall fields in eastern China is mild and the most favorable angles are E-ENE and WNW-W. In addition, the nugget effect is not suitable for spatial interpolation of the data in the study area.

References

Ali, A., Abtew, W., Horn, S., Khanal, N. 2000. Temporal and spatial characterization of rainfall over Central and South Florida. *Journal of American Water Resources Association*, **36** (4): 833 - 848.

Antonic, O., Krizan, J., Marki, A., Bukovec, D. 2001. Spatiotemporal interpolation of climate variables over large region of complex terrain using neural networks. *Ecological Modelling*, **138** (1 - 3): 255 - 263.

Biau, G., Zorita, E., Storch, H., Wackernagel, H., 1999. Estimation of precipitation by kriging in the EOF space of the sea level pressure field. *Journal of Climate*, **12**(4): 1070 - 1085.

Burrough, P. A., 1987. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, p. 147 - 165.

Golden Software, Inc., 1999. Surfer 7.0 Help.

Goovaerts, P., 2000. Geostatistical approaches for interpolating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, **228** (1 - 2): 113 - 129.

Jarvis, C. H., Stuart, H., 2001. A comparison among strategies for interpolating maximum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method, *Journal of Applied Meteorology*, **40** (6): 1075 - 1084.

Li, X., Chen, G., Lu, L., 2000. Comparison of spatial interpolation methods, *Advance in Earth Sciences*, **15** (3): 260 - 265 (in Chinese).

Merino, G. G., Jones, D., Stooksbury, D., Hubbard, K. 2001. Determination of semivariogram models to krige hourly and daily solar irradiance in western Nebraska. *Journal of Applied Meteorology*, **40** (6): 1085 - 1094.

Oliver, M. A., Webster, R., 1990. Kriging: a method of interpolating for geographical information system. *International Journal of Geographical Information System*, **4**(3): 313 - 332.

Robeson, S. M., 1997. Spherical methods for spatial interpolation: review and evaluation. *Cartography and Geographic Information System*, **24**(1): 3 - 20.

Zhang C., Chen B., Wu L., 1995. Geographic Information System. Higher Education Press, Beijing, p. 71 - 79 (in Chinese).