

## Spatial Interpolation of Daily Precipitation in China: 1951–2005

Deliang CHEN<sup>\*1</sup>, Tinghai OU<sup>1,2</sup>, Lebing GONG<sup>3</sup>, Chong-Yu XU<sup>3,4</sup>, LI Weijing<sup>5</sup> (李维京),  
Chang-Hoi HO<sup>6</sup>, and QIAN Weihong<sup>7</sup> (钱维宏)

<sup>1</sup>*Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden*

<sup>2</sup>*School of Geography, Beijing Normal University, Beijing 100875*

<sup>3</sup>*Department of Earth Sciences, Uppsala University, Sweden*

<sup>4</sup>*Department of Geosciences, University of Oslo, Norway*

<sup>5</sup>*Laboratory for Climate Studies/National Climate Center,  
China Meteorological Administration, Beijing 100029*

<sup>6</sup>*School of Earth and Environmental Sciences, Seoul National University, Korea*

<sup>7</sup>*School of Physics, Peking University, Beijing 100871*

(Received 1 September 2009; revised 24 December 2009)

### ABSTRACT

Climate research relies heavily on good quality instrumental data; for modeling efforts gridded data are needed. So far, relatively little effort has been made to create gridded climate data for China. This is especially true for high-resolution daily data. This work, focuses on identifying an accurate method to produce gridded daily precipitation in China based on the observed data at 753 stations for the period 1951–2005. Five interpolation methods, including ordinary nearest neighbor, local polynomial, radial basis function, inverse distance weighting, and ordinary kriging, have been used and compared. Cross-validation shows that the ordinary kriging based on seasonal semi-variograms gives the best performance, closely followed by the inverse distance weighting with a power of 2. Finally the ordinary kriging is chosen to interpolate the station data to a 18 km × 18 km grid system covering the whole country. Precipitation for each 0.5° × 0.5° latitude-longitude block is then obtained by averaging the values at the grid nodes within the block. Owing to the higher station density in the eastern part of the country, the interpolation errors are much smaller than those in the west (west of 100°E). Excluding 145 stations in the western region, the daily, monthly, and annual relative mean absolute errors of the interpolation for the remaining 608 stations are 74%, 29%, and 16%, respectively. The interpolated daily precipitation has been made available on the internet for the scientific community.

**Key words:** daily precipitation, spatial interpolation, ordinary kriging, gridded data, China

**Citation:** Chen, D. L., T. H. Ou, L. Gong, C.-Y. Xu, W. J. Li, C.-H. Ho, and W. H. Qian, 2010: Spatial interpolation of daily precipitation in China: 1951–2005. *Adv. Atmos. Sci.*, **27**(6), 1221–1232, doi: 10.1007/s00376-010-9151-y.

### 1. Introduction

Instrumental climatic observations are important to climate research (Chen et al., 2005; Gao et al., 2007; Gong et al., 2007; Fu et al., 2008) and impact studies (Hulme et al., 1995; Hageback et al., 2005; Song et al., 2006). This is especially true for gridded obser-

vational data that are necessary for model validation (Achberger et al., 2003; Thomas and Herzfeld, 2004). It is usually necessary to perform quality control of the data before they will be used (Feng et al., 2004).

Spatial interpolation methods for monthly precipitation data have been extensively applied to observations in many countries (Price et al., 2000; Chen et

---

\* Correspondence: Deliang CHEN, deliang@gvc.gu.se

al., 2007). In order to utilize geographic information (e.g., elevation, longitude, and latitude) that are correlated with climate data, different kinds of methods have been developed, e.g., regression plus kriging (Johansson et al., 2005), co-kriging (Seo et al., 1990a, b), Gradient plus Inverse Distance Square (GIDS) (Nalder and Wein, 1998), splines (New et al., 1999; Hofierka et al., 2002), optimal interpolation (Sen and Habib, 2000), and neural networks (Antonic et al., 2001). A novel approach applying many features of the topographic data is the Parameter-elevation Regressions on Independent Slopes Model (PRISM model) (Daly et al., 1994). This method has been used in China to interpolate monthly temperature and precipitation data (Zhu et al., 2003; Han et al., 2006). Recently, wind information has also been successfully used for precipitation interpolation in mountainous areas (Johansson and Chen, 2003, 2005). However, it should be kept in mind that these methods are fairly sophisticated, which makes them difficult to apply operationally, although additional collocated geographic and wind information does help improve the quality of interpolation.

While there have been many efforts at producing gridded monthly climatological data (Yan et al., 2005), interpolation and gridding of daily climatic data is rare. Due to the increasing needs of climate studies focusing on extreme events, daily data is becoming more and more important (Liao et al., 2004; Wetterhall et al., 2006). It has been established that on the daily scale, the relationship between precipitation amount and elevation is less clear than at long time scales (Johansson and Chen, 2003). In addition, interpolation of daily precipitation is more difficult compared to the same task for monthly precipitation, because daily precipitation occurs in a discrete pattern both in time and space.

So far relatively little effort has been made to create daily gridded precipitation data for China, while several interpolation-based products have been created for monthly precipitation over some regions in China (Wei and Cao, 1994; He et al., 2005; Cai et al., 2006; Gu et al., 2006; and Liu et al., 2006). Thomas and Herzfeld (2004) used 672 stations during the period 1951–1990 in East Asia to obtain a monthly gridded climate data set for East Asia that includes all of China. In a recent work by Yan et al. (2005), 752 stations were used to obtain a monthly gridding of precipitation in China for the period 1971–2000. Recently, some efforts have been made to interpolate daily precipitation in some regions of China (Zhuang and Wang, 2003; Zhu et al., 2005). However, these studies do not cover all of China. Further, the data lengths used so far are relatively short. An excep-

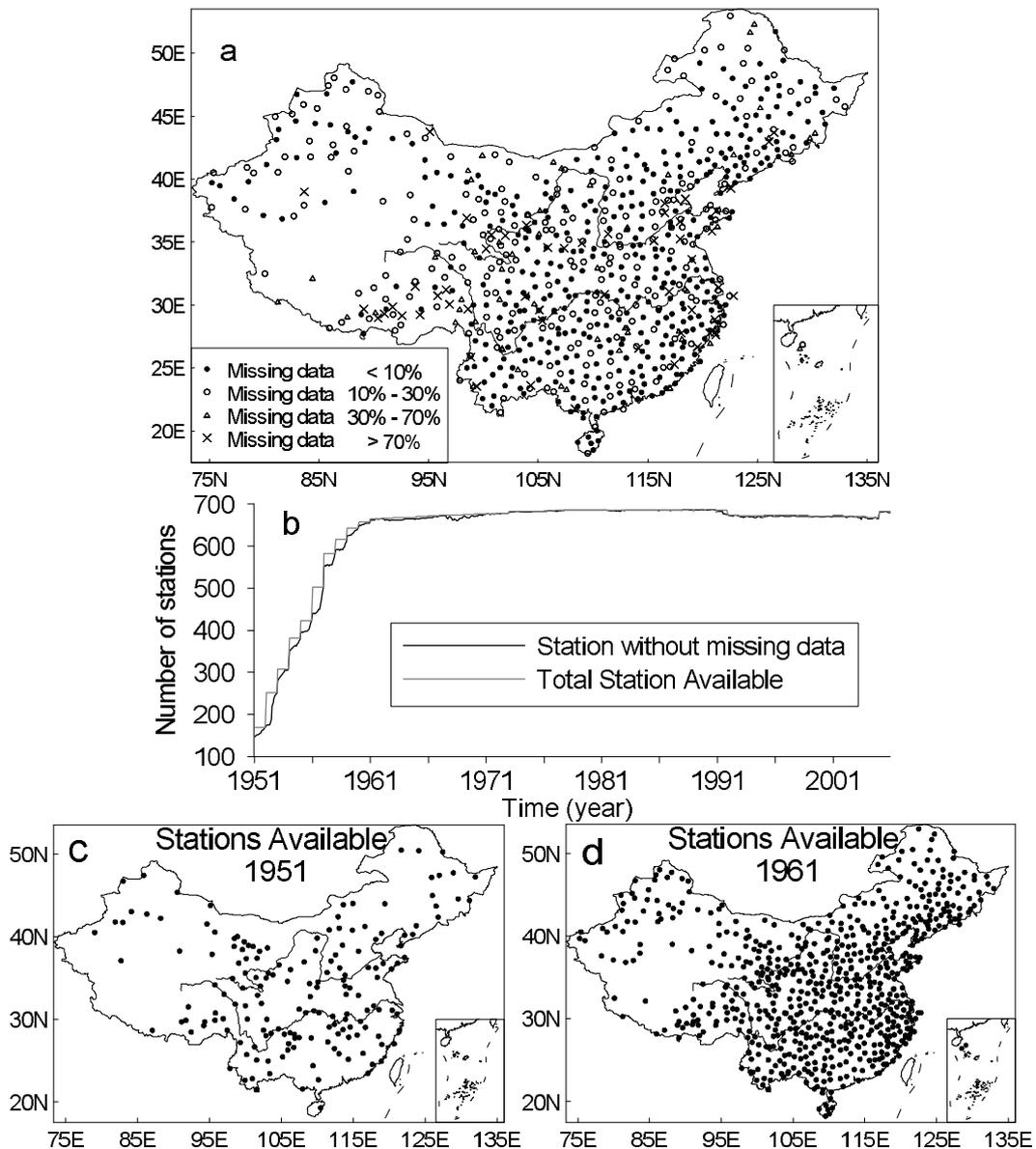
tion is the work by Feng et al. (2004) who generated a  $1^\circ \times 1^\circ$  gridded daily precipitation data set for 1951–2000 covering all of China using the objective analysis method of Cressman (1959). However, the performance of this method was not been tested on daily precipitation data in China before it was used for this purpose. As the interpolation error has not been provided in this case, the interpolation accuracy is not clear for the further application of the gridded precipitation data. A new high-resolution gridded precipitation data set is strongly needed based on newly published daily precipitation data spanning 1951–2005 from 753 stations; the careful selection of an interpolation method is paramount.

This study focuses on the interpolation of daily precipitation data over a long period in China. The overall goal is to find a suitable practical method for creating a high quality gridded data set for China, and to make the product available to the international research community. This method should be simple enough to be implemented in operational use. As has been pointed out by Daly (2006), there is no single method that can work well everywhere, which motivates the approach in this study to test various established methods. Specifically, the objectives of this study are (1) to identify an appropriate interpolation method for daily precipitation in China by comparing several commonly used interpolation methods; (2) to create a gridded data set with high resolution by performing spatial interpolation of daily precipitation in China from 1951 to 2005 using the best interpolation method identified.

## 2. Data and method

### 2.1 Data

Daily precipitation data from a total of 753 national meteorological stations for the period of 1 January 1951 to 31 December 2005 were used for the daily spatial interpolation. The data was provided and quality controlled by the China Meteorological Administration. Quality control of the data is an important part of the study. There are usually two kinds of potential errors to be checked; one is related to outliers and another is linked with homogeneity and consistency. The daily data during 1951–2000 have been checked by Feng et al. (2004) for homogeneity and consistency. In this study we screened the data for outliers, and a few suspicious data were flagged and subsequently treated as missing data. This was done for all stations. The station locations spans from  $16^\circ 32'N$  to  $52^\circ 58'N$ , and from  $75^\circ 14'E$  to  $132^\circ 58'E$  (Fig. 1), and the altitude of the stations varies from 1.1 m above sea level (a.s.l.) to 4800 m a.s.l. The stations cover the



**Fig. 1.** Location of the meteorological stations used in the interpolation: (a) all 753 stations, (b) number of stations in operation varies with time, (c) stations operational from 1951 onward, and (d) stations operational from 1961 onward.

mainland of China; however, the stations are denser and more evenly distributed in the southeast region (Fig. 1a). The mean percentage of missing data during 1951–2005 for the 753 stations is 16.1%. However, stations with a larger percentage of missing data include those that were established before or around 1960 and these are mainly located in eastern part of the country. After 1961, missing data are rare for all the stations. Further, the network's density varied with time (Figs. 1b–1d). There was a steady increase in the number of the stations from 1951–1961 from which time onward the number of the stations has stabilized. Thus, it is important to note that the data coverage is not

homogenous until 1961. It is thus expected that the interpolation quality for 1951–1960 is not as good as during 1961–2005. The average distance between any two stations is 72 km, with a maximal distance of 366 km and a minimal distance of 4 km.

## 2.2 Coordinate transformation

The distance information between stations is required for characterizing the decay of spatial correlations; it also plays an important role in grid interpolation. In some precipitation studies over China, the geographic coordinates in longitude and latitude are used directly for interpolation and mapping (Chen and

**Table 1.** Averaged daily MAE (mm d<sup>-1</sup>) and RMAE (%) (in the bracket) during 1951–2005, based on cross-validation for the eight methods. The bold numbers indicate the best method (smallest error).

	All days	Spring days	Summer days	Autumn days	Winter days
Nearest neighbor	2.00 (94)	1.68 (87)	4.31 (101)	1.54 (96)	0.42 (94)
Local polynomial	1.79 (87)	1.50 (80)	3.87 (91)	1.39 (89)	0.37 (88)
Radial basis	1.87 (90)	1.54 (82)	4.11 (97)	1.43 (92)	0.38 (91)
IDW1	1.74 (83)	1.47 (77)	3.76 (88)	1.34 (85)	0.37 (84)
IDW2	1.71 (82)	1.44 (75)	3.71 (87)	1.31 (83)	0.36 (82)
IDW3	1.71 (82)	1.43 (75)	3.71 (87)	1.31 (83)	0.35 (82)
IDW4	1.72 (82)	1.44 (75)	3.74 (88)	1.32 (84)	0.36 (84)
Kriging	<b>1.70 (82)</b>	<b>1.42 (75)</b>	<b>3.71 (87)</b>	<b>1.30 (83)</b>	<b>0.35 (82)</b>

Gao, 2004; Gemmer et al., 2004), which results in differential treatment of stations in the far south and far north. For example, along the 20°N parallel, 1 degree of longitude is around 100 km, while at 50°N, 1 degree of longitude is only about 70 km. In order to preserve the stations' distance relationships on the curved earth surface, an "equidistant projection" [a map projection in which the distances between one or two points and every other point on the map differ from the corresponding distances on the sphere by only a constant scaling factor (Snyder, 1987)] is used to transform the geographic coordinates into Cartesian ones. A number of different projections, including a cylindrical equidistant projection, an azimuthally equidistant projection, a conic equidistant projection, and a polyconic projection were tested for their performances in terms of equidistant transformation (Fig. 2). The polyconic projection was used to transform the geographic coordinates of the 753 stations since it gave the best result for China (Fig. 2e) and because it is also one of the most common mapping methods used in China.

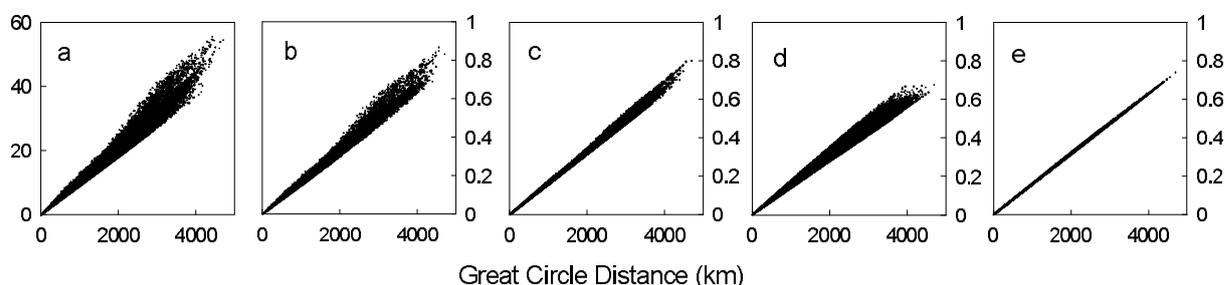
### 2.3 Comparison of interpolation methods

Five commonly used interpolation methods [ordinary kriging (OK), inverse distance weighting (IDW),

radial basis functions, local polynomials, and nearest neighbor] are compared for their interpolation quality. The IDW method was tried with four different powers ranging from 1 to 4, so the method has four variants denoted as IDW1, IDW2, IDW3, and IDW4. The OK method requires a semi-variogram model. A default linear and an exponential model were used. Altogether, eight methods were tested (Table 1).

OK is a geostatistical method that has been widely used in interpolation of climate variables (e.g., Johansson and Chen, 2003). It is a kind of moving average with weights determined by a variogram model that is built based on observations.

The IDW method is also a weighted average interpolation (Franke, 1982), and can be either an exact or a smoothing interpolator. The difference between OK and IDW is that IDW has a predefined weight function while OK uses the observed spatial structure to determine the weights. Data are weighted during interpolation such that the influence of one point relative to another declines with distance from the grid node. When calculating a grid node value, the weights assigned to the data points are fractions, and the sum of all the weights is equal to 1.0. In this paper the smoothing parameter has been set to 0, which indi-



**Fig. 2.** Comparison of the performances of 4 equidistant projections, where the  $x$ -axis is the great circle distance between a pair of stations; the  $y$ -axis is the distance between a pair of stations calculated by the projected coordinates on a Cartesian plane. The following projections were used as the Cartesian coordinates: (a) geographic coordinates, (b) cylindrical equidistant projection, (c) azimuthal equidistant projection, (d) conic equidistant projection, and (e) polyconic projection.

cates no smoothing. The search radius was set to 2800 km with a maximum of 250 stations allowed for use in interpolating a grid node. To find the best functional form of the weight as a function of distance, the power was varied from 1 to 4.

Radial basis function (Carlson and Foley, 1991) interpolation is actually a diverse group of data interpolation methods. All of the Radial Basis Function methods are exact interpolators, so they attempt to honor the data. The basis kernel functions are analogous to variograms in kriging. The basis kernel functions define the optimal set of weights to be applied to the data points when interpolating a grid node. In this paper, the multiquadratic method has been tested. The search radius and the number of stations used for interpolation are the same as in IDW interpolation.

The local polynomial (Yilmaz, 2007) method assigns values to grid nodes by using a weighted least squares fit with data within the grid node's search ellipse. For each grid node, the neighboring data are identified by the user-specified sector search. Using only these identified data, a local polynomial is fit using weighted least squares, and the grid node value is set equal to this value. The polynomial order has been set to 1 and the power is 2 in this study. The search radius and the number of stations used for interpolation are again the same as in IDW interpolation.

The nearest neighbor (Yanalak, 2003) algorithm simply selects the value of the nearest point, and does not consider the values of other neighboring points at all. This method is useful and effective when data are already evenly spaced and there are only a few missing values.

#### 2.4 Determination of semi-variogram model

A semi-variogram model is used to determine the weights for each of the nearby stations in the estimation method by OK. It is possible to make a semi-variogram model for each day. However, this is too time-consuming, especially when the method is used operationally. Thus, we decided to use a seasonal semi-variogram model by averaging all of the daily semi-variograms during a season and comparing the results of the two alternatives. Here, the standard definition of the seasons is adopted, i.e., spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). The seasonal models are justified by the strong seasonality of monsoon climate in China.

The procedure is as follows: (1) calculate the semi-variogram models for each day of the last 30 (1971–2000) years; (2) group these daily semi-variogram models into the four seasons; (3) average the semi-variogram models of each season and fit a model for

each season.

#### 2.5 Validation of interpolation results

Cross-validation was used to evaluate the quality of the interpolations (Chen et al., 2007). To quantify the validation results, Mean Absolute Error (MAE), Mean Bias Error (MBE), and correlation coefficient ( $R$ ) are used. MAE and MBE are defined as follows:

$$\text{MAE} = \frac{1}{n} \left[ \sum_{i=1}^n |P_{\text{obs},i} - P_{\text{int},i}| \right] \quad (1)$$

$$\text{MBE} = \frac{1}{n} \left[ \sum_{i=1}^n (P_{\text{obs},i} - P_{\text{int},i}) \right] \quad (2)$$

where  $n$  is the number of the stations or days,  $P_{\text{obs},i}$  and  $P_{\text{int},i}$  the observed and interpolated precipitation values at station or on day  $i'$ . In the first case, the error statistics are defined with respect to space (spatial errors), whereas the statistics in the second case describe the errors with respect to time (temporal errors).

The errors are also normalized with respect to the observed means to make the comparison between different stations and times feasible. Further, the relative errors are denoted as relative MAE (RMAE) and relative MBE (RMBE). RMAE and RMBE are defined as:

$$\text{RMAE} = \frac{1}{n} \frac{\left[ \sum_{i=1}^n |P_{\text{obs},i} - P_{\text{int},i}| \right]}{\left[ \frac{1}{n} \sum_{i=1}^n P_{\text{obs},i} \right]} \quad (3)$$

$$\text{RMBE} = \frac{1}{n} \frac{\left[ \sum_{i=1}^n (P_{\text{obs},i} - P_{\text{int},i}) \right]}{\left[ \frac{1}{n} \sum_{i=1}^n P_{\text{obs},i} \right]} \quad (4)$$

The monthly, seasonal, and annual errors and correlation coefficients describing the spatial errors are obtained by summing the daily values first and then making the average. The temporal error was also calculated for each station. For example, daily temporal error in terms of MAE for a station was first calculated using Eq. (1), where  $n$  is the number of days; monthly and annual MAEs were then calculated by first obtaining the monthly and annual time series of the observed and interpolated precipitation values for the station in question and then applying Eq. (1). Finally, daily, monthly, and annual RMAEs were obtained by dividing the daily, monthly, and annual MAEs with the corresponding observed daily, monthly, and annual mean precipitation amounts, respectively.

## 2.6 Grid interpolation and block construction

A fine grid with 26000 nodes covering the whole country is built up based on the projected Cartesian plane for interpolation. Nodes are evenly distributed to form  $18 \text{ km} \times 18 \text{ km}$  squares. Then, 4500  $0.5^\circ \times 0.5^\circ$  blocks are built in geographic coordinates, and each of the block values was obtained by averaging the values of the grid nodes that fall in that block.

## 3. Results

### 3.1 Semi-variogram model

For the kriging estimation, a semi-variogram model is used to obtain estimates for the weighting parameters. Here we have compared the validation results of the OK methods using daily and seasonal semi-variogram models. The comparison shows that the OK result with daily semi-variogram models does not substantially improve the results of obtained using OK with seasonal models (not shown). Therefore, the seasonal models were used in all the interpolations discussed in the coming sections.

Figure 3 shows the averaged experimental semi-variograms and the fitted models. For the fitted seasonal models, an exponential model  $r(h) = C(1 - e^{-h})$  is found to fit best, where  $C$  is the scale of the semi-variogram and  $h$  is the relative separation/lag distance. The parameters for the four seasons were estimated separately:  $C=59 \text{ mm}^2 \text{ d}^{-2}$ ,  $h = 230 \text{ km}$  for spring,  $C = 170 \text{ mm}^2 \text{ d}^{-2}$ ,  $h = 160 \text{ km}$  for summer,  $C = 50 \text{ mm}^2 \text{ d}^{-2}$ ,  $h = 230 \text{ km}$  for autumn, and  $C = 9 \text{ mm}^2 \text{ d}^{-2}$ ,  $h = 350 \text{ km}$  for winter.

It can be seen that the semi-variograms for spring and autumn are similar while those for winter and summer are substantially different. Precipitation in summer is often much higher than in winter in China due to the dominating summer monsoon in the region (Qian et al., 2003), which indicates that the spatial correlation decreases with increasing precipitation amount. In other words, the spatial precipitation distribution in winter is of large scale, while summer pre-

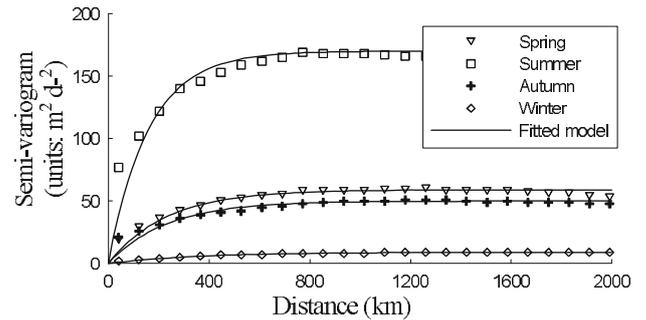


Fig. 3. Seasonal semi-variogram.

cipitation events are much more locally isolated. The semi-variogram models show that the correlation between stations decreases as the distance increases, and the correlation reaches a fairly constant level when the distance is around 400 km. This also has implications for methods other than OK, since a search radius is needed by other methods, too.

### 3.2 Performance of the eight methods

Eight methods were validated for the whole period by cross-validation and the error statistics are summarized in Tables 1–3. It is obvious that the differences in the performance of the eight methods are large.

In terms of MAE and RMAE, the best method is OK (Table 1). It is obvious that the performances of the IDW methods are very close to that of OK, and the differences between the two methods are very small (Tables 2 and 3). In general, all the methods show a very small bias as expressed by MBE and RMBE (Table 2). The biases from the OK method are almost always the smallest and are very close to the theoretical no-bias constraints. In terms of correlation in time, the kriging method and the IDW2 method share the top position in the performance list. Therefore, it is concluded that kriging performs the best while IDW2 is practically as good as kriging.

Because the OK and IDW methods rank first and second in terms of the error statistics, respectively, the cross-validation results for spatial errors averaged over

Table 2. Same as Table 1 but for the averaged daily MBE ( $\text{mm d}^{-1}$ ) and RMBE (%) (in the bracket).

	All days	Spring days	Summer days	Autumn days	Winter days
Nearest neighbor	-0.02 (-1)	-0.01 (-1)	-0.04 (-1)	-0.04 (-2)	0.00 (-1)
Local polynomial	-0.01 (-1)	0.00 (-1)	<b>-0.01 (0)</b>	-0.02 (-2)	0.00 (-1)
Radial basis	-0.01 (0)	0.00 (0)	-0.03 (-1)	-0.02 (-1)	0.00 (0)
IDW1	-0.01 (0)	-0.01 (0)	-0.03 (0)	-0.01 (1)	0.00 (1)
IDW2	-0.02 (0)	-0.01 (0)	-0.04 (-1)	-0.02 (0)	0.00 (1)
IDW3	-0.02 (0)	-0.01 (0)	-0.04 (-1)	-0.02 (-1)	0.00 (0)
IDW4	-0.02 (-1)	-0.01 (0)	-0.04 (-1)	-0.03 (-1)	0.00 (0)
Kriging	<b>-0.01 (0)</b>	<b>0.00 (0)</b>	-0.02 (0)	<b>-0.01 (0)</b>	<b>0.00 (0)</b>

**Table 3.** Averaged daily correlation coefficient during 1951–2005, based on cross-validation for the eight methods. The bold numbers indicate the best method (largest correlation coefficient).

	All days	Spring days	Summer days	Autumn days	Winter days
Nearest neighbor	0.57	0.61	0.51	0.56	0.59
Local polynomial	0.63	0.68	0.59	0.62	0.65
Radial basis	0.63	0.68	0.57	0.62	0.66
IDW1	0.66	0.70	0.61	0.65	0.67
IDW2	0.66	0.70	<b>0.62</b>	0.65	<b>0.68</b>
IDW3	0.65	0.70	0.61	0.65	0.67
IDW4	0.65	0.69	0.60	0.64	0.66
Kriging	<b>0.66</b>	<b>0.70</b>	0.61	<b>0.65</b>	0.68

a one month interval during the whole period for the two methods are also plotted in Fig. 4, which shows more detailed information. The figure displays the temporal evolution of the difference between the two methods. It is interesting to note that the differences between the two are systematic and stable over time. Furthermore, all the error statistics are fairly stable over time starting from 1958. The relatively large errors in the early period may be linked to the sparser network before 1961. This clearly demonstrates the need to create a denser network.

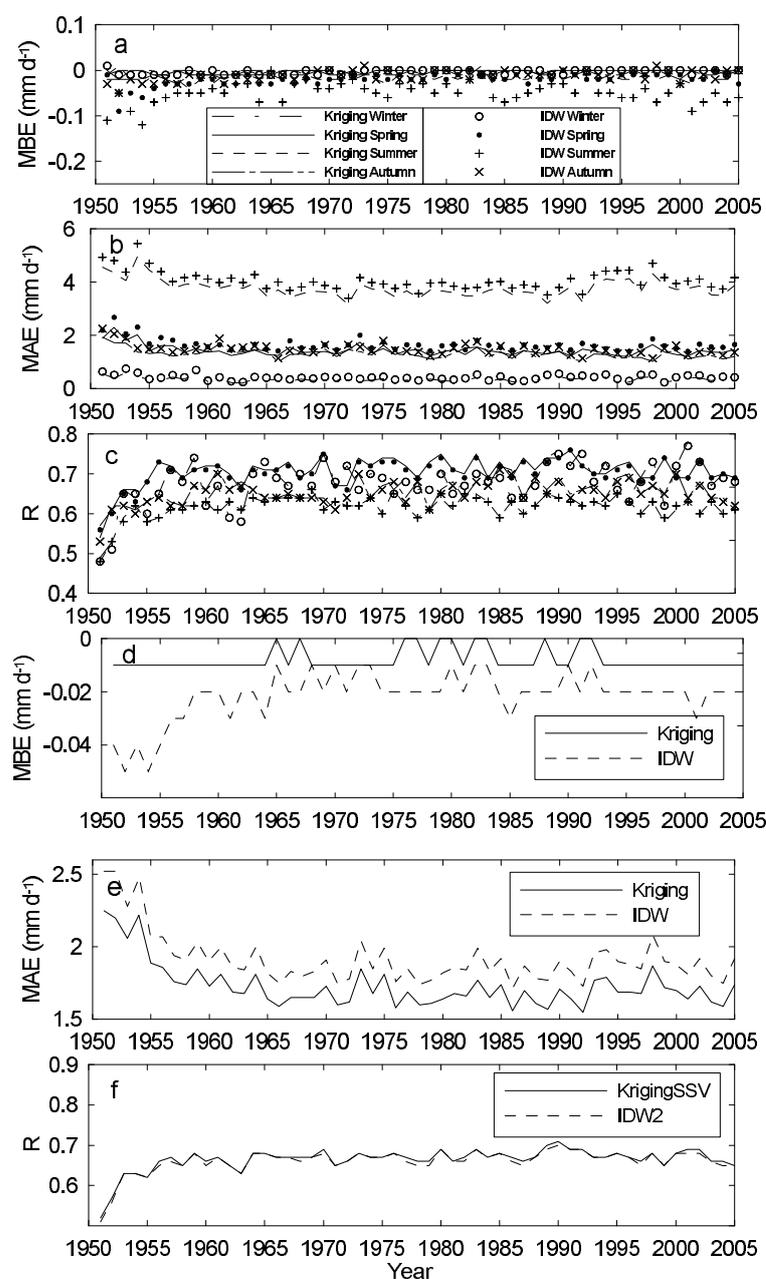
It has long been discussed that kriging is not suitable for monthly precipitation interpolation (Nalder and Wein, 1998), because the assumption of second-order stationarity is usually not met. This is especially true for daily precipitation. However, our result shows that the OK method still has a systematically better performance than IDW2 in most non-stationary cases (Fig. 4).

The fact that the kriging method performs relatively well may be due to the following reasons: (1) although the global precipitation field is not stationary, the local field as defined by the correlation length of the semi-variogram, can be regarded as reasonably stationary; (2) if the spatial field to be interpolated is both non-stationary and non-continuous with high spatial correlation, kriging may indeed be a poor method. However, the daily precipitation is almost discrete, and the correlations between stations are generally weak. As a result, the weights assigned to different stations will not differ much, because no station will fall in the part of the semi-variogram with small lag-distance where the slope of the semi-variogram decreases rapidly. In other words, the interpolation result is most likely not sensitive to the semi-variogram used, which means that differences in the semi-variogram will not cause substantial difference in the weights assigned for the stations even if the semi-variogram can be wrong for some regions under a global non-stationary condition.

### 3.3 Temporal and spatial errors of OK and IDW2

It is now established that the OK method gives the overall best results; this technique was thus chosen for performing the subsequent grid interpolations for the all-China precipitation network data, although the performance of the IDW2 method is not substantially different from the OK method. To understand the error statistics of the OK method and to find possible ways to improve the performance, some analysis of the error statistics is necessary. Figure 5 shows the daily spatial error statistics of the cross-validation for each day during 1951–2005 to identify how the spatial errors vary with spatially averaged precipitation rate and percentage of rainy station over the whole country. The clouds of points in the plots are broad, but a few features still emerge. The maximum error appears to be inversely related to precipitation intensity, which means that the probability of having big interpolation errors under conditions of heavy precipitation is relatively low compared with conditions when there is little precipitation. The range of spatial correlations under conditions with small amounts of precipitation is very large, while the minimum spatial correlation increases with increasing precipitation intensity, indicating the spatial structure is better represented when there is heavy precipitation. Similar results are found with respect to the percentage of rainy stations, which is not surprising since the two factors are well correlated.

To see the above relation more clearly, the monthly mean statistics of the spatial errors are calculated and the results are plotted in Fig. 6. We can see that when the daily mean precipitation amount increases, the mean spatial error first decreases and then increases, with 1–1.5 mm d<sup>-1</sup> or 20%–25% rainy stations as a critical point. This “first decrease and then increase” trend reflects the influence of the two major factors that determine the interpolation quality. One is the percentage of rainy stations: at the non-rainy stations, the estimation error must be a positive er-

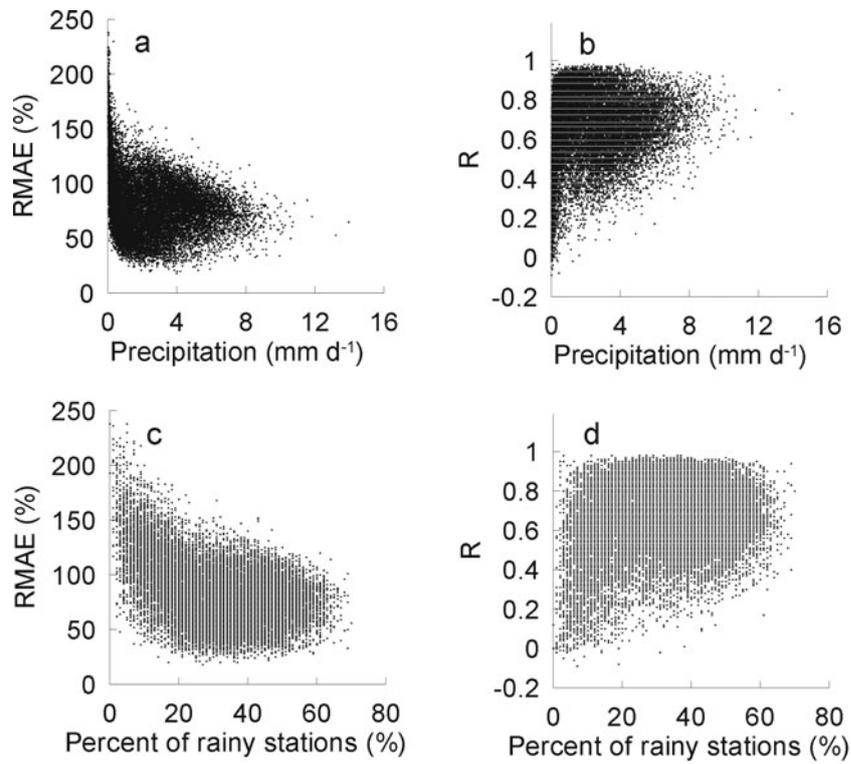


**Fig. 4.** Comparison of seasonal and annual spatial MBE, MAE, and correlation coefficient for OK and IDW2 during 1951–2005.

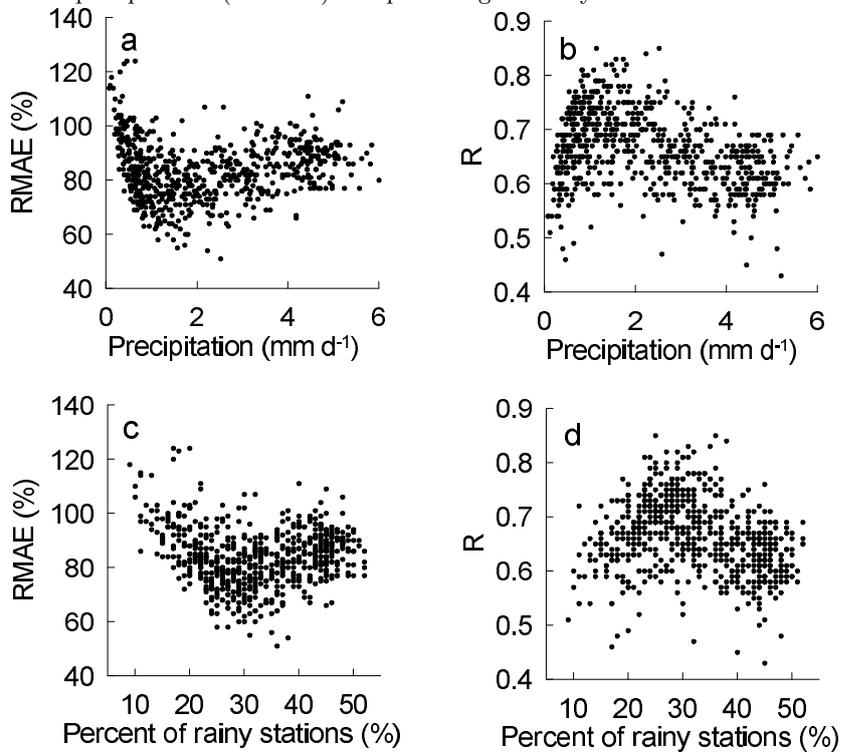
ror, and more positive errors at non-rainy stations will cause a more serious underestimation at rainy stations so that the unbiased rule for kriging is not violated. The second factor is the stationarity of the spatial precipitation field: daily mean precipitation amount is often proportional to the standard deviation of the day's precipitation spatial distribution and therefore proportional to the degree of non-stationarity. Thus, the degree of stationarity will decrease with increasing precipitation amount, which in turn reduces the interpolation quality.

As shown by Fig. 6, when the daily mean precipitation amount is between 0 and  $1 \text{ mm d}^{-1}$ , the first factor, the percentage of rainy stations, will dominate; when the precipitation amount is greater than  $1 \text{ mm d}^{-1}$ , the second factor will dominate. In addition, an increase of precipitation amount will sometime cause a decrease of correlation length as described by the semi-variogram model. Thus, the number of stations that can be used for interpolation will decrease, which also makes the interpolation more challenging.

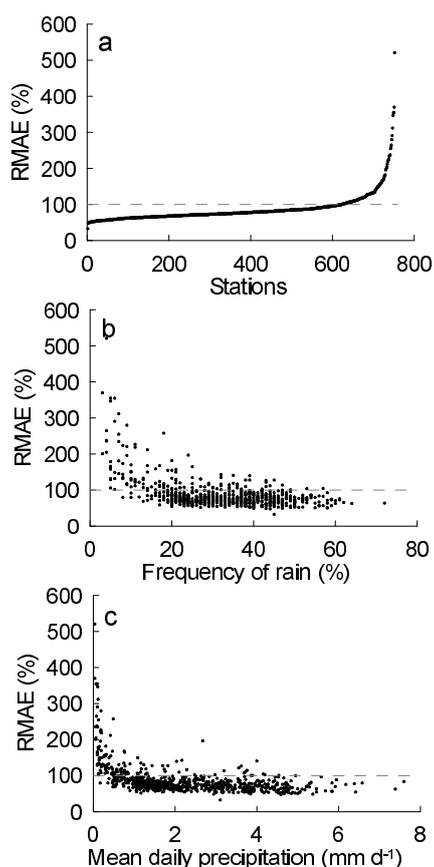
In general, when considering the spatial error, the



**Fig. 5.** Daily spatial error of the interpolation with OK as a function of daily mean precipitation ( $\text{mm d}^{-1}$ ) and percentage of rainy stations.



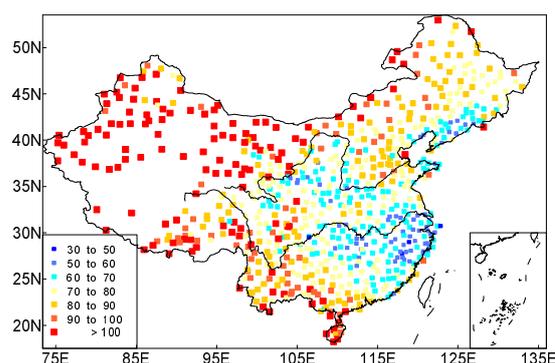
**Fig. 6.** Monthly mean spatial error of the interpolation with OK as a function of monthly mean precipitation expressed in  $\text{mm d}^{-1}$  and percentage of rainy stations.



**Fig. 7.** Mean daily error of the interpolation with OK as a function of stations (a), frequency of rainy days of each station (b), and the mean precipitation of each station (c).

most important influencing factor is the stationarity of the precipitation distribution, as reflected in the mean precipitation amount. The interpolation quality will substantially decrease with the increasing degree of non-stationarity, i.e. with increasing precipitation amounts.

Figure 7 is provided to elucidate how the temporal errors are distributed among the stations and to show how these depend on precipitation intensity and frequency. Compared to the spatial error, the opposite is found for the temporal error where the frequency of rain is the dominant factor (see Fig. 7b vs. Figs. 7b and 7c). As can be seen in the figures, the interpolation quality will substantially increase with increasing precipitation amount, i.e. with increasing frequency of rain (Figs. 7b and 7c). However, the increase of the interpolation quality becomes inconsequential when the mean daily precipitation amount is larger than  $1 \text{ mm d}^{-1}$ , or when the frequency of rain is larger than 20%. This is obviously caused by an underestimation of the semi-variogram scale for the heavy rain region (under a highly non-stationary field).



**Fig. 8.** Spatial distribution of the mean daily temporal errors (RMAE, %) during 1951–2005.

Finally, the spatial distribution of the mean daily errors (RMAE) is plotted in Fig. 8. Due to the relatively sparse station density in the west, the errors in the west are relatively large. If the 145 stations (19% of 753 stations) in the northwest arid area are excluded from the statistics, the daily, monthly, and annual (temporal) RMAEs for the other 608 stations are 74%, 29%, and 16%, respectively. In a similar study carried out in Alberta, Canada (Shen et al., 2001), where 927 stations were available in an area covering 0.638 million  $\text{km}^2$ , the daily RMAE was found to range from 50% to 60%. The daily RMAE shown in our study is a little larger than the error of daily precipitation interpolation in Shen et al. (2001). This is mainly due to the lower density of stations and the larger errors in southwest China where the influence of elevation on the spatial distribution of precipitation needs further investigation (Fig. 8).

#### 4. Conclusions

To meet the need of gridded daily precipitation data for both climate research and model validation, daily observations at 753 stations all over China during 1951–2005 have been used to generate daily gridded precipitation data in China. Eight methods were tested and their performances in terms of interpolation quality were compared by cross-validation. Some important conclusions are listed as follows:

- Among the eight commonly used interpolation methods, OK and IWD2 were ranked highest in terms of interpolation quality of the daily precipitation amount in China, based on the cross-validation during 1951–2005.
- A detailed comparison between the performances of the top two interpolation methods for the period 1951–2005 shows that the ordinary kriging using a seasonal semi-variogram has a systematically better performance than the inverse distance weighting method. However, the difference

in the performance is not substantial.

- Ordinary kriging was chosen to interpolate the station data during 1951–2005 to a 18 km × 18 km grid system covering the whole country. A 0.5° × 0.5° in latitude and longitude grid is then assembled by block interpolation by averaging the grid nodes falling in each block. Owing to the higher station density in the eastern part of the country, the interpolation errors are much smaller than those in the west. Excluding the 145 stations (19% of the total 753 stations) in the western area (west of 100°E), the daily, monthly and annual relative mean absolute errors (RMAE) for the remaining 608 stations are 74%, 29%, and 16%, respectively. The interpolated daily precipitation data set has been made available on the internet at <http://rcg.gvc.gu.se/> for the scientific community.
- Two major factors that influence the interpolation quality are the frequency of rain and the stationarity of the precipitation distribution. Both factors have good correlation with the daily mean precipitation amount. These two factors counteract each other, and their relative importance is controlled by the amount of rain. When the daily mean precipitation amount is greater than 1 mm d<sup>-1</sup>, the non-stationarity of the precipitation distribution will either become the dominant factor and increase the interpolation error (when considering the spatial error), or eliminate the advantage brought about by a high frequency of rain (when considering the temporal error).

**Acknowledgements.** This work is supported by the Swedish Foundation for International Cooperation in Research and High Education through a grant to D. L. Chen. C.-H. Ho is supported by CATER 2006-4204.

## REFERENCES

- Achberger, C., M. L. Linderson, and D. Chen, 2003: Performance of the Rossby Centre regional atmospheric model in southern Sweden: comparison of simulated and observed. *Theor. Appl. Climatol.*, **76**, 219–234.
- Antonic, O., J. Krizan, A. Marki, and D. Bukovec, 2001: Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. *Ecological Modelling*, **138**, 255–263.
- Cai, F., H. B. Yu, L. L. Jiao, K. Tang, H. Q. Ming, and B. Liu, 2006: Comparison of precision of spatial interpolation of precipitation factors: A case study in northeastern China. *Resources Science*, **28**, 73–79. (in Chinese with English abstract)
- Carlson, R. E., and T. A. Foley, 1991: The parameter  $R^2$  in multiquadric interpolation. *Computers & Mathematics With Applications*, **21**, 29–42.
- Chen, D., and G. Gao, 2004: Impact of climate change on runoffs from Han Jiang and Gan Jiang in the Yangtze River basin. *Journal of Lake Sciences*, **15**, 105–114. (in Chinese with English abstract).
- Chen, D., G. Gao, C. Y. Xu, J. Guo, and G. Ren, 2005: Comparison of Thornthwaite method and Pan data with the standard Penman-Monteith estimates of potential evapotranspiration for China. *Climate Research*, **28**, 123–132.
- Chen, D., L. Gong, C. Xu, and S. Halldin, 2007: A high-resolution, gridded dataset for monthly temperature normals (1971–2000) in Sweden. *Geografiska Annaler*, **89A**, 249–261.
- Cressman G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.
- Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, **26**, 707–721.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158.
- Feng, S., Q. Hu, and W. H. Qian, 2004: Quality control of daily meteorological data in China, 1951–2000: A new dataset. *International Journal of Climatology*, **24**, 853–870.
- Franke, R., 1982: Scattered data interpolation: Test of some methods. *Mathematics of Computations*, **33**, 181–200.
- Fu, J., W. Qian, X. Lin, and D. Chen, 2008: Trends of graded precipitation days in China from 1961 to 2000. *Adv. Atmos. Sci.*, **25**, 267–278, doi: 10.1007/s00376-008-0267-2.
- Gao, G., D. Chen, C. Y. Xu, and E. Simelton, 2007: Trend of estimated actual evapotranspiration over China during 1960–2002. *J. Geophys. Res.*, **112**, D11120, doi: 10.1029/2006JD008010.
- Gong, D. Y., C. H. Ho, D. Chen, Y. Qian, Y. S. Choi, and J. Kim, 2007: Weekly cycle of aerosol-meteorology interaction over China. *J. Geophys. Res.*, **112**, D22202, doi: 10.1029/2007JD008888.
- Gu, Z. H., P. J. Shi, and J. Chen, 2006: Precipitation interpolation research over regions with sparse meteorological stations: A case study in Xilin Gole League. *Journal of Beijing Normal University (Natural Science)*, **42**, 204–208. (in Chinese with English abstract)
- Gemmer, M., S. Becker, and T. Jiang, 2004: Observed monthly precipitation trends in China 1951–2002. *Theor. Appl. Climatol.*, **77**, 39–45.
- Hageback, J., J. Sundberg, M. Ostwald, D. Chen, Y. Xie, and P. Knutsson, 2005: Climate variations and land use in Danangou watershed, China—Examples of small-scale farmers adaptation. *Climatic Change*, **72**, 189–212.
- Han, Q. H., W. M. Chen, P. C. Wang, Y. K. Wang, and

- C. J. Yang, 2006: Application of PRISM spatial interpolation to the precipitation distribution in the Songhuajiang river basin. *Journal of Nanjing Institute of Meteorology*, **1**, 56–61. (in Chinese with English abstract)
- He, H. Y., Z. H. Guo, W. F. Xiao, and Q. S. Guo, 2005: Mapping monthly precipitation for Tibetan plateau with GIS and multivariate analysis based in DEM data. *Acta Ecologica Sinica*, **25**, 2933–2938. (in Chinese with English abstract)
- Hoferka, J., J. Parajks, H. Mitasova, and L. Mitas, 2002: Multivariate interpolation of precipitation using regularized spline with tension. *Transactions in GIS*, **6**, 135–150.
- Hulme, M., D. Conway, P. D. Jones, T. Jiang, E. M. Barrow, and C. Turney, 1995: Construction of a 1961–1990 European climatology for climate change modelling and impact applications. *International Journal of Climatology*, **15**, 1333–1363.
- Johansson, B., and D. Chen, 2003: The influence of wind and topography on precipitation distribution in Sweden: statistical analysis and modeling. *International Journal of Climatology*, **23**, 1523–1535.
- Johansson, B., and D. Chen, 2005: Estimation of areal precipitation for runoff modelling using wind data: A case study in Sweden. *Climate Research*, **29**, 53–61.
- Liao, Y., Q. Zhang, and D. Chen, 2004: Stochastic modeling of daily precipitation in China. *Journal of Geographical Sciences*, **14**, 417–426.
- Liu, D. W., Z. M. Feng, and Y. Z. Yang, 2006: Selection of spatial interpolation methods for precipitation in the Haihe River Basin. *Geo-Information Science*, **8**, 75–83. (in Chinese with English abstract).
- New, M., M. Hulme, and P. D. Jones, 1999: Representing twentieth century space-time climate variability. Part 1: development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856.
- Nalder, I. A., and R. W. Wein, 1998: Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, **92**, 211–225.
- Price, D. T., D. W. McKenney, I. A. Nalder, M. F. Hutchinson, and J. L. Kesteven, 2000: A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology*, **101**, 81–94.
- Qian, W. H., D. Chen, Y. Zhu, and H. Y. Shen, 2003: Temporal and spatial variability of dryness/wetness in China during the last 530 Years. *Theor. Appl. Climatol.*, **76**, 13–29.
- Seo, D. J., W. F. Krajewski, and D. S. Bowles, 1990a: Stochastic interpolation of rainfall data from rain gages and radar using cokriging, 1, design of experiments. *Water Resour. Res.*, **26**, 469–477.
- Seo, D. J., W. F. Krajewski, A. Azimi-Zonooz, and D. S. Bowles, 1990b: Stochastic interpolation of rainfall data from rain gages and radar using cokriging, 2, results. *Water Resour. Res.*, **26**, 915–924.
- Sen, Z., and Z. Habib, 2000: Spatial analysis of monthly precipitation in Turkey. *Theor. Appl. Climatol.*, **67**, 81–96.
- Shen, S. S. P., P. Dzikowski, G. Li, and D. Griffith, 2001: Interpolation of 1961–97 daily temperature and precipitation data onto Alberta polygons of ecodistrict and soil landscapes of Canada. *Journal of Applied Meteorology*, **40**, 2162–2177.
- Snyder, J. P., 1987: Map Projections—A working manual. U. S. Geological Survey Professional Paper 1395, U. S. Government Printing Office, Washington, DC, 383pp.
- Song, Y., D. Chen, and W. Dong, 2006: Influence of climate on winter wheat productivity in different climate regions of China during 1961–2000. *Climate Research*, **32**, 219–227.
- Thomas, A., and U. C. Herzfeld, 2004: REGEOTOP: New climatic data fields for East Asia based on localized relief information and geostatistical methods. *International Journal of Climatology*, **24**, 1282–1306.
- Wei, F. Y., and H. X. Cao, 1994: Gridded processing and analyzing for monthly precipitation and temperature data in China. *Meteorological Monthly*, **20**, 26–30. (in Chinese with English abstract).
- Wetterhall, F., A. Bárdossy, D. Chen, S. Halldin, and C. Y. Xu, 2006: Daily precipitation downscaling techniques in different climate regions in China. *Water Resour. Res.*, **42**, W11423, doi: 10.1029/2005WR004573.
- Yan, H., H. A. Nix, M. F. Hutchinson, and T. H. Booth, 2005: Spatial interpolation of monthly mean climate data for China. *International Journal of Climatology*, **25**, 1369–1379.
- Yanalak, M., 2003: Effect of gridding method on digital terrain model profile data based on scattered data. *Journal of Computing in Civil Engineering*, **17**, 58–67.
- Yilmaz, H. M., 2007: The effect of interpolation methods in surface definition: An experimental study. *Earth Surface Processes and Landforms*, **32**, 1346–1361.
- Zhu, H., T. Luo, and C. Daly, 2003: Validation of simulated grid data sets of China's temperature and precipitation with high spatial resolution. *Geographical Research*, **22**, 349–259. (in Chinese with English abstract).
- Zhu, Q. A., W. C. Zhang, and D. Z. Zhao, 2005: Topography-based spatial daily precipitation interpolation by means of PRISM and Thiessen polygon analysis. *Scientia Geographica Sinica*, **25**, 233–238. (in Chinese with English abstract).
- Zhuang, L. W., and S. L. Wang, 2003: Spatial interpolation methods of daily weather data in Northeast China. *Quarterly Journal of Applied Meteorology*, **14**, 605–615. (in Chinese with English abstract).