

Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan

Feng-Wen Chen · Chen-Wuing Liu

Received: 12 October 2011/Revised: 6 February 2012/Accepted: 13 February 2012/Published online: 25 February 2012
© Springer-Verlag 2012

Abstract In this article, we used the inverse distance weighting (IDW) method to estimate the rainfall distribution in the middle of Taiwan. We evaluated the relationship between interpolation accuracy and two critical parameters of IDW: power (α value), and a radius of influence (search radius). A total of 46 rainfall stations and rainfall data between 1981 and 2010 were used in this study, of which the 12 rainfall stations belonging to the Taichung Irrigation Association (TIA) were used for cross-validation. To obtain optimal interpolation data of rainfall, the value of the radius of influence, and the control parameter- α were determined by root mean squared error. The results show that the optimal parameters for IDW in interpolating rainfall data have a radius of influence up to 10–30 km in most cases. However, the optimal α values varied between zero and five. Rainfall data of interpolation using IDW can obtain more accurate results during the dry season than in the flood season. High correlation coefficient values of over 0.95 confirmed IDW as a suitable method of spatial interpolation to predict the probable rainfall data in the middle of Taiwan.

Keywords Inverse distance weighting (IDW) · Spatial interpolation · Rainfall data · Omission

Introduction

Rainfall is a highly significant piece of hydrologic data. Such data are recorded as observational data through comprehensively designed rainfall station networks. However, rainfall records are often incomplete because of missing rainfall data in the measured period, or insufficient rainfall stations in the study region. To resolve the problems of such partial rainfall data, probable rainfall data can be estimated through spatial interpolation techniques.

Various spatial interpolation techniques have already been employed in related fields. Such techniques can be divided into geographical statistics and non-geographical statistics. Examples include nearest neighbor (NN), Thiessen polygons, splines and local trend surfaces, global polynomial (GP), local polynomial (LP), trend surface analysis (TSA), radial basic function (RBF), inverse distance weighting (IDW), and geographically weighted regression proposed by Fotheringham et al. (2002), which are all classified as non-geographical statistics. On the other hand, various forms of Kriging method are classified as geographical statistics (Lam 1983; Jeffrey et al. 2001; Price et al. 2000; Li and Heap 2008; Yeh et al. 2011).

As this article aims to discuss the spatial interpolation of rainfall, the research literature carried out by Naoum and Tsanis (2004) on the interpolation of rainfall in the island of Crete, Greece was reviewed. The research had developed a new Geographical Information System (GIS)-based spatial interpolation module that adopts a multiple linear regression (MLR) technique. This technique can be compared with other methods, such as `splin_regularized`, `spline_tension`, IDW, kriging, and second-order polynomial. When estimating precipitation at ungauged locations, the MLR models provided better estimations than the other spatial interpolation techniques. Li et al. (2006) used the

F.-W. Chen · C.-W. Liu (✉)
Department of Bioenvironmental Systems Engineering, National
Taiwan University, Taipei 10617, Taiwan
e-mail: cwliu@ntu.edu.tw

F.-W. Chen
Agricultural Engineering Research Center, Zhongli 32061,
Taiwan
e-mail: chenfw@aerc.org.tw

annual precipitation over a span of 30 years between 1961 and 1990 from 2114 meteorological stations in China. The data were compared with its respective nearby regions and analyzed through spline, ordinary kriging (OK), and IDW. The result in cross-validation tests shows that the precision of interpolated results are very high. The relative mean errors of three methods were 8.31, 8.76 and 8.76%, respectively ranking as $OK > IDW = \text{spline}$. Similar results of $OK > \text{Spline} > IDW$ (the mean absolute error, MAE are 42.94, 44.79, and 49.86) The study carried out by Chu et al. (2008) reported a similar trend.

Segond et al. (2007) indicated that high spatial and temporal rainfall resolutions are required for urban drainage and urban flood modeling applications. Through IDW, the spatial rainfall field can be obtained when data over a whole catchment are interpolated. When using such method, the results were proven satisfactory as the stimulated data at individual sites preserved properties which mimicked the observed statistics at an acceptable level for practical purposes. Garcia et al. (2008) reported that Multiquadric–biharmonic (MQB) methods surpass IDW methods in terms of interpolation accuracy for both convective and mixed/stratiform events in the study regarding the North American monsoon season over a dense gauge network in the southwestern United States. However, it is found that the order in the IDW method is more important as under certain conditions, results obtained are just as accurate as the MQB method. Dong et al. (2009) used OK, co-kriging (CK), and IDW to interpolate daily precipitation in Qingjiang river basin of China. Daily precipitation data from 36 rainfall stations in June 2006 were analyzed, and the result demonstrated that CK was superior to OK and IDW. The author explained that IDW was unable to take the spatial dependencies of adjacent rain gauges in the basin into account. A similar conclusion was also reached in another case study in Xinjiang, China. The study suggested that the CK method is the most superior compared with RBF, IDW, Kriging methods based on the results of annual precipitation in 75 weather stations from 1995 to 2008 (Zhong 2010). A model composed of OK and entropy with probability distribution function is proposed to relocate the rainfall network and to obtain the optimal design with the minimum number of rain gauges by Yeh et al. (2011).

Nevertheless, several researchers have contrasting views in this area of study. Dirks et al. (1998) compared four spatial interpolation methods using rainfall data from a network of 13 rain gauges on Norfolk Island. The more computationally demanding method of Kriging provided no significant advantage over any of the much simpler IDW, Thiessen, or areal-mean methods. This further indicates that in order to assimilate some characteristics of spatially varying rainfall, the inverse distance method is the

most advantageous for interpolations using spatially dense networks. In another case study, the daily precipitation data from 72 meteorological stations between 1961 and 2000 in Northeast China were analyzed using OK and IDW methods with the weighting of longitude and latitude, and gradient of height plus IDW (GIDW). For daily precipitation, the results showed that the precision of the evaluated value with IDW is greater than OK and GIDW (Zhuang and Wang 2003). Hsieh et al. (2006) used daily summer rainfall records from 20 rain gauges stations between 1990 and 2000 to predict the spatial rainfall distribution in the Shih-Men Watershed in Taiwan using two schemes as OK and IDW. The results indicated that IDW (mean error = 0.04) produced more accurate representations of spatial distribution of rainfall than OK (mean error = 0.54). Kurtzman et al. (2009) aimed at improving the spatial interpolation of daily precipitation for hydrologic models. Different parameterizations of IDW and a local weighted regression (LWR) method were tested in a mountainous terrain in the eastern Mediterranean using 16 years of daily data. The LWR took into account of the weighting factors of elevation which are the explanatory variable and distance, elevation factors, and aspect difference. The IDW interpolation was preferred over the LWR scheme in 27 out of 31 validation gauges. Wu et al. (2010) analyzed and compared five typical interpolation models: IDW, OK, GP, LP, and RBF. The results show that OK and IDW are suitable methods for maximum and minimum precipitations respectively; the results were consistent with the 30 years worth of data in 599 climate stations situated in Texas, US as analyzed by Kong and Tong (2008). A similar research was carried out by Li et al. (2010) utilizing the mean yearly precipitation of 72 meteorological stations from 1971 to 2008 in Zhejiang, China. Different interpolation methods such as the combining stepwise regression and IDW, kriging, spline, and trend were tested. The result demonstrated that the combination of stepwise regression and IDW showed the highest accuracy in prediction, and was better than other methods.

From the above review and comparison of the related researches on spatial interpolation techniques of rainfall, a conclusion can be drawn. According to the comparisons, each method has its advantages and disadvantages based on its objectives, and hence the optimal interpolation method to be adopted varies for different proposals. In general, OK is only suitable for normal distributions; the advantage the IDW method is its usefulness when the distribution of the estimated parameters is not a normal distribution.

In this article, the IDW method is used to interpolate the spatial rainfall distribution in the middle of Taiwan. The choice of the IDW exponent was found to be more significant than the choice of whether or not to use elevation as explanatory data (Kurtzman et al. 2009). In most cases, the

critical influence parameter of IDW is the distance. For this reason, elevation of rainfall stations is not considered in this study. The study aims at improving interpolation accuracy of rainfall using IDW, which is concerned with parameters adjustment including the power (α value) and search radius.

Materials and methods

Study area and data

In this study, the region of the middle of Taiwan was chosen as the main research area. There are 46 rainfall stations distributed in this region. Figure 1 shows the schematic diagram of the rainfall stations spatial distribution in the middle of Taiwan. The rainfall stations are managed by two organizations. The 33 rainfall stations, which are shown as blue points in Fig. 1, are managed by Water Resources Agency (WRA), Ministry of Economic Affairs, while the remaining 13 rainfall stations, which are shown as red points in Fig. 1, are managed by TIA.

For the purpose of using IDW to interpolate spatial rainfall, long-term observed rainfall data were necessary

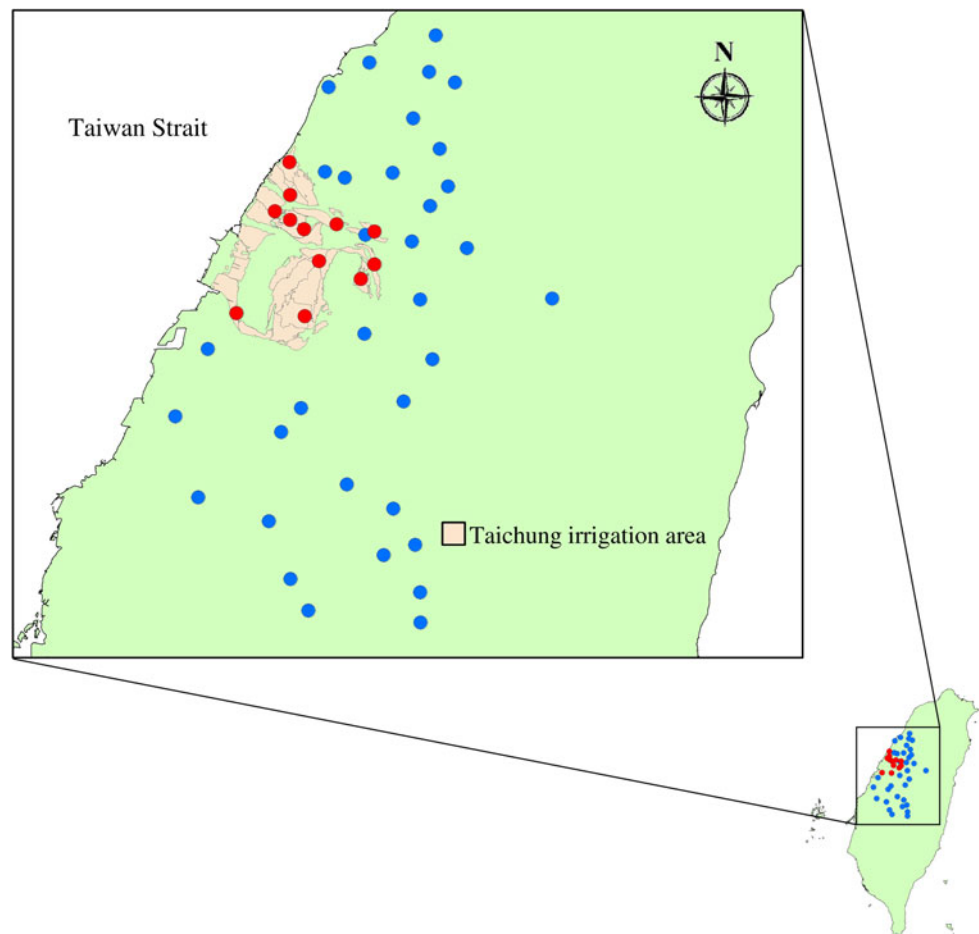
for analysis in the process. Therefore, the daily rainfall data of 30 years from 1981 to 2010 were adopted in this study.

IDW

IDW is based on the concept of Tobler's first law (the first law of geography) from 1970. It was defined as *everything is related to everything else, but near things are more related than distant things*. The IDW was developed by the U.S. National Weather Service in 1972 and is classified as a deterministic method. This is due to the lack of requirement in the calculation to meet specific statistical assumptions, thus IDW is different from stochastic methods (e.g., Kriging and TRA).

The IDW method is also for multivariate interpolation. Its general idea is based on the assumption that the attribute value of an unsampled point is the weighted average of known values within the neighborhood (Lu and Wong 2008). This involves the process of assigning values to unknown points by using values from a scattered set of known points. The value at the unknown point is a weighted sum of the values of N known points. In this study, the IDW method is used to interpolate spatial data,

Fig. 1 Location of 46 rainfall stations in the middle of Taiwan. *Blue and red dots* denote the rainfall stations managed by the WRA and Taichung Irrigation Association, respectively. (Color figure online)



which is based on a concept of distance weighting. It can be used to estimate the unknown spatial rainfall data from the known data of sites that are adjacent to the unknown site (Bedient and Huber 1992; Burrough and McDonnell 1998; Goovaerts 2000; Li and Heap 2008). The IDW formulas are given as Eqs. 1 and 2.

$$\hat{R}_p = \sum_{i=1}^N w_i R_i \quad (1)$$

$$w_i = \frac{d_i^{-\alpha}}{\sum_{i=1}^N d_i^{-\alpha}} \quad (2)$$

where \hat{R}_p means the unknown rainfall data (mm); R_i means the rainfall data of known rainfall stations (mm); N means the amount of rainfall stations; w_i means the weighting of each rainfall stations; d_i means the distance from each rainfall stations to the unknown site; α means the power, and is also a control parameter, generally assumed as two as used by Zhu and Jia (2004) and Lin and Yu (2008), or as six as set by Gemmer et al. (2004). Several researches (e.g., Simanton and Osborn 1980; Tung 1983) have experimented with variations in a power, examining its effects on the spatial distribution of information from precipitation observations. For this reason, α value is conducted in the range of zero to five with an incremental interval value of 0.1 in this article.

Cross-validation

As IDW is the chosen method to interpolate spatial rainfall data for this article, cross-validation is essential to validate critical parameters that could affect the interpolation accuracy of rainfall data. In this case, α value and search radius were evaluated for optimal parameters. This insures the overall utility of the IDW models and enables optimal data prediction that is comparable to the observed data.

Cross-validation, also called rotation estimation, is a technique for assessing how generalized the results of a statistical analysis are with respect to an independent dataset. Common types of cross-validation methods include k -fold cross-validation, twofold cross-validation, repeated random sub-sampling validation, leave-one-out cross-validation (LOOCV), etc. It is mainly used in settings where the objective is to gain a prediction, and estimating how accurately a predictive model will perform in practice (Devijver and Kittler 1982; Seaman 1983; Geisser 1993; Kohavi 1995). Cross-validation has been widely applied in studying the accuracy of prediction methods in precipitation fields. Related research studies include Gyalistras (2003), Feng et al. (2004), Lloyd (2005), Li et al. (2006), Hsieh et al. (2006), Chu et al. (2008), Wang et al. (2008), and Kong and Tong (2008).

Cross-validation can be defined as a method for estimating the accuracy of an inducer by dividing the data into k mutually exclusive subsets (the “folds”) of approximately equal size. The inducer is trained and tested k times. Each time it is trained on the data-set minus a fold and tested on that fold. The accuracy estimate is the average accuracy for the k folds (Cressie 1993; Kohavi and Provost 1998). In general, LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. This is the same as a K -fold cross-validation with K being equal to the number of observations in the original sample. Because of considering tenfold cross-validation is commonly used (McLachlan et al. 2004). Thus, for the cross-validation purpose of this study, from the total of 46 rainfall stations, the 12 rainfall stations that belonged to the Taichung Irrigation Association (TIA), within Taichung irrigation area (Fig. 2) were adopted for this study.

Performance assessment

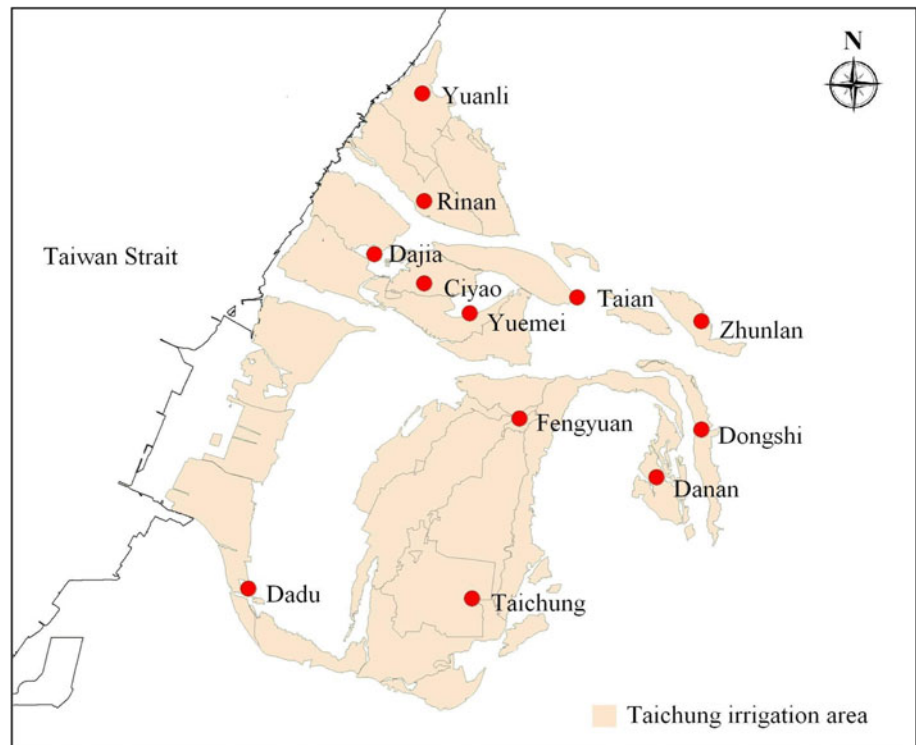
In this study, the root mean squared error (RMSE) was adopted to assess the IDW models performances. The RMSE is also called root-mean-square deviation (RMSD), a measure frequently used on the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. RMSE is a robust measure of accuracy. These individual differences are also called residuals, and the RMSE is served to aggregate them into a single measure of predictive power. The RMSE is applied widely in various fields as follows: in hydrogeology, RMSE is used to evaluate the calibration of a groundwater model; in meteorology, to see how effectively a mathematical model predicts the behavior of the atmosphere; in GIS, the RMSE is one of the measures used to assess the accuracy of spatial analysis and remote sensing. In this study, the RMSE was used to evaluate the optimal model of IDW. At the same time, coefficient of correlation (r) was also used for evaluate whether the estimated data fits observed data. The formulas of RMSE and r are considered by Phogat et al. (2010) and Traore et al. (2010) and given as Eqs. 3–4:

$$\text{RMSE}_i = \sqrt{\frac{1}{n} \sum_{t=1}^n [R_i(t) - \hat{R}_i(t)]^2}; \quad \text{RMSE}_i \geq 0. \quad (3)$$

$$r = \frac{\sum_{t=1}^n (R_i(t) - \bar{R}_i) (\hat{R}_i(t) - \bar{\hat{R}}_i)}{\sqrt{\sum_{t=1}^n (R_i(t) - \bar{R}_i)^2 \sum_{t=1}^n (\hat{R}_i(t) - \bar{\hat{R}}_i)^2}}; \quad (4)$$

$$1 \geq r \geq -1$$

Fig. 2 Distribution of 12 rainfall stations for cross validation



where: $\hat{R}_i(t)$ means spatial rainfall values interpolated using IDW in the unknown rainfall station (i); $R_i(t)$ means observed rainfall data in the unknown rainfall station (i); n means numbers of year (t) adopted, n was equal to 30 in this study.

Analysis procedure

There are 46 rainfall stations used for interpolate rainfall data by IDW models, adopting cross-validation as an appropriate method to assess the accuracy of spatial interpolated rainfall data. 25% of the 46 rainfall stations were selected for a cross-validation process. The selection of rainfall stations was based on the conditions of isotropic and maximum search radius to increase the evaluated groups in a cross-validation process. For this reason, the 12 rainfall stations that are managed by the TIA were selected to conduct the model testing of validity using cross-validation. These 12 rainfall stations were selected in following of Taichung, Zhunlan, Taian, Yuemei, Ciyao, Yuanli, Rinan, Dajia, Danan, Dongshi, Fengyuan and Dadu. Rainfall data required were continuous daily data in the period of 1981–2010 (30 years). Table 1 shows the distance between respective rainfall stations, and is used to calculate the weighting of individual rainfall station to the objective rainfall station. To determine the optimal search radii (O.S.R.) of IDW of the 12 rainfall stations, 11 search radii (10–110 km) were executed (Table 2). The italicized

numbers in Table 2 represent the optimal number of rainfall stations nearby the objective that were selected for daily rainfall interpolation. Figure 3 shows the schematic diagram of different groups in Taichung rainfall station as an example. It illustrates that different rainfall stations within different selected search radii can be used for interpolate rainfall data. For example, 19 and 45 rainfall stations falls within the selected search radius of 30 and 70 km for rainfall data interpolation. We compared and analyzed the data with seven groups, assessing the relationship between prediction accuracy with search radius (the number of selected rainfall stations). This was used for further calculations in cross validation. Finally, RMSE was used for determining optimal parameters: α value and search radius of IDW. This is done through data verification testing which must be conducted subsequent to all procedures of cross-validation. The data verification testing then compared the data interpolated with observed data. To express the applicability of IDW models, the results of rainfall data interpolation using IDW were all determined by r .

Results and discussions

To interpolate the unknown rainfall data, 12 rainfall stations (Table 2) were assumed as unknowns and rainfall data estimated using IDW in a common parameter of $\alpha = 0-5.0$. As the adopted data expressed as daily rainfall, all interpolated data results were also expressed as daily

Table 1 Distance (km) between 46 rainfall stations to the 12 objective rainfall stations for calculation of the individual weighting

Rainfall stations	Taichung	Zhunlan	Taian	Yuemei	Ciyao	Yuanli	Rinan	Dajia	Danan	Dongshi	Fengyuan	Dadu
Taichung	–	23.2	20.6	18.4	20.6	32.8	25.9	23.2	14.3	18.4	12.0	14.5
Zhunlan	23.2	–	8.2	15.0	18.1	23.3	19.5	21.6	10.5	7.0	13.3	34.0
Taian	20.6	8.2	–	7.0	9.9	16.6	11.7	13.4	12.7	11.7	8.7	28.4
Yuemei	18.4	15.0	7.0	–	3.5	14.6	7.8	7.3	16.1	16.7	7.5	22.9
Ciyao	20.6	18.1	9.9	3.5	–	12.3	5.3	3.7	19.6	20.3	10.7	22.8
Yuanli	32.8	23.3	16.6	14.6	12.3	–	7.0	10.8	29.1	28.2	21.9	34.0
Rinan	25.9	19.5	11.7	7.8	5.3	7.0	–	4.7	23.3	23.2	15.3	27.5
Dajia	23.2	21.6	13.4	7.3	3.7	10.8	4.7	–	23.3	24.0	14.2	23.1
Danan	14.3	10.5	12.7	16.1	19.6	29.1	23.3	23.3	–	4.2	9.6	27.4
Dongshi	18.4	7.0	11.7	16.7	20.3	28.2	23.2	24.0	4.2	–	11.8	31.1
Fengyuan	12.0	13.3	8.7	7.5	10.7	21.9	15.3	14.2	9.6	11.8	–	20.7
Dadu	14.5	34.0	28.4	22.9	22.8	34.0	27.5	23.1	27.4	31.1	20.7	–
Baileng	24.6	17.4	23.9	28.8	32.3	40.2	35.4	36.1	13.2	12.2	22.9	39.0
Nanchuang-1	59.0	36.0	39.2	44.7	45.6	39.0	42.4	47.0	46.3	42.3	47.7	67.5
Shanhuhu	65.8	43.7	45.3	49.8	50.0	41.1	45.9	50.6	54.1	50.3	54.0	72.6
Taho	58.2	35.9	37.9	42.7	43.2	35.3	39.5	44.2	46.4	42.5	46.5	65.6
Tatan	55.6	35.9	35.1	38.0	37.5	27.1	32.8	37.4	46.0	42.9	43.5	60.2
Henglungshan	45.7	22.4	27.2	33.5	35.3	32.1	33.3	37.5	32.4	28.2	35.0	55.5
Tahu-1	35.7	13.1	16.2	22.3	24.0	22.0	22.3	26.3	23.5	19.8	24.4	44.6
Hexing	47.9	25.4	27.8	33.1	33.9	27.9	30.8	35.4	35.9	32.1	36.3	55.8
Sanyi-2	30.6	13.1	10.1	14.0	14.7	12.2	12.2	16.5	21.8	19.5	18.6	36.8
Nanshishan	48.9	32.2	29.2	30.6	29.4	18.0	24.3	28.7	41.3	38.9	37.0	51.8
Tapingtien	30.9	16.5	11.4	13.0	12.6	7.8	8.9	13.5	24.0	22.3	19.0	35.4
Shuanychi-2	27.7	8.2	16.4	23.1	26.3	31.0	27.7	29.8	13.5	9.3	20.2	40.2
Hsiangpi-1	35.5	13.1	20.3	27.3	29.9	31.3	29.8	33.0	21.4	17.2	26.4	47.0
Sungan	41.1	18.4	25.0	31.9	34.3	34.1	33.6	37.2	27.0	22.8	31.7	52.4
Choulun-2	21.6	2.0	6.5	13.1	16.3	22.3	18.1	19.9	9.5	6.6	11.4	32.1
Hsuenling	37.4	20.0	28.2	34.9	38.1	41.9	39.2	41.6	23.5	20.0	31.5	50.9
Toupienkeng	13.2	21.8	24.0	25.6	28.9	39.8	33.4	32.2	11.6	14.9	18.2	27.5
Cuiluan	52.7	40.3	48.4	54.7	58.1	62.9	59.8	61.8	40.9	38.4	50.1	67.1
Liufenliao	25.1	46.9	45.7	43.4	45.1	57.3	50.4	46.9	36.6	40.7	37.2	27.0
Tsaotun-4	19.5	40.6	39.7	38.0	40.0	52.3	45.3	42.2	30.2	34.2	31.4	24.4
Peishan-2	27.7	36.6	40.2	42.2	45.4	56.3	50.0	48.8	27.5	29.7	34.8	40.1
Chingliou-1	28.7	29.8	35.2	38.8	42.3	51.7	46.2	45.9	22.8	23.6	31.8	42.8
Toupien	21.7	43.3	38.1	32.6	32.5	43.3	37.1	32.5	35.7	39.7	30.1	9.8
Wanhsing-2	34.8	57.7	53.3	48.2	48.3	59.2	53.0	48.5	49.0	53.2	44.9	25.5
Lungshen Bridge	45.0	58.9	61.5	62.3	65.1	76.8	70.1	68.0	49.2	52.0	54.8	53.2
Chichi-2	36.9	54.0	55.3	54.9	57.4	69.5	62.6	60.0	43.7	47.1	47.8	43.3
Tuntou-2	55.9	75.9	75.9	74.3	76.2	88.5	81.5	78.2	65.4	69.1	67.8	57.6
Wangxiang	63.6	77.2	80.1	81.0	83.8	95.5	88.8	86.6	67.7	70.3	73.5	71.0
Tungpu	69.5	83.5	86.4	87.1	89.8	101.6	94.9	92.6	74.0	76.6	79.7	76.4
Hsiluan	54.0	67.0	70.1	71.1	73.9	85.5	78.9	76.9	57.6	60.1	63.6	62.2
Neimaopu-2	53.5	68.7	71.0	71.3	73.9	85.8	79.0	76.6	58.8	61.8	63.9	60.2
Hsilo-2	44.7	67.7	65.0	61.2	62.0	73.8	67.1	62.9	57.8	62.0	56.4	40.0
Hsiaolin-2	62.5	81.7	82.2	81.0	83.0	95.3	88.3	85.1	71.3	74.8	74.2	65.0
Linne-1	44.2	65.4	64.7	62.5	64.1	76.4	69.4	65.9	55.0	59.0	56.2	44.7

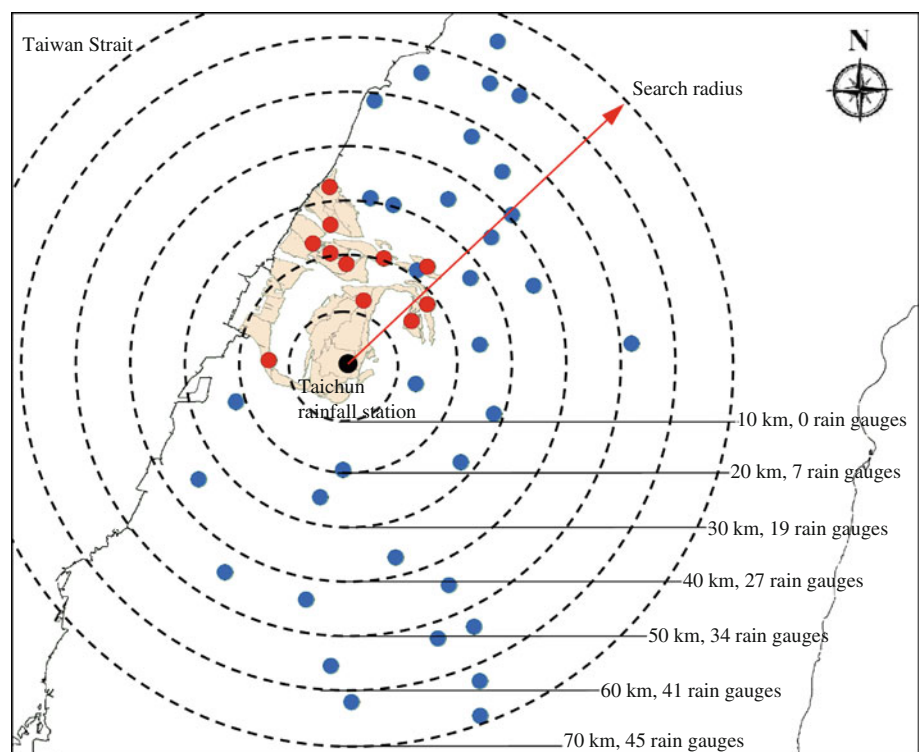
Unit: km

Table 2 Optimal numbers of rainfall stations evaluated by different search radii (10–110 km) of 12 objective rainfall stations

Group search radius	Rainfall stations											
	Taichung	Zhunlan	Taian	Yuemei	Ciyao	Yuanli	Rinan	Dajia	Danan	Dongshi	Fengyuan	Dadu
Group 1 (10 km)	0 ^a	4	5	5	4	2	5	3	3	4	3	1
Group 2 (20 km)	7	16	14	13	11	8	10	9	11	14	13	2
Group 3 (30 km)	19	23	24	19	19	16	18	17	23	23	19	13
Group 4 (40 km)	27	29	30	28	27	25	27	25	28	29	29	21
Group 5 (50 km)	34	34	34	34	34	29	31	33	36	34	34	28
Group 6 (60 km)	41	37	36	36	36	34	35	35	41	38	39	35
Group 7 (70 km)	45	41	39	39	39	36	38	39	43	42	42	42
Group 8 (80 km)		43	42	42	42	39	41	42	45	45	45	45
Group 9 (90 km)		45	45	45	45	42	44	44				
(Group 10 (100 km))						44	45	45				
Group 11 (110 km)						45						
Number of groups	7	9	9	9	9	11	10	10	8	8	8	8
Optimal number of stations	7	4	14	19	4	8	5	9	11	4	45	21

^a No any rainfall stations within the effect radius

Fig. 3 Schematic diagram of rainfall station groups of different search radius—use Taichung rainfall station for example



rainfall data. Every different group of each rainfall station was estimated individually using only the observed data respective to individual rainfall station’s search radius. After interpolation of daily rainfall levels, data were then compiled accordingly to create data expressed as ten-day rainfall, monthly rainfall and annual rainfall.

To identify the optimal parameters of IDW, α value and research radius, and the minimum RMSE were calculated.

Table 3 shows the minimum RMSE in the forms of both monthly rainfall and annually rainfall. At the same time, the O.S.R. and α value were recorded in the condition of the minimum RMSE. Comparing the annual O.S.R. and monthly α values, the results showed two phenomena. Firstly, in a viewpoint of annual O.S.R., 75% O.S.R. were within 10–20 km, there was only one anomaly (80 km) which occurred in Fengyuan rainfall station; it deemed the

Table 3 Optimal parameters of IDW for interpolation of spatial rainfall data

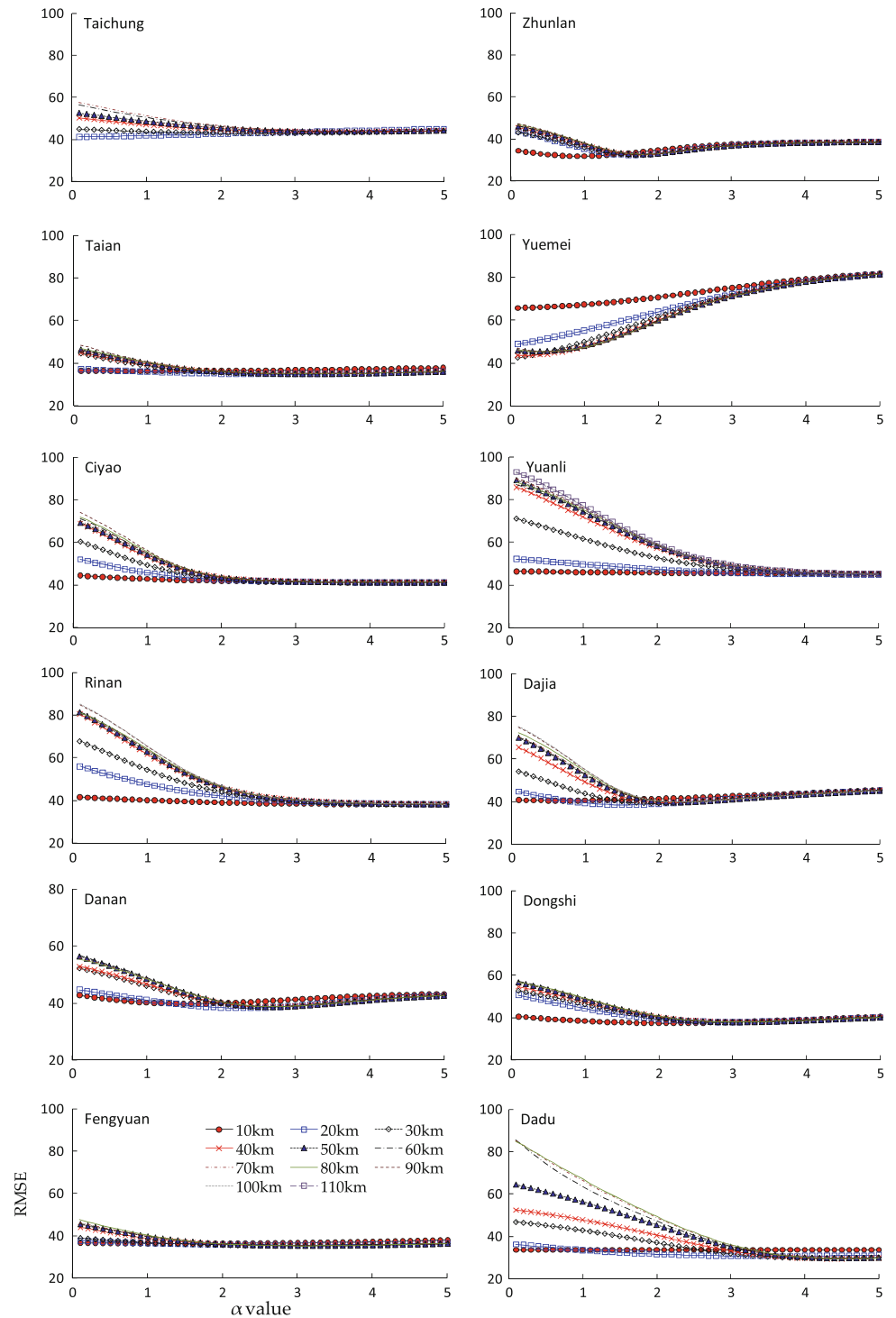
Rainfall stations	Item	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Annual
Taichung	RMSE	5.7	22.0	20.8	35.5	61.9	79.2	104.5	73.5	59.5	10.8	3.3	9.8	41.2
	α	2.7	0.7	0.1	0.1	0.1	2.4	0.1	0.1	1.6	0.1	1.9	2.3	0.1
	O.S.R.	30	30	20	20	20	70	20	20	30	20	20	50	20
Zhunlan	RMSE	7.3	14.6	12.9	30.9	56.3	54.4	61.4	73.2	24.7	13.1	5.1	8.0	31.8
	α	1.8	1.0	1.6	1.2	0.2	1.0	1.1	1.2	2.1	2.0	0.5	1.5	0.9
	O.S.R.	20	10	50	30	10	10	10	40	70	20	10	10	10
Taian	RMSE	6.6	28.5	20.1	25.3	45.9	82.4	57.8	73.5	36.7	9.7	7.7	7.4	34.8
	α	3.2	1.7	1.8	0.1	1.8	2.7	2.0	1.5	1.2	5.0	4.0	2.2	2.8
	O.S.R.	10	20	20	10	20	20	30	50	30	10	10	10	20
Yuemei	RMSE	7.8	17.5	15.7	29.2	58.0	99.1	94.5	79.2	37.8	20.5	5.0	7.0	42.4
	α	1.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.1	1.6	0.1
	O.S.R.	40	30	40	30	30	40	30	90	30	40	20	90	30
Ciyao	RMSE	16.3	22.0	30.1	21.4	67.1	119.9	52.8	88.9	39.4	9.8	6.3	8.4	41.2
	α	0.1	5.0	0.7	0.1	4.5	1.8	5.0	3.7	5.0	5.0	0.2	0.1	4.7
	O.S.R.	20	10	20	10	10	10	10	10	90	20	10	10	10
Yuanli	RMSE	10.2	26.8	31.6	73.6	56.9	48.3	66.5	98.9	40.9	18.0	20.4	11.3	44.9
	α	5.0	5.0	1.4	5.0	5.0	5.0	0.1	5.0	1.9	0.1	0.1	5.0	5.0
	O.S.R.	10	20	20	10	20	10	10	10	20	10	30	10	20
Rinan	RMSE	13.2	21.1	23.4	25.2	41.3	52.9	42.6	48.7	141.1	5.3	7.4	6.2	38.1
	α	0.1	0.6	1.3	0.1	0.1	5.0	5.0	5.0	5.0	5.0	0.5	2.9	5.0
	O.S.R.	10	20	10	10	10	10	10	10	10	60	10	10	10
Dajia	RMSE	7.1	49.7	19.1	43.9	45.0	71.2	61.8	88.2	32.1	8.2	9.5	8.0	38.5
	α	0.7	0.1	0.1	1.3	0.4	1.3	2.0	2.5	1.7	1.7	0.3	1.6	1.5
	O.S.R.	10	20	10	30	20	20	20	20	30	30	10	60	20
Danan	RMSE	8.0	15.4	29.1	25.9	36.8	77.3	76.1	85.6	47.4	12.4	11.0	14.6	38.2
	α	3.1	0.1	0.1	2.0	1.8	3.7	3.2	0.9	1.9	0.8	3.2	0.1	2.3
	O.S.R.	20	10	20	40	20	30	20	10	20	10	20	20	20
Dongshi	RMSE	6.2	17.4	25.4	49.3	35.7	53.0	49.9	110.9	43.3	15.9	6.9	4.9	37.5
	α	2.1	1.1	2.0	5.0	2.3	1.9	0.1	5.0	2.1	5.0	0.5	1.0	2.1
	O.S.R.	70	20	20	10	30	10	10	10	10	60	10	70	10
Fengyuan	RMSE	6.4	14.4	20.3	21.8	40.8	88.0	77.4	80.5	27.2	6.8	5.2	6.9	35.3
	α	1.0	0.1	3.2	3.7	0.9	3.8	3.1	1.2	3.0	1.1	2.3	0.6	3.1
	O.S.R.	20	10	30	20	80	10	10	30	10	40	20	20	80
Dadu	RMSE	16.3	10.5	15.0	23.4	41.8	40.8	49.5	65.9	48.1	8.0	4.4	7.3	29.5
	α	0.3	4.9	0.6	2.0	0.1	3.6	0.1–5.0	5.0	3.9	0.1–5.0	1.2	1.8	4.4
	O.S.R.	20	40	20	20	20	30	10	30	40	10	20	30	40

O.S.R. optimal search radius

use of numerous rainfall stations unnecessary for data interpolation under most conditions. Nevertheless, the greater result accuracy occurred when only four to five rainfall stations were considered such as Zhunlan, Ciyao, Rinan, and Dongshi rainfall stations as opposed to incorporating data from all 45 rainfall stations. A similar trend also occurred with monthly rainfall levels. The results revealed that the interpolation accuracy of rainfall was greater with increasing rainfall stations and to an optimal up-limit. However, the interpolation accuracy can become inferior

when the number of rainfall stations considered exceeds the optimal value. In all cases of this article, when considering all 45 stations, less than 20 stations, and less than 10 stations, the optimal number of rainfall stations 8, 83 and 58%, respectively. Excessive number of rainfall stations considered for interpolate rainfall could cause the data to become meaningless. Such results are similar to previous researches such as those conducted by Li et al. (2006), Lin and Yu (2008), and Chu et al. (2008) which the optimal number of rainfall stations were 10, 13 and 15 respectively.

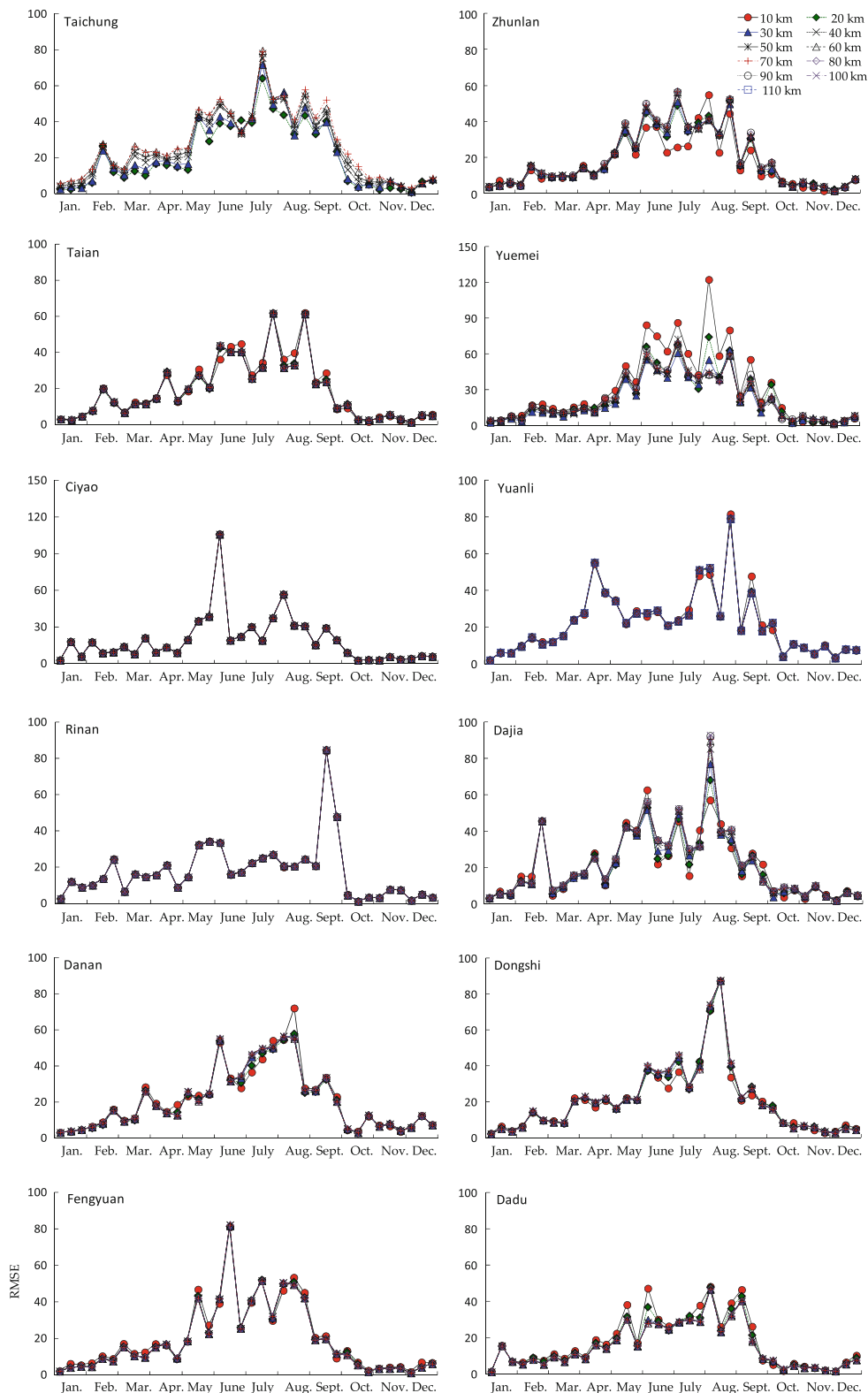
Fig. 4 RMSE variation of different search radii (10–110 km) and α value (0–5) of 12 rainfall stations



The second phenomenon was that the optimal α value varied greatly from zero to five. There was only a 1.92% probability met and that the optimal α was equal to 2.0. It revealed that the prediction accuracy of rainfall could not be the optimal when using α value at 2.0. Even at a 2.0 α value, it is only a general consent but cannot be considered scientific. The result in this article have identical views

with several researches including one by Chu et al. (2008) reported the optimal α value was equal to 3.31 when 15 rainfall stations were considered to measure a case of Gansu, China. Wang et al. (2008) also reported a case in China where the prediction accuracy of annual rainfall had the highest significance when the α value was considered in the range of three to five. However, Dirks et al. (1998)

Fig. 5 RMSE variation of different search radii (10–110 km) using annual optimal α value in each month of 12 rainfall stations



showed that the exact choice of the numerical value of the power has minimal effect on the resulting errors providing data using the α values within the range of 1.5–4. It was an inconsistent argument with this article, because several

optimal α values of this article occurred in the range 0–1.5 and even at 4–5. Table 2 also showed that the annual RMSE existed in the range of 29.5–44.9. Figure 4 displays a series of 12 sub-diagrams on the RMSE variation at

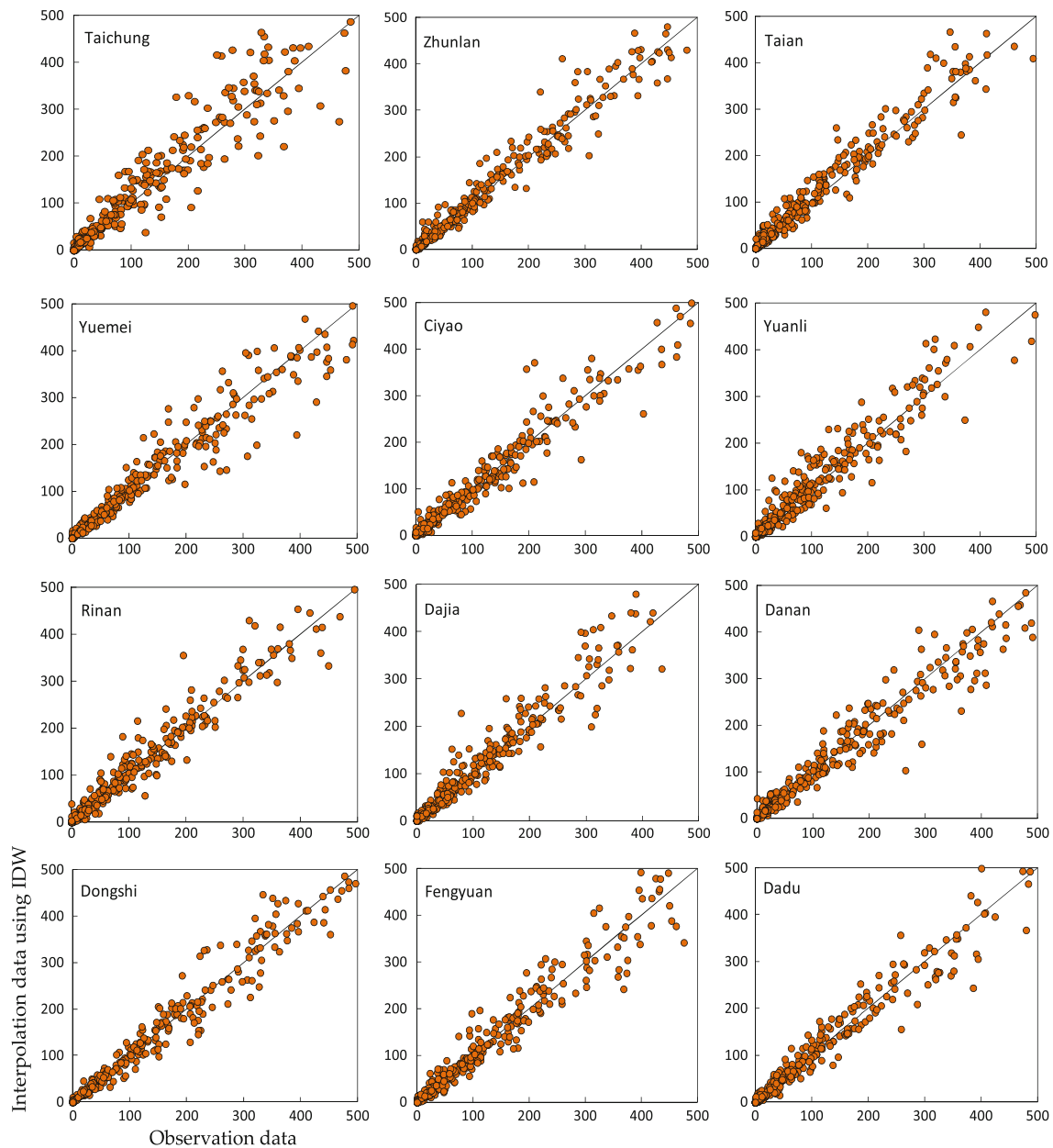


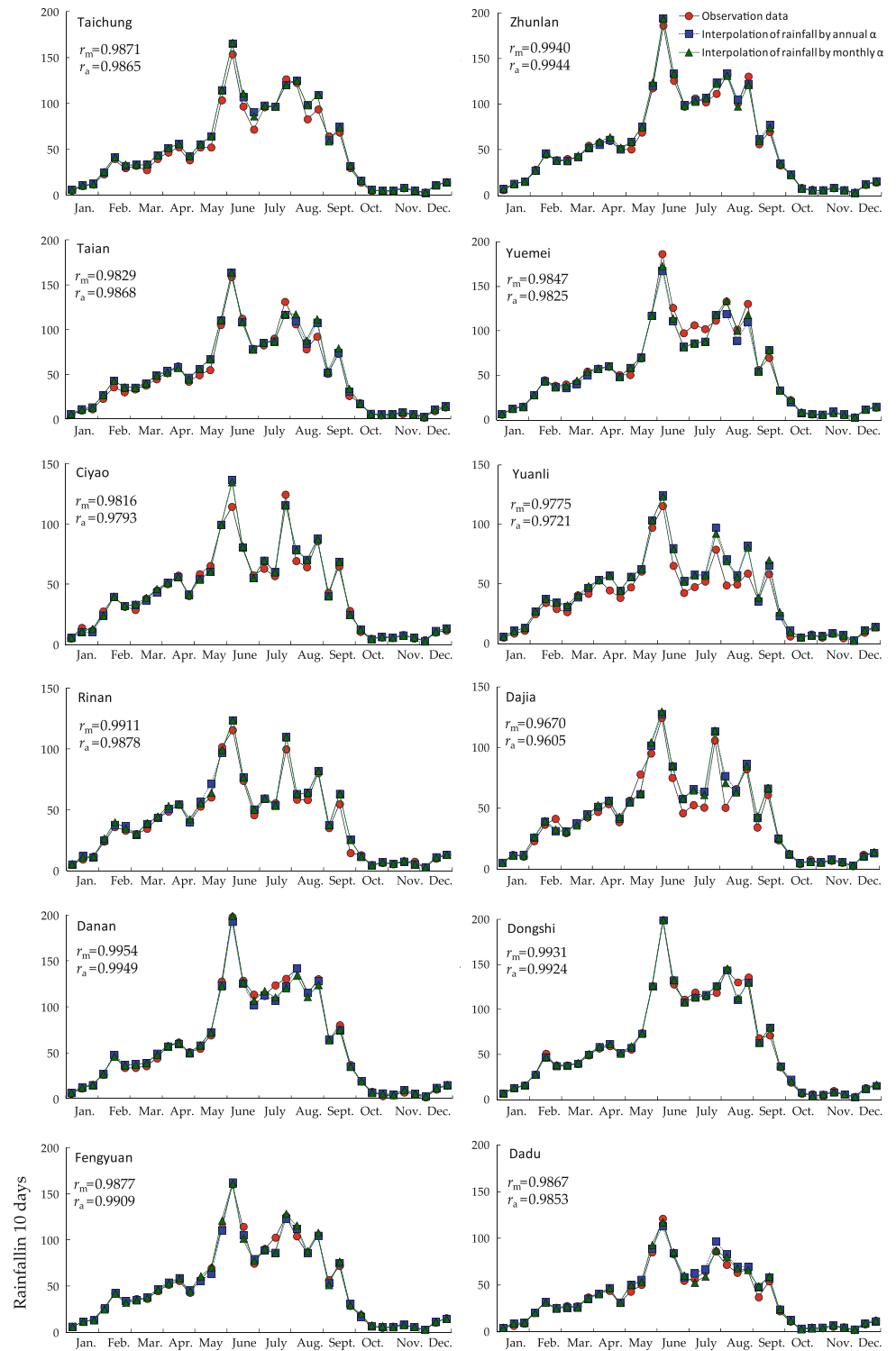
Fig. 6 Scatter plots of monthly interpolation of spatial rainfall data

different search radii (10–110 km) and α value (from zero to five with an increment interval of 0.1). The finding was a large RMSE variation of different groups (search radius) occurred when α approach zero, the RMSE variation also reduced with the increase in α . This showed that the minimum variation occurred at the largest α value, regardless of the number of rainfall stations used for interpolation. However, the minimum variation of RMSE remains uncertain on based of the optimal α and search radius. Therefore the optimal α and search radius must be further measured.

Figure 5 diagrams the RMSE variation of different groups using annual optimal α value in each month. These

results show that all 12 rainfall stations follows the same trend: the RMSE value was relatively lower in dry seasons (from October to April) than in flood seasons (from May to September), it denoted that the data obtained during dry seasons are more accurate than in flood seasons. This point of view in this study is consistent with the study done by Kong and Tong (2008). It would therefore consider that spatial rainfall interpolated in flood season is inferior to the data interpolated during dry seasons due to extreme rainfall events. However, another research had indicated that IDW was better than kriging and suggested for spatial rainfall prediction in summer (Hsieh et al. 2006). The relationship

Fig. 7 Comparison of observation and interpolation of spatial rainfall data using optimal monthly and annual α values



between the interpolated rainfall values and the true observed data was also evaluated. Figure 6 showed a significant accuracy of the predicted data during low rainfall value, and the deviation increased gradually with increased rainfall. The trend could explain the cause of high RMSE values occurring during flood seasons (Fig. 5). The

phenomenon implied large deviations were caused from extreme rainfall events in the flood season.

Finally, for the purpose of evaluating the suitability of using IDW for interpolated data, daily rainfall data interpolated were accumulated into 36 ten-day rainfall values from 1 year, and was compared with the observed data.

The coefficient of correlation (r) was utilized as an indicator to evaluate the fittingness of IDW. Cross-validation methods were applied to 12 optimal monthly α values and a single annual one on an individual basis. The r can be expressed as r_m and r_a . 30-years worth of average data (1981–2010) presented in the form of 10-day interpolated rainfall were compared with observed data in Fig. 7. The result showed that r_m and r_a were all higher than 0.95 in 12 rainfall stations. It was evident that, rainfall interpolations using IDW showed significant similarities with the observed data, either using optimal monthly α values or as independent annual values. Therefore an argument can be drawn from this study: that IDW is a suitable method for rainfall interpolation under the conditions that optimal α and search radius must be measured.

Conclusion

In this article, the authors have three findings about using IDW for interpolate spatial rainfall: (i) The predicted accuracy of rainfall interpolated can be improved through the α value adjustment, and that the α value usually is not equivalent to two. Therefore, for the purpose of increasing prediction accuracy, searching an optimal α value as a preparation step is necessary. (ii) The number of known rainfall station is also another influential parameter; most cases show that the prediction accuracy increases with the increasing numbers of known rainfall station. However, the accuracy of rainfall data interpolation could be reduced by the interference from the use of excessive rainfall stations. Nevertheless, radius of influence is important to effective interpolation of rainfall data. The optimal result is based on only using rainfall stations within the radius of influence. (iii) Application of IDW for spatial rainfall data interpolation, results show the prediction accuracy are better during dry seasons (October to April) than in flood seasons (May–September). It reveals that IDW has significant prediction ability in small rainfall events than in extreme rainfall events. In summary, through analyzing the optimization steps of α value and radius of influence, IDW is deemed as a suitable spatial interpolation method of rainfall.

References

- Bedient PB, Huber WC (1992) Hydrology and floodplain analysis, 2nd edn. Addison-Wesley, Reading
- Burrough PA, McDonnell RA (1998) Principles of geographical information systems. Oxford University Press, Oxford
- Chu SL, Zhou ZY, Yuan L, Chen QG (2008) Study on spatial precipitation interpolation methods. *Pratacult Sci* 25(6):19–23
- Cressie N (1993) *Statics for spatial data* (revised edition). Wiley, New York
- Devijver PA, Kittler J (1982) *Pattern recognition: a statistical approach*. Prentice-hall, London
- Dirks KN, Hay JE, Stow CD (1998) High resolution studies of rainfall on Norfolk Island Part II: interpolation of rainfall data. *J Hydrol* 208:187–193
- Dong XH, Bo HJ, Deng X, Su J, Wang X (2009) Rainfall spatial interpolation methods and their applications to Qingjiang river basin. *J China Three Gorges Univ (Nat Sci)* 31(6):6–10
- Feng ZM, Yang YZ, Ding XQ, Lin ZH (2004) Optimization of the spatial interpolation methods for climate resources. *Geograph Res* 23(5):357–364
- Fotheringham AS, Brunsdon C, Charlton M (2002) *Geographically weighted regression: the analysis of spatially varying relationship*. Wiley, New York
- Garcia M, Peters-Lidard CD, Goodrich DC (2008) Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. *Water Resour Res* 44:W05S13(1–14)
- Geisser S (1993) *Predictive inference*. Chapman and Hall, New York
- Gemmer M, Becker S, Jiang T (2004) Observed monthly precipitation trends in China 1951–2002. *Theor Appl Climatol* 77:39–45
- Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J Hydrol* 228:113–129
- Gyalistras D (2003) Development and validation of a high-resolution monthly gridded temperature and precipitation data set for Switzerland (1951–2000). *Clim Res* 25(1):55–83
- Hsieh HH, Cheng SJ, Liou JY, Chou SC, Siao BR (2006) Characterization of spatially distributed summer daily rainfall. *J Chin Agric Eng* 52(1):47–55
- Jeffrey SJ, Carter JO, Moodie KB, Beswick AR (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ Model Softw* 16(4):309–330
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the fourteenth international joint conference on artificial intelligence*, vol 2, no 12. Morgan Kaufmann, San Mateo, pp 1137–1143
- Kohavi R, Provost F (1998) Special issue on applications of machine learning and the knowledge discovery process. *Mach Learn* 30(2–3):271–274
- Kong YF, Tong WW (2008) Spatial exploration and interpolation of the surface precipitation data. *Geograph Res* 27(5):1097–1108
- Kurtzman D, Navon S, Morin E (2009) Improving interpolation of daily precipitation for hydrologic modeling: spatial patterns of preferred interpolators. *Hydrol Process* 23:3281–3291
- Lam NS-N (1983) Spatial interpolation methods: a review. *Am Cartogr* 10:129–140
- Li J, Heap AD (2008) *Spatial interpolation methods: a review for environmental scientists*. Geoscience Australia, Record. Geoscience Australia, Canberra
- Li JL, Zhang J, Zhang C, Chen QG (2006) Analyze and compare the spatial interpolation methods for climate factor. *Pratacult Sci* 23(8):6–11
- Li B, Huang JF, Jin ZF, Liu ZY (2010) Methods for calculation precipitation spatial distribution of Zhejiang Province based on GIS. *J Zhejiang Univ (Sci Ed)* 27(2):239–244
- Lin XS, Yu Q (2008) Study on the spatial interpolation of agroclimatic resources in Chongqing. *J Anhui Agric* 36(30):13431–13463, 13470
- Lloyd CD (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J Hydrol* 308:128–150
- Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci* 34(9):1044–1055

- McLachlan GJ, Ambrose K-A, Do C (2004) Analyzing microarray gene expression data. Wiley, New York
- Naoum S, Tsanis IK (2004) A multiple linear regression GIS module using spatial variables to model orographic rainfall. *J Hydroinform* 6:39–56
- Phogat V, Yadav AK, Malik RS, Kumar S, Cox J (2010) Simulation of salt and water movement and estimation of water productivity of rice crop irrigated with saline water. *Paddy Water Environ* 8:333–346
- Price DT, McKenney DW, Nalder IA, Hutchison MF, Kesteven JL (2000) A comparison of two statistical methods for interpolation of Canadian monthly mean climate data. *Agric Meteorol* 101:81–94
- Seaman RS (1983) Objective analysis accuracies of statistical interpolation and successive correction schemes. *J Aust Meteorol Mag* 31:225–240
- Segond M-L, Neokleous N, Makropoulos C, Onof C, Maksimovic C (2007) Simulation and spatio-temporal disaggregation of multi-site rainfall data for urban drainage applications. *Hydrol Sci J* 52(5):917–935
- Simanton JR, Osborn HB (1980) Reciprocal-distance estimate of point rainfall. *J Hydraul Eng* 106(7):1242–1246
- Traore S, Wang YM, Kan CE, Kerh T, Leu JM (2010) A mixture neural methodology for computing rice consumptive water requirements in Fada N’Gourma Region, Eastern Burkina Faso. *Paddy Water Environ* 8:165–173
- Tung YK (1983) Point rainfall estimation for a mountainous region. *J Hydraul Eng* 109(10):1386–1393
- Wang Y, Li CK, Chen L, Zheng SN (2008) Analysis on impact of weight to spatial interpolation methods. *J Hunan Univ Sci Technol (Nat Sci Ed)* 23(4):77–80
- Wu L, Wu XJ, Xiao CC, Tian Y (2010) On temporal and spatial error distribution of five precipitation interpolation models. *Geogr Geo-Inf Sci* 26(3):19–24
- Yeh HC, Chen YC, Wei C, Chen RH (2011) Entropy and kriging approach to rainfall network design. *Paddy Water Environ* 9:343–355
- Zhong JJ (2010) A comparative study of spatial interpolation precision of annual average precipitation based on GIS in Xinjiang. *Desert Oasis Meteorol* 4(4):51–54
- Zhu HY, Jia SF (2004) Uncertainty in the spatial interpolation of rainfall data. *Prog Geogr* 23(2):34–42
- Zhuang LW, Wang SL (2003) Spatial interpolation methods of daily weather data in Northeast China. *Quart J Appl Meteorol* 14(5):605–616