



# Applying fuzzy theory and genetic algorithm to interpolate precipitation

C.L. Chang<sup>a,\*</sup>, S.L. Lo<sup>a</sup>, S.L. Yu<sup>b</sup>

<sup>a</sup>Research Center for Environmental Pollution Prevention and Control Technology, Graduate Institute of Environmental Engineering, National Taiwan University, No. 71, Chou-Shan Road, Taipei 106 Taiwan, ROC

<sup>b</sup>Department of Civil Engineering, University of Virginia, Charlottesville, VA, USA

Received 22 December 2003; revised 14 February 2005; accepted 18 March 2005

## Abstract

A watershed management program is usually based on the results of watershed modeling. Accurate modeling results are decided by the appropriate parameters and input data. Rainfall is the most important input for watershed modeling. Precipitation characteristics, such as rainfall intensity and duration, usually exhibit significant spatial variation, even within small watersheds. Therefore, properly describing the spatial variation of rainfall is essential for predicting the water movement in a watershed. Varied circumstances require a variety of suitable methods for interpolating and estimating precipitation. In this study, a modified method, combining the inverse distance method and fuzzy theory, was applied to precipitation interpolation. Meanwhile, genetic algorithm (GA) was used to determine the parameters of fuzzy membership functions, which represent the relationship between the location without rainfall records and its surrounding rainfall gauges. The objective in the optimization process is to minimize the estimated error of precipitation. The results show that the estimated error is usually reduced by this method. Particularly, when there are large and irregular elevation differences between the interpolated area and its vicinal rainfall gauging stations, it is important to consider the effect of elevation differences, in addition to the effect of horizontal distances. Reliable modeling results can substantially lower the cost for the watershed management strategy.

© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Fuzzy theory; Genetic algorithm; Inverse distance method; Precipitation interpolation

## 1. Introduction

Effective watershed management strategies depend upon accurate model results. Although the cost and effort of a hydrology and water quality modeling study is typically a small fraction of the total

management program cost, the cost of implementing an inefficient strategy based on faulty modeling may be much larger. Because so much is at stake, reliable model results are very important (Lung, 2001).

All the watershed responses, such as runoff, soil erosion, and the variation of water quality in a reservoir, result from rainfall. In other words, rainfall is the most important input for watershed modeling, including hydrology and water quality modeling. Rainfall characteristics are usually spatially varying,

\* Corresponding author. Tel.: +886 2 23625373; fax: +886 2 23628764.

E-mail address: [f89541201@ntu.edu.tw](mailto:f89541201@ntu.edu.tw) (C.L. Chang).

even in a small watershed; so accurately describing the spatial variation of rainfall is quite important for predicting the water movement in a watershed. The lack of spatial information on rainfall will be a serious obstacle to the development of modeling technology (Vicente, 1996). Using rainfall data from a single gauge to represent the rainfall characteristics of an entire watershed will cause much uncertainty when modeling non-point source pollution (Chaubey et al., 1999). Even small-scale spatial variations in rainfall can greatly influence runoff (Faures et al., 1995).

Some model frameworks, through the design of grid inputs and the sub-basin divisions, allow the users to input information separately for each partition. It is an efficient way to describe the spatial variation in a watershed. However, gauging stations, such as rainfall gauges and flow gauges, are impossible to be set up everywhere in a watershed. Therefore, it is essential to use limited monitored data to interpolate unknown information.

The results of spatially distributed models can be substantially influenced by interpolation errors (Donald and Danny, 1996). Although the interpolation methods are manifold, there is no single method suitable to be applied in every circumstance (Nalder and Wein, 1998). The Kriging method, the optimal interpolation method, and the weighted method are commonly used to estimate precipitation. The accuracy of estimation, of course, is related to the density of gauging stations. Among the interpolation methods mentioned above, the Kriging method, based on statistic theory (Dirks et al., 1998), and the optimal interpolation method (Tabios and Salas, 1985) are necessary for more complicated data. Consequently, if the gauging station data is limited, the estimated error will be great. As a remedy, the weighted method is more flexible for adjusting the weighting factors to account for the relative influence exerted by each gauge. Thus, it can greatly reduce the estimated error due to the limited number of gauges.

The inverse distance method is one of the weighted methods, and the horizontal distances between the estimated area and its surrounding gauges determine the weighting factors used in this method. Some studies have shown that it is necessary to consider the effect of multiple factors, in addition to horizontal distances, on the relative importance of each gauging station (Bartier and Keller, 1996). Moreover, the

membership function in the fuzzy set theory (Zadeh, 1965) is a new concept rather than a traditional mathematical way to describe a problem. In this study, the modified inverse distance method combined with the fuzzy theory for estimating precipitation was discussed, and the membership function was applied to represent the relationship between the area without rainfall records and the considered rainfall gauges in its neighborhood. Meanwhile, the genetic algorithm, an efficient tool for optimization analysis, was applied to determine the parameters of membership functions.

## 2. Methods

### 2.1. The inverse distance method

The inverse distance method, which is also called the inverse distance weighted (IDW) interpolation, is a general technique for interpolating. This method has been used widely in many different fields, such as hydrology, earth science (Ware et al., 1991; Ashraf et al., 1997; Cheng, 1998), etc. The basic equation for the inverse distance method is

$$k_{xy} = \frac{\sum_{i=1}^N k_i w_i}{\sum_{i=1}^N w_i} \quad (1)$$

where  $k_i$  is the control value for  $i^{\text{th}}$  sample point,  $w_i$  represents a weight determining the relative importance of individual control point  $k_i$  in the interpolation process,  $k_{xy}$  is the point to be estimated, and  $N$  is the number of sample points (Bartier and Keller, 1996). This concept is also commonly applied to estimate average precipitation and interpolate unknown rainfall. In the case, when each control point has the same relative importance, the inverse distance method is identical to the arithmetic average method for estimating precipitation. In another case, sometimes the relative importance can be expressed as a binary switch. Using this approach,  $w_i$  is equal to 1 for the several control points nearest to the point to be interpolated, or for the set of control points within some radius of the point being interpolated, and  $w_i$  is given by 0 otherwise (Bartier and Keller, 1996). This notion is similar to the Thiessen Polygons method. However, the arithmetic average method sometimes

cannot represent the influence of each control point; also, the Thiessen Polygons method usually over-expresses the influence.

An alternative weighting strategy giving near points more influence than distant points is based on a formula using the inverse of distance to a power, such as

$$w_i = d_{xyi}^{-m} \quad (2)$$

where  $d_{xyi}$  is the distance between  $k_{xy}$  and  $k_i$ , and  $m$  is an exponent given by the users (Bartier and Keller, 1996), and also named the order of distances. The inverse distance method is flexible due to the adjustable nature of the order of distances (Chang et al., 2003). Then, Eq. (1) can be rewritten as

$$k_{xy} = \frac{\sum_{i=1}^N k_i d_{xyi}^{-m}}{\sum_{i=1}^N d_{xyi}^{-m}} \quad (3)$$

Also, the weighting factor,  $W_i$ , which represents the relative influence, can be defined as Eq. (4). The sum of the weighting factors of each rainfall gauging station in the neighborhood is equal to one.

$$W_i = \frac{w_i}{\sum_{i=1}^N w_i} = \frac{d_{xyi}^{-m}}{\sum_{i=1}^N d_{xyi}^{-m}} \quad (4)$$

Although the theory is not complicated when used to decide the weighting factors and interpolate precipitation, it is still a heavy job to modify the order of distances and lower the estimated error. After determining the weighting factors, the average precipitation can be estimated. The basic calculation of the IDW interpolation for estimating precipitation is expressed as

$$P_p = \sum_{i=1}^N (W_i P_i) = \frac{\sum_{i=1}^N P_i d_{pi}^{-m}}{\sum_{i=1}^N d_{pi}^{-m}} \quad (5)$$

where  $P_p$  is the interpolated precipitation in the area  $p$ ;  $P_i$  is the precipitation of rainfall gauge  $i$ ;  $W_i$  is the weighting factor that represents the relative influence of gauging station  $i$ , and  $d_{pi}$  is the distance between the area  $p$  and the rainfall gauge  $i$ .

The IDW interpolation is univariate with a single influence factor, namely horizontal distance. This technique assumes that the interpolation area is uniform rather than variable (Ware et al., 1991). Therefore, it cannot be applied in an area with abrupt

changes in elevation, which would create a major obstacle to estimating unknown information. Subsequently, precipitation multivariate IDW interpolation, a modified version for considering additional independent variables, was developed to improve upon the previous method. The modified equation can be given by

$$k_{xy} = \frac{\sum_{i=1}^N k_i w_i(v_1, \dots, v_x)}{\sum_{i=1}^N w_i(v_1, \dots, v_x)} \quad (6)$$

where the weights  $w_i$  are determined by the variables  $v_1, \dots, v_x$ . A multivariate version based on Eq. (3) can be redefined as

$$k_{xy} = \frac{\sum_{i=1}^N k_i d_{xyi}^{-m} w_i(v_1, \dots, v_x)}{\sum_{i=1}^N d_{xyi}^{-m} w_i(v_1, \dots, v_x)} \quad (7)$$

In this equation, it is assumed that there are two independent weights controlling the interpolation process, namely the inverse distance weight, and a second weight that represents the influence of all other factors. Of course, these two weights can be combined into a single weight (Bartier and Keller, 1996). The weighting effect of two independent factors—horizontal distances and elevation differences between the interpolated area and its surrounding rainfall gauging stations is discussed herein.

## 2.2. Fuzzy theory

Fuzzy theory, first introduced by Zadeh (1965), has been applied in various engineering applications to deal with imprecise information (Klir and Yuan, 1995). Recently, fuzzy multi-objective programming, fuzzy uncertainty analysis, fuzzy decision system, fuzzy optimal model, etc. have been commonly used in the fields of environmental science, hydrology and water resources (Bardossy et al., 1990; Lagacherie et al., 1997; Wu et al., 1997; Perret and Prasher, 1998; Cheng, 1999; Yu and Yang, 2000; Zhu and Mackay, 2001; Cheng et al., 2002). Traditional binary value, such as true and false, are not as useful as a variable with many values between these two extremes (Klir and Folger, 1988). Fuzzy logic addresses these situations by allowing variables to be ‘partially true’ and/or ‘partially false’ (Abdel-Kader et al., 1998). Meanwhile, it provides a more appropriate way to

describe the real world as well as an adaptable process to solve problems.

The weighting factors in the IDW interpolation method and the membership function in the fuzzy theory express the same idea in different words. The relative influence of each rainfall gauge is an abstract factor, so it is not easy to use absolutely binary classification, i.e. completely related and completely unrelated classes. Also, the effect factor is certainly not unique. Therefore, combining various effect factors would be difficult, since the scale and the influence characteristic of different effect factors are by no means analogous.

The fuzzy membership function used in this paper indicates the relative importance of each rainfall gauge. Mathematically, let  $U = \{d_1, \dots, d_i, \dots, d_N\}$  be a universal set of objects  $d$ , namely the variable of horizontal distances, and  $V = \{h_1, \dots, h_i, \dots, h_N\}$  be a general set of objects  $h$ , i.e. the variable of elevation differences. Then, fuzzy sets  $P_d$  in  $U$  and  $P_h$  in  $V$  are defined as:

$$P_d = \{d, \mu_{P_d}(d)\} \quad \forall d \in U \quad (8)$$

$$P_h = \{h, \mu_{P_h}(h)\} \quad \forall h \in V \quad (9)$$

where  $\mu_{P_d}$  is called the membership grade of  $d$  in  $P_d$ , i.e. the membership function of horizontal distances, which represents the relative importance of each surrounding rainfall gauge due to the effect of horizontal distances.  $\mu_{P_h}$  is called the membership grade of  $h$  in  $P_h$ , i.e. the membership function of elevation differences, which shows the relative importance of each vicinal gauging station due to the effect of elevation differences.

The value of membership function  $\mu$  varies from 0 to 1, and represents the degree of importance and influence from non-membership to full-membership.

This study considered two affect factors including the horizontal distances and the elevation differences between the interpolated area and its adjacent rainfall gauging stations. Their membership functions are separately defined as Eqs. (10) and (11), and shown as Fig. 1.

$$\mu_{P_d}(d) = \begin{cases} 1 & 0 \leq d \leq d_{\text{close}} \\ d^{-m} & d > d_{\text{close}} \end{cases} \quad (10)$$

$$\mu_{P_h}(h) = \begin{cases} 1 & 0 \leq h \leq h_{\text{close}} \\ h^{-n} & h > h_{\text{close}} \end{cases} \quad (11)$$

where  $m$  and  $n$  are the order of horizontal distances and the order of elevation differences separately. Also,  $d_{\text{close}}$  and  $h_{\text{close}}$  define a control area. The rainfall gauges in the control area is much more similar to the interpolated area, so it is supposed to be with full effect on the interpolated area. In this study,  $d_{\text{close}}$  and  $h_{\text{close}}$  were assumed to be zero. This assumption means that it exists precipitation difference between any two locations.

Significantly, the membership function will change following the variation of the power of the inverse distances. As shown in Fig. 1, when the order of horizontal distances, the value of  $m$ , increases, the profile of the membership function will change from curve 1 to curve 4. Similarly, when the order of elevation differences, the value of  $n$ , increases,

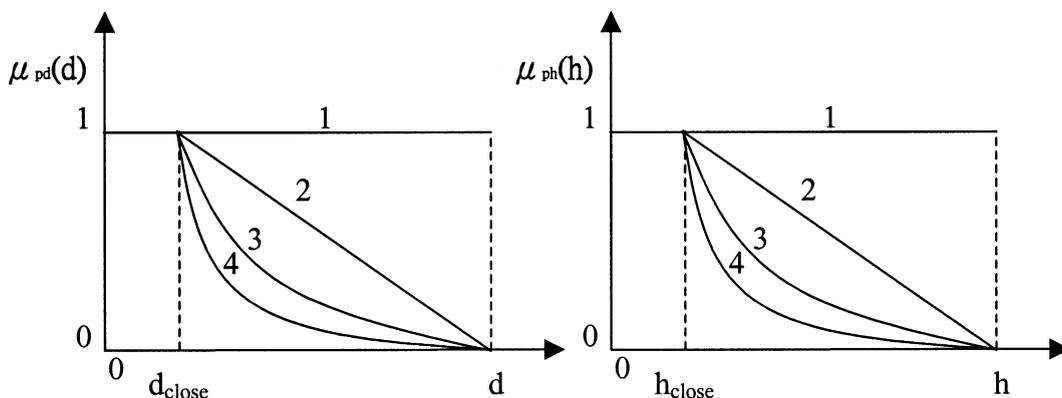


Fig. 1. The profile of the membership function of horizontal distances and elevation differences.

the profile of the membership function will also shift from curve 1 to curve 4.

The relative importance of each rainfall station is determined by these two membership functions. Although the tendencies of these two membership functions are resemble, the scales of horizontal distances and elevation differences are different. In this study, normalization technology (Craig and Karen, 1995) was applied for avoiding the misleading analysis due to various scales of horizontal distances and elevation differences. The elements in set  $U$ , and  $V$  are redefined as normalized horizontal distances and elevation differences. Meanwhile, how to combine these membership functions is significantly important, so define the composite fuzzy set as  $W = \{x_1, \dots, x_i, \dots, x_N\}$ , which means a set of rainfall gauges  $x$ , then a fuzzy set  $P$  in  $W$  can be defined as

$$P = \{x, \mu_P(x)\} \quad \forall x \in W \tag{12}$$

where  $\mu_P$  is the membership grade of  $x$  in  $P$ . It is combined the membership degree of horizontal distances with the membership degree of elevation differences, so  $\mu_P(x)$  can be redefined as

$$\mu_P(x) = \mu_P(d, h) \tag{13}$$

The effect trends of horizontal distances and elevation differences are the same, so the integrated operation is supposed to be additive or multiple. In this study, several operations were used to combine these membership functions for representing the integrate effect of horizontal distances and elevation differences. The equations are as follows:

$$\mu_P(d, h) = \sqrt{\mu_{P_d}(d)^2 + \mu_{P_h}(h)^2} \tag{14}$$

$$\mu_P(d, h) = \mu_{P_d}(d) + \mu_{P_h}(h) \tag{15}$$

$$\mu_P(d, h) = \mu_{P_d}(d)\mu_{P_h}(h) \tag{16}$$

$$\mu_P(d, h) = \max[\mu_{P_d}(d), \mu_{P_h}(h)] \tag{17}$$

$$\mu_P(d, h) = \min[\mu_{P_d}(d), \mu_{P_h}(h)] \tag{18}$$

Moreover, the weighting factors can be determined by the membership grades of each rainfall station. It is

calculated by

$$W_i = \frac{\mu_P(d_i, h_i)}{\sum_{i=1}^N \mu_P(d_i, h_i)} \tag{19}$$

Then, the estimated precipitation can be estimated by

$$P_p = \sum_{i=1}^N (W_i P_i) = \frac{\sum_{i=1}^N P_i \mu_P(d_i, h_i)}{\sum_{i=1}^N \mu_P(d_i, h_i)} \tag{20}$$

### 2.3. Genetic algorithm

Genetic algorithm (GA) is inspired by Darwin’s theory about evolution, which strengthens survival ability by the processes of reproduction, crossover and mutation of genes, etc. The algorithm starts with a set of solutions called population, which is analogous to the chromosomes in the natural systems. Gene encoding, the first step to apply genetic algorithm, is a way to create decision variables analogous to the genes in chromosomes. The procedure of GA simulates the processes of reproduction, crossover and mutation to maintain superior solutions and to generate better and better offspring, to make the solutions close to the objective function (Tung et al., 2003). This process is significant to select the individuals in the population and generate new individuals. The objective function discriminates how good each individual

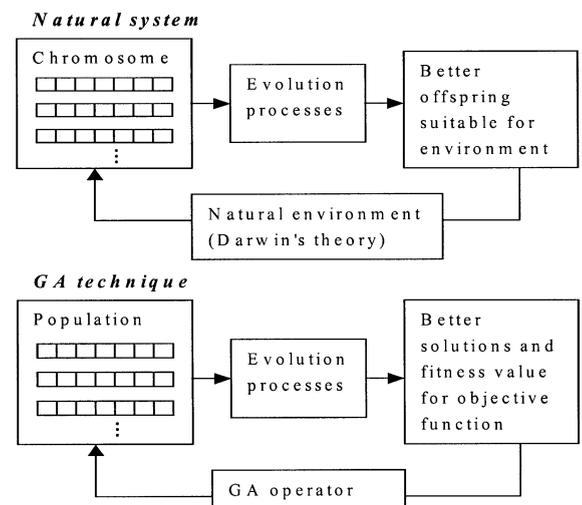


Fig. 2. Comparison of the procedure of Genetic algorithm and Natural system.

is. Fig. 2 shows the procedure of GA. GA has been verified to have more advantages than the classical optimization methods (Cheng et al., 2002). In recent years, it has been a popular technique for solving hydrology and water resources problems (Wang, 1991; Ritzel and Wayland, 1994; Savic et al., 1999; Cheng et al., 2002; Tung et al., 2003).

This study applied GA to adjust the powers of inverse distances including the order of horizontal distances  $m$ , and the order of elevation differences  $n$ , for integrating approximate relative importance grades of each rainfall gauging station, and improving the efficiency on precipitation interpolation. Due to the evolution of genes, the worse individual will be eliminated, and a better individual fit the objective function will be selected automatically and progressively. Hence, using GA can significantly reduce the difficulty of adjusting the powers.

When applying the GA technology, of course, the objective function should be defined first. In this study, the minimum estimated error was the main objective. Some criteria, such as mean error (ME) representing the degree of deviation, mean absolute error (MAE) illustrating the possible maximum deviation, and root mean square error (RMSE) showing the sensitivity of deviation, are popular used in various studies (Hulme et al., 1995; Ashraf et al., 1997; Price et al., 2000) to determine the accuracy of estimated results, and each criterion has different meaning. In order to avoid the variable scale, which would influence the criterion, relative error (RE) is usually used as a criterion to show the dimensionless relative deviation (Vicente, 1996). In this study, daily estimates were used for calculating estimated error of precipitation. The major criteria applied herein are defined as follows:

$$\text{Sum(MAE)} = \sum_{i=1}^{365} |(O_i - P_i)| \quad (21)$$

$$\text{RE} = \frac{\text{Sum(MAE)}}{\sum_{i=1}^{365} P_i} \quad (22)$$

where  $O_i$  is the  $i^{\text{th}}$  observed precipitation, and  $P_i$  is the  $i^{\text{th}}$  predicted precipitation.

In this study, the sizes of population were set as 20 elements, namely there were 20 genes in a chromosome, and kept best 3 members of population. The orders of horizontal distances and elevation differences were

defined as gene groups, and with the real number type between 0 and 10. The mutation and crossover mechanics designated by users might influence the mining results, but in this case the effect is not remarkable. Significantly, GA is a tool to find an approximate value, not bound to be an optimal value. Sometimes, a sudden mutation would cause an unreasonable result, so intelligently artificial judgment is still necessary. Also, the operation processed about 150 times.

#### 2.4. Case study

The Feitsui reservoir watershed is located in Northern Taiwan and has a drainage area of 303 km<sup>2</sup>. The topography of this area is mountainous. The main backbone is the Snow Mountains and its branches. The elevation is between 50 and 1200 m. There are six rainfall stations in the Feitsui reservoir watershed, namely, Pinglin, Shisangu, Feitsui, Jiuqionggen, Bihu and Taiping. Fig. 3 shows the sub-watersheds and the distribution of rainfall stations in the Feitsui reservoir watershed. Table 1 lists the horizontal distances and elevation differences between the rainfall gauging stations each other in the Feitsui reservoir watershed.

The Feitsui reservoir watershed is located on subtropics area. The southwest current in summer does not greatly attack studied area as a result of the block of the Snow Mountains, but the weather is humid due to the influence of the northeast current. The thundershowers usually occur owing to the frequent current convection, and typhoon often brings plentiful rainfall and large rainfall intensity. Therefore, even though understanding rainfall characteristics in the Feitsui reservoir watershed is not a simple work, it is the first step to develop an efficient strategy for watershed management.

This case study was based on the daily rainfall records at each rainfall station in 2002. Separately assuming one rainfall gauge without precipitation data, and using the other five gauging stations to interpolate its precipitation is convenient for comparing the difference between predicted and observed precipitation. The objective function is to minimize the value of Sum(MAE) defined as Eq. (21), which is the sum of absolute daily estimation error. Then, the near optimal powers of inverse distances, i.e. the value of  $m$ , and  $n$ , can be found. In order to validate that the near optimal parameters in the modified inverse

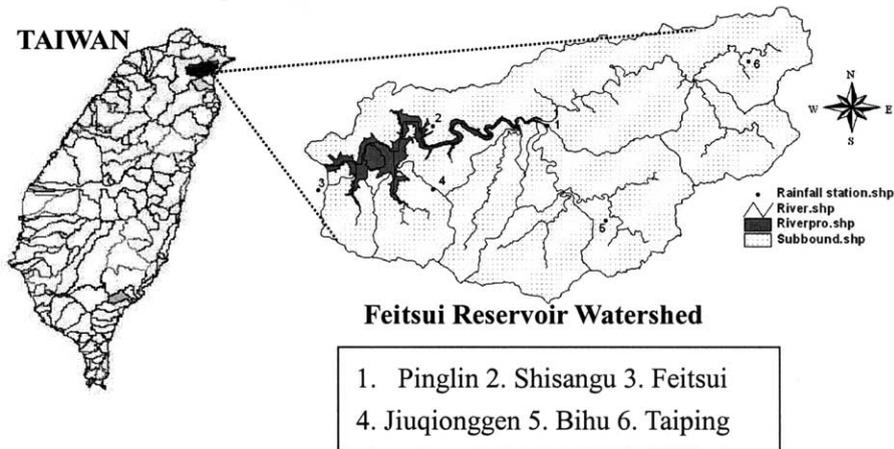


Fig. 3. Sub-watersheds and the distribution of rainfall stations in the Feitsui reservoir watershed.

distance method possess extensive applicability, the daily rainfall records in 2001 were used for parameter verification.

### 3. Results and discussion

#### 3.1. Applying fuzzy sets to interpolate precipitation

Traditional interpolation methods for estimating precipitation, such as the arithmetic average method

and the Thiessen Polygons method, were compared with a modified method that combines the inverse distance method with fuzzy logic. The analysis results show that the estimated error in the precipitation estimated by the modified IDW method is considerably reduced when precipitation interpolation is used, particularly for the western Feitsui reservoir watershed. Table 2 presents the errors estimated by a variety of methods for interpolating precipitation. The Thiessen Polygons method is better than other two methods at the Bihu and Taiping rainfall gauge.

Table 1

The horizontal distances and elevation differences between the rainfall stations each other in the Feitsui reservoir watershed

	Pinglin	Shisangu	Feitsui	Jiuqionggen	Bihu	Taiping
<i>Horizontal distances (m)</i>						
Pinglin	0	6729	14,011	7654	6468	12,377
Shisangu	6729	0	7760	3684	11,558	18,883
Feitsui	14,011	7760	0	6769	17,007	26,382
Jiuqionggen	7654	3684	6769	0	10,317	19,953
Bihu	6468	11,558	17,007	10,317	0	12,514
Taiping	12,377	18,883	26,382	19,953	12,514	0
Ave-d (m)	9448	9723	14,386	9675	11,573	18,022
Std-d (m)	3496	5843	7950	6214	3810	5843
<i>Elevation differences (m)</i>						
Pinglin	0	320	8	168	176	250
Shisangu	320	0	328	152	144	70
Feitsui	8	328	0	176	184	258
Jiuqionggen	168	152	176	0	8	82
Bihu	176	144	184	8	0	74
Taiping	250	70	258	82	74	0
Ave-h (m)	184	203	191	117	117	147
Std-h (m)	116	115	119	71	75	98

Table 2  
The estimated errors of precipitation by several interpolation methods

Methods	Arithmetic		Thiessen		Fuzzy		<i>m</i>	<i>n</i>	Type
	Sum(MAE) (mm)	RE	Sum(MAE) (mm)	RE	Sum(MAE) (mm)	RE			
Pinglin	830	0.44	731	0.39	687	0.36	2	1	Min
Shisangu	1069	0.56	1090	0.57	810	0.42	4	2	Min
Feitsui	1171	0.57	1170	0.57	716	0.35	6	1	Min
Jiuqionggen	768	0.36	748	0.35	556	0.26	6	4	Min
Bihu	1365	0.56	997	0.41	1181	0.49	4	2	Multi
Taiping	1841	0.57	1410	0.43	1714	0.53	1	6	Multi

min, minimum operation; multi, multiplication operation. (Note: The result is based on the daily rainfall records in 2002).

However, at the Pinglin, Shisangu, Feitsui and Jiuqionggen rainfall gauging stations, the modified IDW method, incorporating fuzzy theory, is much better for estimating precipitation. The error estimated using the arithmetic average method is large at all rainfall stations. Adopting the modified method, which combines the IDW method with fuzzy theory, reduces the estimated errors at the Pinglin, Shisangu, Feitsui, Jiuqionggen, Bihu and Taiping rainfall gauging stations by 8%, 14%, 22%, 10%, 8% and 4% from those obtained using the arithmetic average method.

Fig. 3 indicates that the relationship between the Pinglin and Jiuqionggen stations and all of the surrounding rainfall stations is even, explaining why the estimated error on precipitation interpolation can normally be controlled. The estimated error at the Taiping rainfall station cannot effectively be reduced by any method, because the average horizontal distance between the Taiping rainfall gauge and its nearby rainfall stations is so large. Significantly, although the horizontal distances between the Feitsui rainfall station and its neighboring gauges are also large, the modified IDW method, applying fuzzy theory, still substantially reduces the estimated error because the differences between the elevations of the Feitsui rainfall station and its nearby gauges are large and various, and the effect of elevation must be considered in lowering the estimated error. The horizontal distances between the Bihu station and its surrounding rainfall gauges are large, but one rainfall gauge is much closed than the others, so the improvement by the modified fuzzy method in estimating the precipitation is not very remarkable.

Several operations, including square root, addition, multiplication, maximum and minimum, were used to integrate the effects obtained at each rainfall gauging station to combine the membership functions of the horizontal distances and the differences between elevations. The results show that the minimum operation is the best for merging the effect of horizontal distances and differences between elevations; the multiplication operation is second-best.

The approximate optimal orders of the inverse horizontal distances, *m*, at the western rainfall stations in the Feitsui reservoir watershed, always exceed the near-optimal orders of the inverse differences between the elevations, *n*. Therefore, the relative importance of each rainfall gauge, governed by the horizontal distances, declines more quickly toward the farther stations than does that due to the effect of the differences between the elevations. Fig. 4 presents

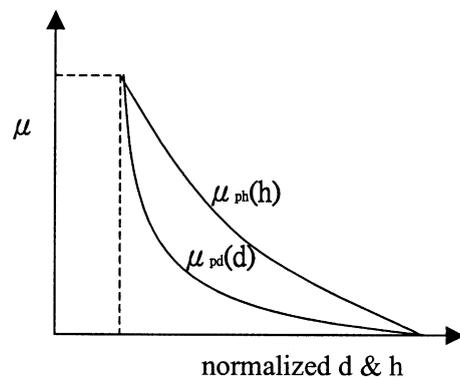


Fig. 4. Comparison of the membership function of horizontal distances and elevation differences.

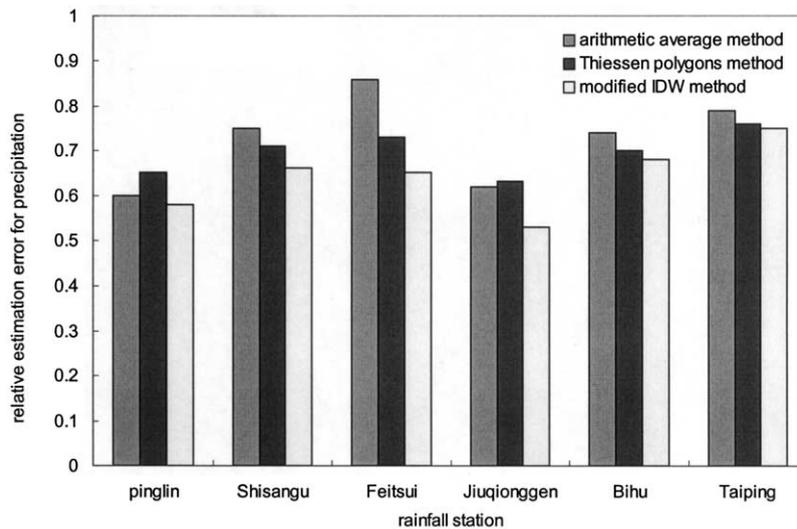


Fig. 5. Relative estimated error of precipitation by several interpolation methods at each rainfall station in the Feitsui reservoir watershed. (Note: The result is based on the daily rainfall records in 2001).

the relative effect of horizontal distances and differences between elevations.

The near optimal powers used in the modified IDW method were sought, based on daily rainfall records from 2002. The rainfall data from 2001 were used to verify the application of these parameters. The results in Fig. 5 show that the modified IDW method can reduce the estimated error of the precipitation, particularly when the standard deviation of the difference between the elevations of the interpolated area and its surrounding rainfall gauges is large.

### 3.2. Tendencies of the powers of the inverse distances

The orders of the inverse distances,  $m$  and  $n$ , affect the accuracy of the precipitation interpolation. Fig. 6 presents the sum of MAE over a whole year versus  $n$  for fixed  $m$ . The curves for the square root, addition and maximum operations are similar. At the Pinglin rainfall station, most of the curves decline first, before reaching a peak with a minimum estimated error, before rising smoothly, ultimately reaching a constant value. Most of the peak points are at an  $n$  value of between 0 and 2, regardless of the value of  $m$ . However, the curves for the multiplication and minimum operations are less regular. Although they

also have a peak point, the range of the optimal values  $n$  varies with  $m$ . The curves at the Shisangu gauge always rise initially, and before reaching a constant. Restated, the estimated errors are always minimum at  $n=0$ . The profile at the Jiuqionggen station is similar to that of the Pinglin gauge, since the spatial distributions of their surrounding rainfall gauges are alike. Likewise, the profile at the Feitsui station resembles that at the Bihu station because their neighboring gauging stations are similarly distributed.

Fig. 7 plots the sum of MAE over a whole year versus  $m$  for a fixed  $n$ . The curves for the square root, addition and maximum operations are similar. The curves for the multiplication and minimum operations differ more, so the minimum estimated errors—the peak points in the profiles—vary with  $n$ . The curves for the multiplication and minimum operations at the Shisangu, Feitsui and Jiuqionggen gauges are unusual. These curves are initially flat at first, and then decline. Accordingly, the near-optimal order of the inverse horizontal distance,  $m$ , always exceeds zero. All the curves at the Taiping rainfall gauging station on each operation are alike. Therefore, the sensitivity of the change in both  $m$  and  $n$  does not strongly influence the sum of MAE over a whole year versus the order of the inverse distance.

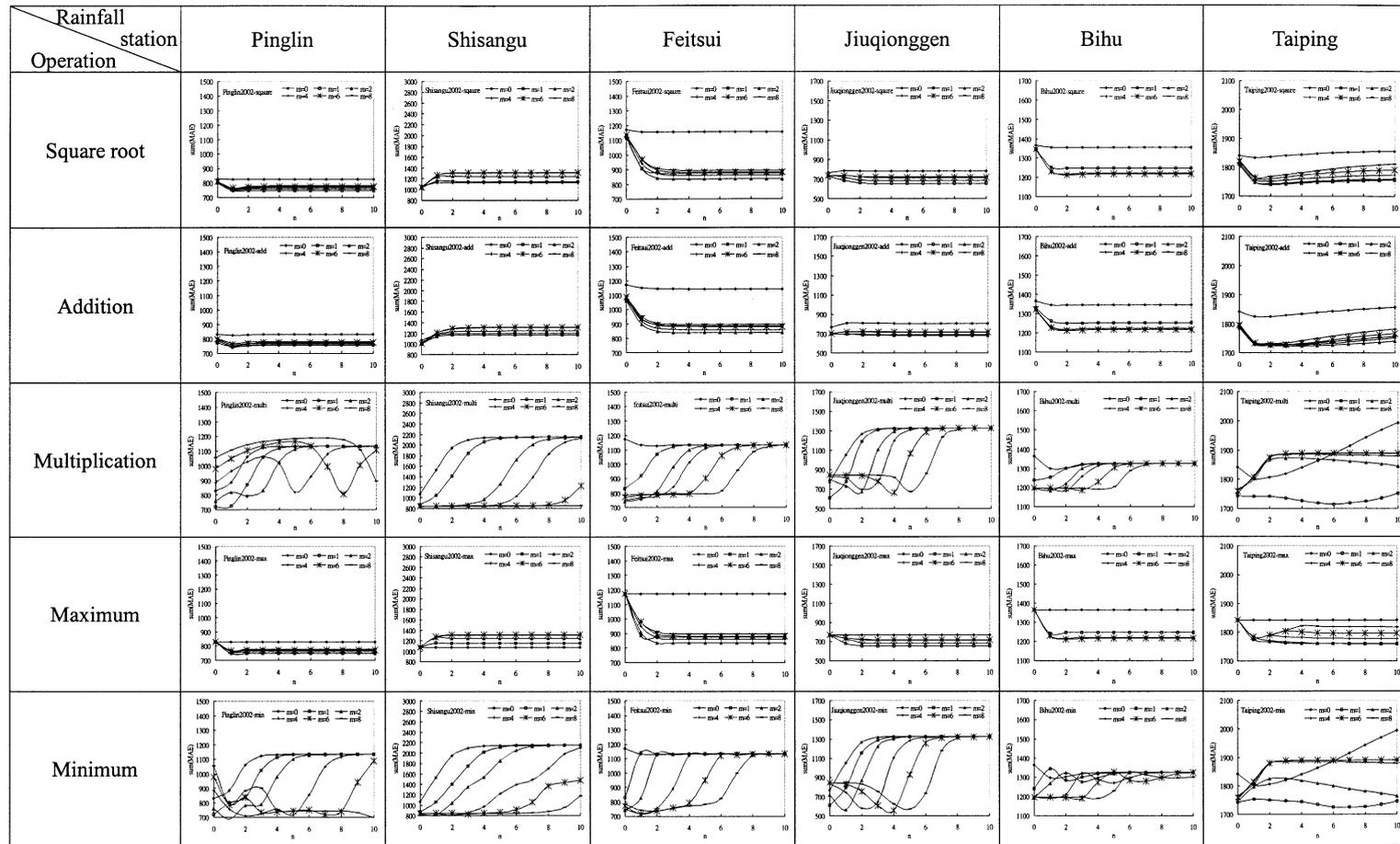


Fig. 6. The profile of sum(MAE) vs.  $n$  at each rainfall station in the Feitsui reservoir watershed.

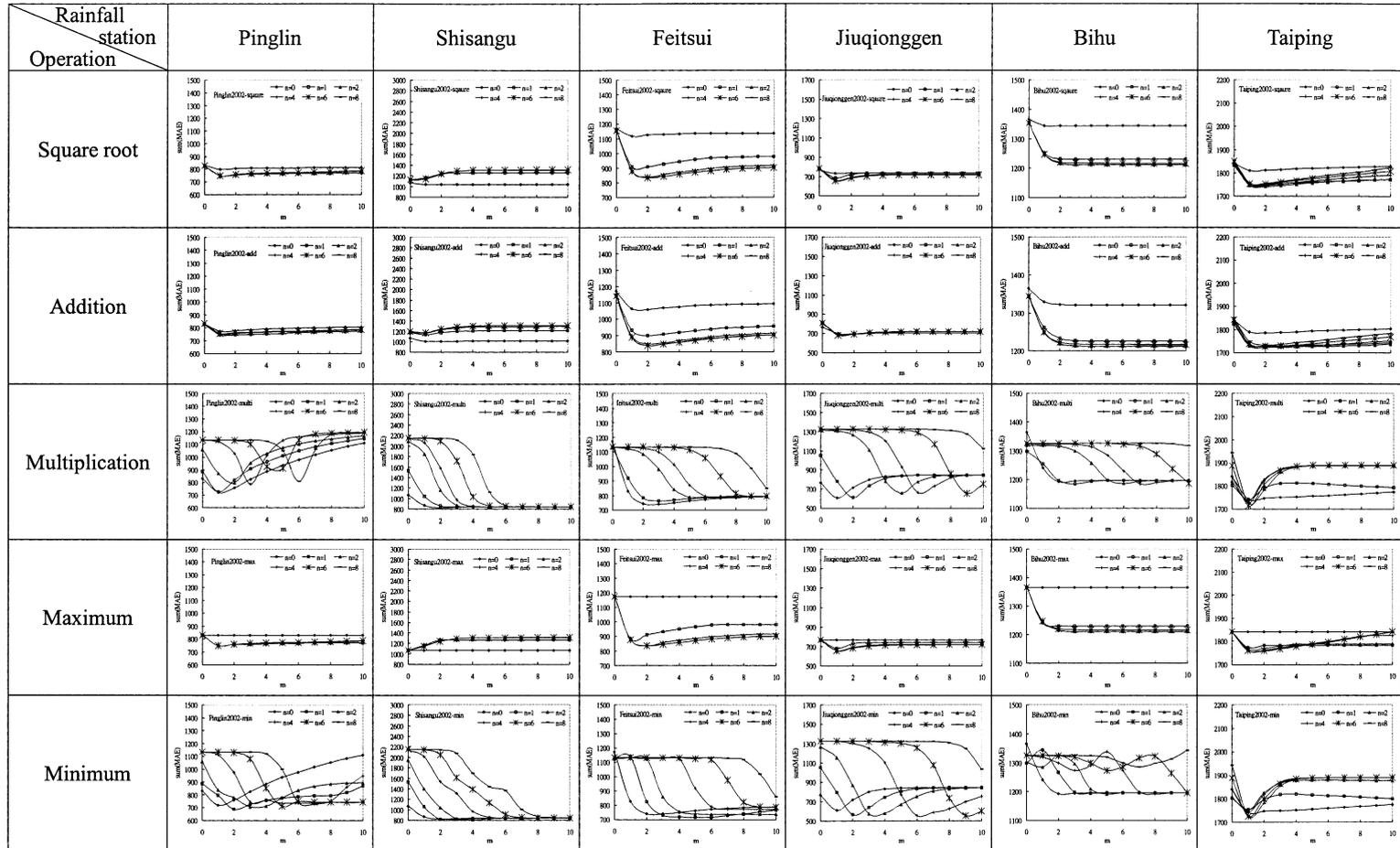


Fig. 7. The profile of sum(MAE) vs.  $m$  at each rainfall station in the Feitsui reservoir watershed.

#### 4. Conclusion

The modified IDW method incorporates not only the effect of the horizontal distances, but also the effect of the difference between elevations. Fuzzy theory is a useful technique for describing the indistinct relative importance of the rainfall gauging stations. Also, GA technology facilitates the search for the near-optimal orders of the inverse distances and the operation that integrates various factors.

Although the modified method discussed herein is not supposed to be suited to any distribution of rainfall stations, the results nevertheless confirm that the method is flexible, and usually much better than traditional methods for estimating precipitation. Certainly, the modified method, which combines the IDW method with fuzzy theory, always outperforms the arithmetic average method. In particular, when the differences between the elevations of the rainfall gauge in question and the surrounding rainfall gauges are significant, the modified IDW method can greatly improve the estimated precipitation. Moreover, the minimum and multiplication operations for integrating the factors that affect the relative importance of the nearby rainfall stations are the most effective for describing the actual situation.

Accurately predicting the future rainfall characteristics is difficult. However, the flexible and valuable modified IDW method discussed in this study is a more effective method for estimating reliable precipitation, which is of value to researchers and watershed managers. The accurate modeling results provide a valuable reference for efforts to develop a watershed management strategy based on appropriate rainfall input.

#### References

- Abdel-Kader, M.G., Dugdale, D., Taylor, P., 1998. Investment Decisions in Advanced Manufacturing Technology—A Fuzzy Set Theory Approach. Ashgate Publishing Company, VT, USA.
- Ashraf, M., Loftis, J.C., Hubbard, K.G., 1997. Application of geostatistics to evaluate partial weather station networks. *Agricultural and Forest Meteorology* 84, 255–271.
- Bardossy, A., Bogardi, I., Duckstein, L., 1990. Fuzzy regression in hydrology. *Water Resources Research* 26 (7), 1497–1508.
- Bartier, P.M., Keller, C.P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Computers & Geosciences* 22 (7), 795–799.
- Chang, C.L., Lo, S.L., Yu, S.L., Hu, C.Y., 2003. Combining variable order inverse distance method and genetic algorithm to precipitation interpolation in Feitsui Reservoir Watershed. Asian Waterqual 2003 Conference, Thailand.
- Chaubey, I., Haan, C.T., Grunwald, S., Salisbury, J.M., 1999. Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology* 220, 48–61.
- Cheng, C.H., 1998. A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems* 95, 307–317.
- Cheng, C.T., 1999. Fuzzy optimal model for the flood control system of the upper and middle reaches of the Yangtze River. *Hydrological Sciences Journal* 44 (4), 573–582.
- Cheng, C.T., Ou, C.P., Chau, K.W., 2002. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology* 268, 72–86.
- Craig, E.H., Karen, A.K., 1995. To normalize or not to normalize? Fat is the question. *Environmental Toxicology and Chemistry* 14 (5), 801–807.
- Dirks, K.N., Hay, J.E., Stow, C.D., Harris, D., 1998. High-resolution studies of rainfall on Norfolk island, part II: interpolation of rainfall data. *Journal of Hydrology* 208, 187–193.
- Donald, L.P., Danny, G.M., 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. *Ecological Modeling* 91, 213–229.
- Faures, J.M., Goodrich, D.C., Woolhiser, D.A., Soroosh, S., 1995. Impact of small-scale spatial variability on runoff modeling. *Journal of Hydrology* 173, 309–326.
- Hulme, M., Conway, D., Jones, P.D., Jiang, T., Barrow, E.M., Turney, C., 1995. Construction of a 1961–1990 European climatology for climate change modeling and impact application. *International Journal of Climatology* 15, 1333–1363.
- Klir, G., Folger, T.A., 1988. *Fuzzy Sets, Uncertainty, and Information*. Prentice-Hall, Englewood Cliffs, NJ.
- Klir, G.J., Yuan, B., 1995. *Fuzzy Sets and Fuzzy Logic-Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.
- Lagacherie, P., Cazemier, D.R., Vangaans, P.F.M., Burrough, P.A., 1997. Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. *Geoderma* 77, 197–216.
- Lung, W.S., 2001. *Water Quality Modeling for Wasteload Allocations and TMDLs*. Wiley, New York.
- Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology* 92, 211–225.
- Perret, J.S., Prasher, S.O., 1998. Applications of fuzzy-logic in the design of vegetated waterways under uncertainty. *Journal of the American Water Resources Association* 34 (6), 1355–1367.
- Price, D.T., Mckenny, D.W., Nalder, I.A., Hutchinson, M.F., Kesteven, J.L., 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology* 101, 81–94.

- Ritzel, B.J., Wayland, E.J., 1994. Using genetic algorithms to solve a multiple objective groundwater pollution containment problem. *Water Resources Research* 30 (5), 1589–1603.
- Savic, D.A., Walters, G.A., Davidson, J.W., 1999. A genetic programming approach to rainfall-runoff modeling. *Water Resources Management* 12, 219–231.
- Tabios, G.O., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Research* 21 (3), 365–380.
- Tung, C.P., Hsu, S.Y., Liu, C.M., Li, J.S., 2003. Application of the genetic algorithm for optimizing operation rules of the LiYuTan reservoir in Taiwan. *Journal of the American Water Resources Association* 39, 649–657.
- Vicente, L.L., 1996. On the effect of uncertainty in spatial distribution of rainfall on catchment modeling. *Catena* 28, 107–119.
- Wang, Q.J., 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resources Research* 27 (9), 2467–2471.
- Ware, C., Knight, W., Wells, D., 1991. Memory intensive algorithms for multibeam bathymetric data. *Computers & Geosciences* 17 (7), 985–993.
- Wu, S.M., Huang, G.H., Guo, H.C., 1997. An interactive inexact-fuzzy approach for multiobjective planning of water-resource systems. *Water Science and Technology* 36 (5), 235–242.
- Yu, P.S., Yang, T.C., 2000. Fuzzy multi-objective function for rainfall-runoff model calibration. *Journal of Hydrology* 238, 1–14.
- Zadeh, L., 1965. Fuzzy sets. *Information and Control* 8, 338–353.
- Zhu, A.X., Mackay, D.S., 2001. Effects of spatial detail of soil information on watershed modeling. *Journal of Hydrology* 248, 54–77.