THE PARAMETER OPTIMIZATION IN THE INVERSE DISTANCE METHOD BY GENETIC ALGORITHM FOR ESTIMATING PRECIPITATION

CHIA-LING CHANG[1,*], SHANG-LIEN LO[1] and SHAW-L YU[2]

[1]*Graduate Institute of Environmental Engineering, National Taiwan University, Taipei 106, Chinese Taiwan;* [2]*Department of Civil Engineering, University of Virginia, Charlottesville, Virginia, USA*
(*author for correspondence, e-mail: f89541201@ntu.edu.tw)*

Abstract. The inverse distance method, one of the commonly used methods for analyzing spatial variation of rainfall, is flexible if the order of distances in the method is adjustable. By applying the genetic algorithm (GA), the optimal order of distances can be found to minimize the difference between estimated and measured precipitation data. A case study of the Feitsui reservoir watershed in Taiwan is described in the present paper. The results show that the variability of the order of distances is small when the topography of rainfall stations is uniform. Moreover, when rainfall characteristic is uniform, the horizontal distance between rainfall stations and interpolated locations is the major factor influencing the order of distances. The results also verify that the variable-order inverse distance method is more suitable than the arithmetic average method and the Thiessen Polygons method in describing the spatial variation of rainfall. The efficiency and reliability of hydrologic modeling and hence of general water resource management can be significantly improved by more accurate rainfall data interpolated by the variable-order inverse distance method.

Keywords: genetic algorithm, inverse distance, precipitation interpolation, rainfall, spatial variation, watershed hydrology

1. Introduction

Accuracy of input data and parameters in modeling decides if the modeling results are reliable. The input data and parameters needed in watershed modeling are always with spatial variation. Incomplete understanding of spatial variability and the lack of adequate data to characterize spatially varying inputs and state variables would hamper the development of distributed modeling technology. Particularly, the scarcity of spatial variation of rainfall information would cause much uncertainty in discharge and erosion predictions (Vicente, 1996). Even at small-scale spatial variability of precipitation can translate into large variations in simulated runoff (Faures *et al.*, 1995). A large uncertainty in the estimated parameters resulted from spatial variability of rainfall. The uncertainty in the estimated parameters using the rainfall observed by a single gauge could exceed the rainfall measurement error (Chaubey *et al.*, 1999).

The grid input system is usually designed in models to convenient the expression of spatial variability. By the grid input, we can describe the data with spatial variation

in a watershed. However, although the modeling with grid input design is a profitable way to represent spatial variation in a watershed, spatial information is not enough everywhere. The monitor stations such as rainfall stations and flow stations are not all over each grid. Therefore, using the limited monitoring data to interpolate the unknown information is important to adequately describe the spatial variability of parameters.

There is not any interpolation method suitable for every circumstance (Nalder and Wein, 1998). The results of spatially distributed models would be highly influenced by the interpolation errors (Donald and Danny, 1996). Several methods, including the optimal interpolation method, the Kriging method and the weighted mean method, etc., have been applied to analyze spatial variation of rainfall for many years. The optimal interpolation method (Tabios, 1985) is complicated, and completely monitored data are required. The complication of required information would increase the difficulty of analyses, particularly when the data are not enough. In the Kriging method, statistical hypotheses are made in evaluating and identifying the multidimensional spatial structure of the hydrological process of interest (Todini and Ferraresi, 1996; Dirks *et al.*, 1998a, b). If rainfall stations are not sufficient, the statistical data will be limited, and the Kriging method will not be suitable for estimating precipitation. Moreover, the weighted mean method (Troutman, 1983) is a popular interpolation method. The unknown precipitation or average precipitation can be calculated after deciding the weighting factors for each surrounding rainfall station. The arithmetic average method, the Thiessen Polygons method, the K-nearest method and the inverse distance method all belong to the weighted mean method. The only difference between these methods is how to allot weighting factors for the neighboring rainfall gauges.

The order of distances is used in the inverse distance method to vary the allotment of weighting factors for the vicinal rainfall stations. In other words, the relative influences of each surrounding rainfall station on precipitation estimation will be different when changing the order of distances. Due to the flexibility of the order of distances, the estimation error for precipitation could be highly reduced by adjusting the order of distances to reach the optimal value. The optimal value of the order of distance can make the allotment of weighting factors for each neighboring rainfall gauge suitable for calculating unknown precipitation. The objective of this research was to find the most proper variable order of distances in different circumstances with varied rainfall characteristics and to minimize the difference between estimated and measured precipitation.

2. Methodology

2.1. CASE STUDY

The Feitsui reservoir watershed in northern Taiwan provided the setting for the case study reported herein. It is located on the southeast of the capital city of Taipei
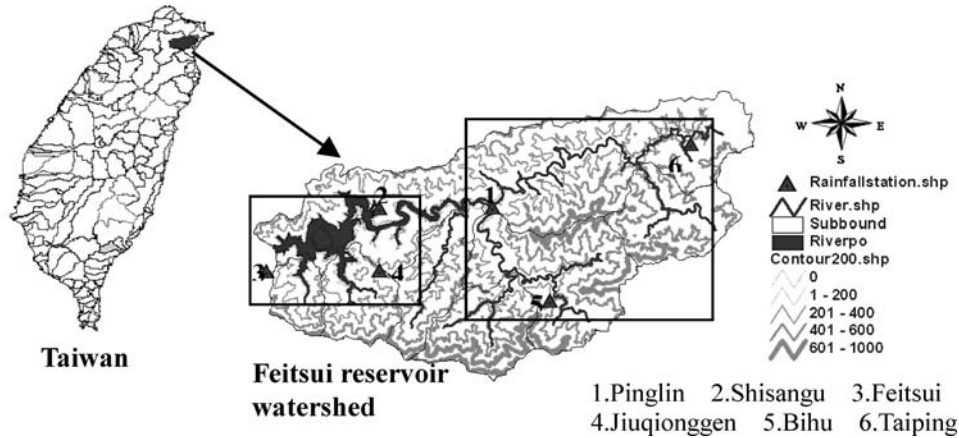
Figure 1. Sub-watersheds and the distribution of rainfall stations in the Feitsui reservoir watershed.

and has a drainage area of 303 square kilometers. The topography of this area is mountainous. The main backbone is the Snow Mountains and its branches. The elevation is between 50 and 1200 meters. There are six rainfall stations in the Feitsui reservoir watershed, namely, Pinglin, Shisangu, Feitsui, Jiuqionggen, Bihu and Taiping. Figure 1 shows the sub-watersheds and the distribution of rainfall stations in the Feitsui reservoir watershed.

In the present study, rainfall data for the six stations from 1995 to 1999 were used to demonstrate the application of the variable-order inverse distance method. Assuming the precipitation at one of the rainfall stations is unknown and applying the rainfall data from the other five rainfall stations to estimate the unknown precipitation is the major method convenient to compare the difference between estimated and measured precipitation. Rainfall characteristics and horizontal distances between interpolated locations and all of the rainfall stations in a watershed are significant factors for the analyses of spatial variation of rainfall. This study considered these two factors and apply genetic algorithm (GA) to search the most suitable order of distances in the variable-order inverse distance method. The adequate weighting factors can be decided after the optimal order of distances is determined and unknown precipitation can be interpolated close to actual precipitation.

2.2. The inverse distance method

The inverse distance method is a commonly used method for watershed precipitation interpolation. The horizontal distances between the location without precipitation records and its surrounding rainfall stations, as Figure 2, and the order of distances decide the weights given to each rainfall station (Bartier and Keller, 1996). Horizontal distances between rainfall stations are constant, but the order of distances
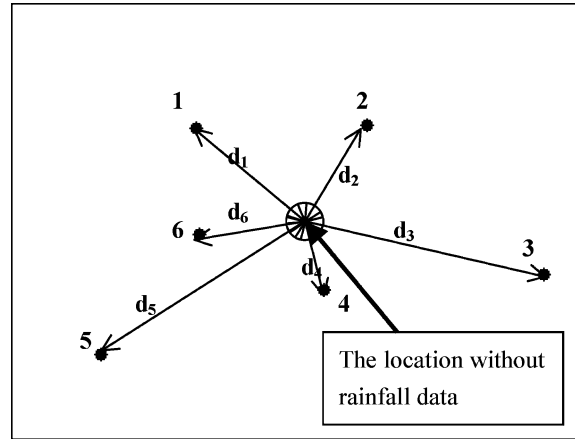
Figure 2. Horizontal distances between the location without precipitation records and its surrounding rainfall stations.

is adjustable. When the order of distances is adjusted to reach an optimal value, the allotment of weighting factors for the surrounding rainfall gauges is suitable to present more actual relative influences of each rainfall station on precipitation estimation, and the accuracy of precipitation interpolation can be improved as well. Due to the alterability of the order of distances in the inverse distance method, this method is renamed as the variable-order inverse distance method. The weight for precipitation at rainfall station i in sub-watershed p, W_{pi}, is given by

$$W_{pi} = \frac{\left(\frac{1}{d_{pi}^m}\right)}{\left(\sum\limits_{i=1}^{n} \frac{1}{d_{pi}^m}\right)} \tag{1}$$

where m is the order of distances; d_{pi} is the distance between the center of sub-watershed p and rainfall station i; n is the total number of rainfall stations. After the weights are decided, unknown precipitation for each sub-watershed could be estimated. The precipitation of sub-watershed p, P_p, is defined as formula (2), in which P_i is the precipitation at rainfall station i.

$$P_p = \sum_{i=1}^{n} (W_{pi} * P_i) \tag{2}$$

The arithmetic average method and the Thiessen Polygons method are special cases of the variable-order inverse distance method. When the value of "m" is zero, the variable-order inverse distance method is the same as the arithmetic average method for precipitation interpolation. The weighting factors for precipitation at each rainfall station for the sub-watershed are the same. On the other hand, when
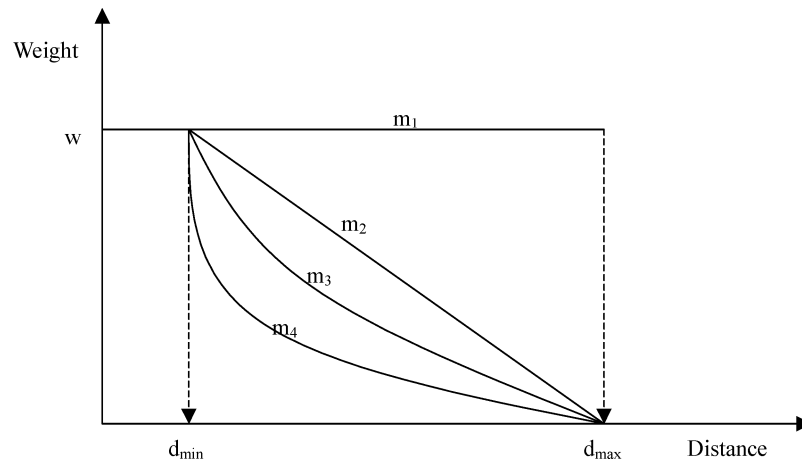
Figure 3. The correlation between the weights for precipitation at each rainfall station and horizontal distances to the location with no rainfall records. (Note: w is a weighting value. $m_1 = 0$; $m_4 > m_3 > m_2 > m_1$).

the value of "m" is close to infinite, the variable-order inverse distance method is similar to the Thiessen Polygons method. As shown in Figure 3, "m_1" is zero, so the weights for precipitation at each rainfall station with any horizontal distance are the same. Moreover, $m_1 < m_2 < m_3 < m_4$, when the value of "m" increases, the estimated precipitation is more influenced by the closer rainfall stations.

2.3. APPLYING GENETIC ALGORITHM

GA is inspired by Darwin's theory about evolution. The algorithm is started with a set of solutions called population. Gene encoding is the first step to apply GA. The object is a genome. After a population of genomes is created, crossover and mutation occur persistently. This process is significant to select the individuals in the population and generate new individuals. The objective function discriminates how good each individual is. The flow chart of GA is shown in Figure 4.

GA has been verified to have more advantages than the classical optimization methods (Goldberg, 1989). In recent years, it has been a popular technique for solving hydrology and water resources problems (Wang, 1991; Cheng *et al.*, 2002). In this work, GA was applied to adjust the value of the order of distances in the inverse distance method to suitably describe spatial variation of rainfall. Due to the evolution of genes, the worse individual can be eliminated, and the optimal individual fit the objective function can be selected automatically. Hence, using the GA can significantly reduce the difficulty of adjusting the order. Moreover, adjusting the order of distances in the inverse distance method can lessen the limitation on analyzing the spatial variation of rainfall.
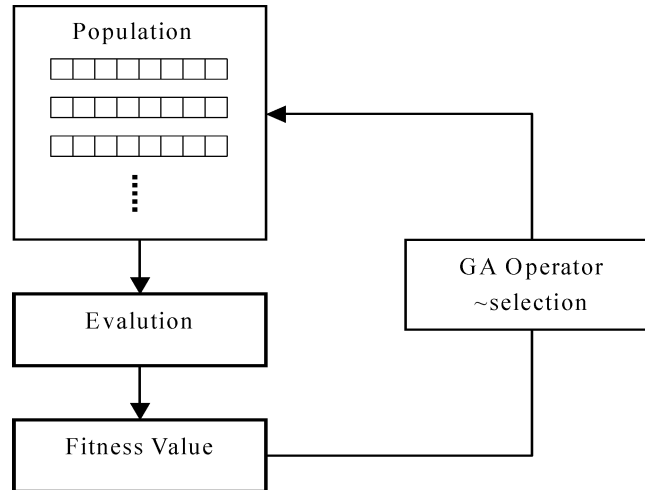
Figure 4. The flow chart of genetic algorithm.

The criterion applied herein is accumulated absolute error (AABSE), which is the sum of the difference between measured and estimated daily precipitation for the five-year period. The major objective in the process of parameter optimization was to minimize the difference between measured precipitation and estimated precipitation. The objective function is shown as follows:

$$\text{Min AABSE} = \sum_{j=1}^{k} |(O_j - P_j)| \tag{3}$$

TABLE I
The optimal value of "m" and the sort of the optimal value of "m"

	Pinglin	Shisangu	Feitsui	Jiuqionggen	Bihu	Taiping	
The optimal m in 1995	4.0	2.6	4.7	1.6	4.9	9.2	
Sort	3	2	4	1	5	6	†
The optimal m in 1996	5.7	2.3	3.7	1.4	19.3	9.9	
Sort	4	2	3	1	6	5	
The optimal m in 1997	20.0	1.7	3.4	0.8	4.3	6.0	
Sort	6	2	3	1	4	5	
The optimal m in 1998	2.1	1.6	2.9	2.3	19.7	19.9	
Sort	3	1	4	2	6	5	
The optimal m in 1999	3.1	1.8	3.4	1.6	4.6	9.3	
Sort	3	2	4	1	5	6	†

Note. The same sort of the optimal value of "m" is noted by the mark "†".
Sort→The least value of the optimal "m" is given by one.

where O_j is the *jth* observed precipitation, and P_j is the *jth* predicted precipitation; k is the number of rainfall records over the five years. 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 order of horizontal distances was defined as a gene group, and with the real number type between 0–20. The mutation and crossover mechanics designated by users might influence the mining results, but in this case the effect is not remarkable. The operation processed about 100 times.

3. Results and Discussion

The optimal values of the order of distances on the locations of each rainfall station in the Feitsui reservoir watershed from 1995 to 1999 are found by GA and presented in Table I. The results show that the value of variable "m" is close to 2 for Shisangu rainfall station, between 3 and 5 for Feitsui rainfall station and between 1 and 2 for

TABLE II
Rainfall characteristics of six rainfall stations in the Feitsui reservoir watershed

Years	Rainfall characteristics	Rainfall stations					
		Pinglin	Shisangu	Feitsui	Jiuqionggen	Bihu	Taiping
1995	Mean daily precipitation	5.74	5.10	5.88	5.94	6.43	10.56
	Standard deviation of daily precipitation	11.42	10.12	11.61	11.45	15.22	20.43
1996	Mean daily precipitation	9.28	7.67	7.99	9.05	11.06	16.90
	Standard deviation of daily precipitation	30.23	23.22	24.22	30.21	34.90	39.71
1997	Mean daily precipitation	7.62	7.43	9.61	9.29	7.74	11.04
	Standard deviation of daily precipitation	29.35	17.22	22.85	29.40	26.48	28.07
1998	Mean daily precipitation	14.38	13.13	14.58	12.96	15.24	21.64
	Standard deviation of daily precipitation	37.24	29.06	33.43	28.16	42.99	50.70
1999	Mean daily precipitation	7.11	6.90	7.01	5.42	7.77	11.97
	Standard deviation of daily precipitation	16.94	14.96	15.85	11.17	16.14	24.00

Unit: mm
Note: The standard deviation of daily precipitation, "S", is defined by

$$S = \sqrt{\frac{\sum_{i=1}^{k} (P_i - \bar{P}_i)^2}{k - 1}}$$

where k is the number of rainfall records; P_i is the ith rainfall record (daily estimates); $\bar{P}_i$ is the average precipitation of all the rainfall records over a time-period.

Jiuqionggen rainfall station. The locations of Shisangu, Feitsui and Jiuqionggen rainfall stations are near the towns of Xindian and Shiding. As shown in Figure 1, the variability and roughness of topography in this area is less than those in the area of Pinglin, Bihu and Taiping rainfall stations. Hence, the values of the variable "m" are more uniform and not extreme.

The main effects on the order of distances in the inverse distance method, the variable "m" in formula (1), are rainfall characteristics and horizontal distances
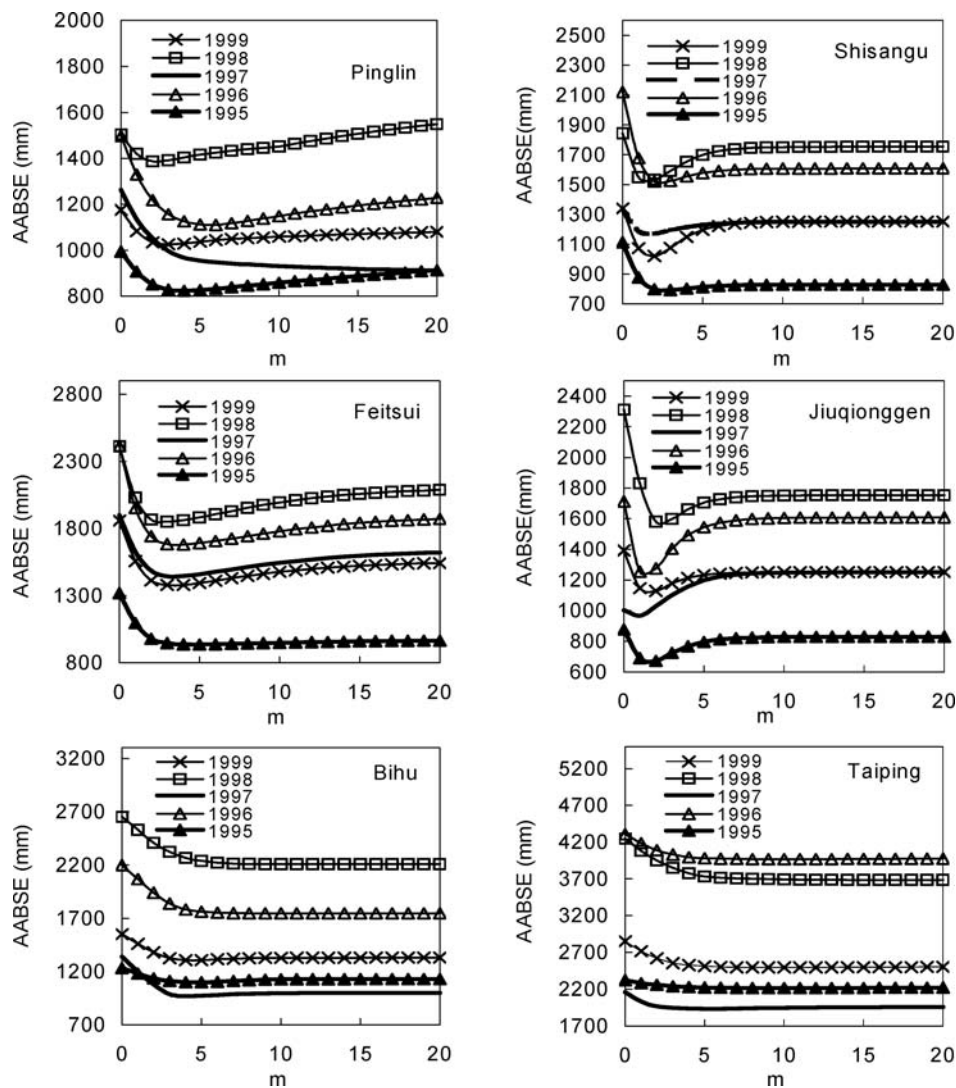

Figure 5. Accumulated absolute errors (AABSE) of precipitation interpolation for different "m" values at the rainfall stations in the Feitsui reservoir watershed.

between rainfall stations and interpolated locations. In the situation with uniform rainfall characteristic, the value of "m" always varies only following the difference of horizontal distances between the location with no rainfall records and all of the rainfall stations in a watershed. Table II shows the rainfall data from 1995 to 1999, including mean daily precipitation and standard deviation of daily precipitation.

The standard deviations of daily precipitation in 1995 and 1999 are between 10 and 20 mm, much smaller than those in 1996, 1997 and 1998. Therefore, when the smallest value of the optimal "m" is determined as 1, the sorts of the optimal orders of distances among these rainfall stations are the same in 1995 and 1999, marked as "†" in Table I. On the contrary, the standard deviations of daily precipitation in 1996, 1997 and 1998 are bigger than in 1995 and 1999, and some extreme situations
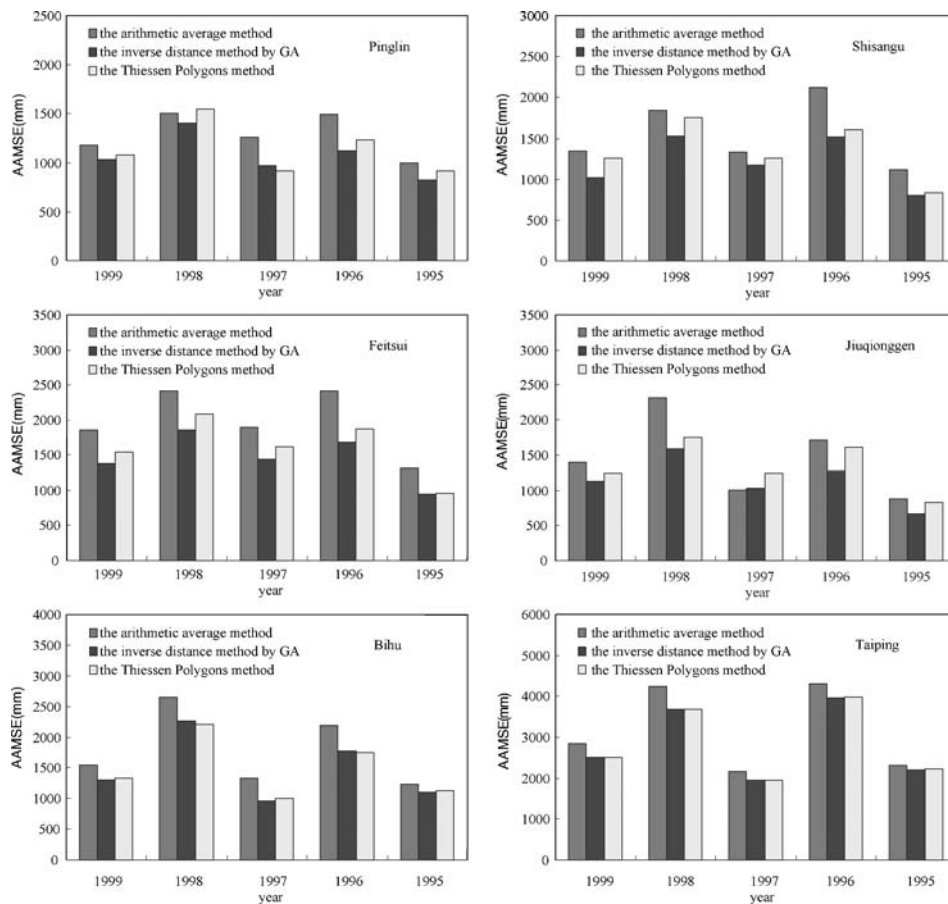

Figure 6. Comparison of between the method combining the inverse distance methods and other traditional methods for estimating precipitation.

occur, like bigger than 50 mm. In these situations, not only horizontal distances between interpolated locations and rainfall stations, but also other factors would influence the value of variable "m". Consequently, there is no regularity on the sort of the optimal orders of distances among these rainfall stations.

Figure 5 shows the relationships between the AABSE and the value of "m". Although the optimal value of "m" is different, theoretically it should be between zero and infinite. The results from the present analysis show that the optimal value of "m" is neither zero nor infinite, and that most of the optimal "m" values found in the present study take on values greater than zero and less than five. Also, the trends representing the relation between AABSE and the variable "m" are the same. The value of "m" larger than 20 is considered an extreme situation. It means there would be some other factors influencing on the variable "m" and the estimation error for precipitation would be noticeable if we do not consider these effects. Figure 6 shows that the arithmetic average method on precipitation interpolation usually causes larger AABSE than any other methods, and the Thiessen Polygons method is usually not the optimal method to truly describe rainfall spatial variation. Meanwhile, the results also verify the variable-order inverse distance method is more suitable than the arithmetic average method and the Thiessen Polygons method to describe the spatial variation of rainfall.

4. Conclusions

This research selected the Feitsui reservoir watershed for case study. Combining the variable-order inverse distance method and GA is a way to find the optimal analysis on spatial variation of rainfall. The results are concluded as follows:

- The variability of the order of distances is small as the topography of rainfall stations is uniform.
- When the standard deviation of precipitation is small and rainfall characteristic is uniform, the horizontal distance between rainfall stations and interpolated locations is the major factor influencing the order of distances.
- The variable-order inverse distance method is more suitable than the arithmetic average method and the Thiessen Polygons method for describing the spatial variation of rainfall.

References

Bartier, P. M. and Keller, C. P.: 1996, 'Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW)', *Comput. Geosci.* **22**(7), 795–799.

Chaubey, I., Haan, C. T., Grunwald, S. and Salisbury, J. M.: 1999, 'Uncertainty in the model parameters due to spatial variability of rainfall', *J. Hydrol.* **220**, 48–61.

Cheng, C. T., Ou, C. P. and Chau, K. W.: 2002, 'Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration', *J. Hydrol.* **268**, 72–86.

Dirks, K. N., Hay, J. E., Stow, C. D. and Harris, D.: 1998, 'High-resolution studies of rainfall on Norfolk island, Part I: The spatial variability of rainfall', *J. Hydrol.* **208**, 163–186.

Dirks, K. N., Hay, J. E., Stow, C. D. and Harris, D.: 1998, 'High-resolution studies of rainfall on Norfolk island, Part II: Interpolation of rainfall data', *J. Hydrol.* **208**, 187–193.

Donald, L. Ph. and Danny, G. M.: 1996, 'Spatial uncertainty analysis: Propagation of interpolation errors in spatially distributed models', *Ecol. Model.* **91**, 213–229.

Faures, J. M., Goodrich, D. C., Davis, A. W. and Soroosh, S.: 1995, 'Impact of small-scale spatial variability on runoff modeling', *J. Hydrol.* **173**, 309–326.

Goldberd, D. E.: 1989, *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley, MA.

Nalder, I. A. and Wein, R. W.: 1998, 'Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest', *Agric. For. Meteorol.* **92**, 211–225.

Tabios, G. O. and Salas, J. D.: 1985, 'A comparative analysis of techniques for spatial interpolation of precipitation', *Water Resour. Res.* **21**, 365–380.

Todini, E. and Ferraresi, M.: 1996, 'Influence of parameter estimation uncertainty in Kriging', *J. Hydrol.* **175**, 555–566.

Troutman, B. M.: 1983, 'Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation', *Water Resour. Res.* **19**(3), 791–810.

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 Resour. Res.* **27**(9), 2467–2471.