A mathematical spatial interpolation method for the estimation of convective rainfall distribution over small watersheds

Shengtang Zhang1†, Jingzhou Zhang1, Yin Liu2, Yuanchen Liu1

1College of Earth Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266590, P. R. China
2College of Mining and Safety Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266590, P. R. China

ABSTRACT
Rainfall is one of crucial factors that impact on our environment. Rainfall data is important in water resources management, flood forecasting, and designing hydraulic structures. However, it is not available in some rural watersheds without rain gauges. Thus, effective ways of interpolating the available records are needed. Despite many widely used spatial interpolation methods, few studies have investigated rainfall center characteristics. Based on the theory that the spatial distribution of convective rainfall event has a definite center with maximum rainfall, we present a mathematical interpolation method to estimate convective rainfall distribution and indicate the rainfall center location and the center rainfall volume. We apply the method to estimate three convective rainfall events in Santa Catalina Island where reliable hydrological data is available. A cross-validation technique is used to evaluate the method. The result shows that the method will suffer from high relative error in two situations: 1) when estimating the minimum rainfall and 2) when estimating an external site. For all other situations, the method's performance is reasonable and acceptable. Since the method is based on a continuous function, it can provide distributed rainfall data for distributed hydrological model sand indicate statistical characteristics of given areas via mathematical calculation.

Keywords: Rainstorm disaster, Rainfall distribution, Spatial interpolation

1. Introduction

Rainfall is the crucial factor impact on the environment, especially, the flash convective rainfall which can trigger landslides. Rainfall data is widely used in hydrologic design as an important input parameter. Continuous rainfall data in grid format is required to run distributed models for hydrological and agricultural research, including water resource management, flood forecasting, climate change studies, water balance computations, soil moisture modeling for crop production, and irrigation scheduling. The World Meteorological Organization (WMO) conducted a comparative study of various models and indicated that rainfall distribution assumptions and the determination of the form of rainfall were the most important factors in producing accurate estimates of the runoff volume [1]. It has been reported that in order to forecast the stream flow from a basin, quality rainfall inputs are more important than the complexity of the hydrological model [2].

Traditionally, rainfall has been measured by rain gauges that provide information at points scattered over watersheds. Gauge densities typically vary between 1 per 25 km² to 1 gauge for an area over 10,000 km². Unfortunately, rain gauge density is often quite low due to the high cost of monitoring. In recent years, ground-based radar and meteorological satellites have been used to measure rainfall information. Radar data overcome the problem of sparseness in the rain gauge network but are not reliable for the assessment of rain amounts [3]. While meteorological satellites provide quality data for understanding the spatial pattern of rainfall, the estimates of rainfall levels that are produced inherit errors from the nature of the devices and the algorithms used for converting the radiometric measurements to rainfall [4].

It has been highlighted in the literature that rainfall volume and its spatial pattern is the dominant controlling factors of runoff generation and therefore need to be captured properly [5, 6]. Additionally, the locations of the limited rain gauges within the watershed play a crucial role in capturing and understanding how the spatial variability of rainfall influences runoff [7]. Thus, effective ways of interpolating the available records need to be established.

Based on the idea that measured values nearer to the prediction location have more influence on the predicted value than those farther away, spatial interpolation is generally achieved by estimating a regionalized value at ungauged points from a weighted combination of observed regionalized values. The methods currently
The rainfall spatial distribution model is used to define the relationship of rainfall with spatial location. The rainfall distribution pattern always has a rainfall center. The rainfall center has the maximum rainfall volume and the rainfall volume decreases with distance from the center. Accordingly, an exponential function is constructed which is geometrically represented as concentric elliptic curves. The function estimates the rainfall $p_i$ at the location $(x, y)$ as follows:

$$p_i = p_0 e^{-n \sqrt{(x-x_0)^2 + (y-y_0)^2}} (n, m > 0)$$ (1)

where $(x_0, y_0)$ is the location of the rainfall center and $p_0$ is the rainfall volume of the rainfall center. $n$ and $m$ are the coefficients depending on the spatial distribution characteristic of rainfall event. The rainfall $p_i$ has the maximum volume $p_0$, where the point $(x, y)$ equals point $(x_0, y_0)$ according to Eq. (1). The rainfall volume $p$ becomes small and trends to zero when the point $(x, y)$ is far away from the rainfall center $(x_0, y_0)$. The points which have the same rainfall value $p_i$ form an oval on a flat surface. The oval can be formulated with Eq. (2).

$$\left( \frac{\ln(p_i/p_0)}{\sqrt{2nm}} \right)^2 = \frac{(x-x_0)^2}{2} + \frac{(y-y_0)^2}{2m}$$ (2)

The rainfall isolines are thus a group of concentric ovals, and the rainfall center has the maximum rainfall value. These characteristics of Eq. (1) match the observed rainfall distribution characters which are described above. However, the rainfall fields are usually more complex than a group of concentric ovals with a definite center. This situation can be regarded as distribution drift under orographic effects.

### 2.2. Model Modification Via Orographic Effects

Eq. (1) is simply a mathematical interpolation method of estimating rainfall on a flat earth surface, and the rainfall spatial distribution...
only connects with the location \((x, y)\) of the points. The orographic effects are ignored. In fact, the earth surface is not an ideal flat plane. Rainfall can be greatly influenced by the orographic effects, and elevation is an especially important factor in mountainous areas. Elevation variation enhances or retards rainfall, and high altitude is the key factor influencing rainfall. Inclusion of elevation information in the model seems to improve estimation efficiency in a complex orography.

This paper proposes that the orographic effects modify rainfall based on the estimation result of Eq. (1). Considering a given rain gauge \((x_r, y_r, z_r)\), the rainfall difference \(\Delta p\) between the observed value and that calculated via Eq. (1) can be regarded as the result of neglecting the elevation factor. Therefore, \(\Delta p\) is the function of elevation \(z\).

\[
p_r = f(z) = \Delta p
\]

For one rainfall event, every rain gauge has data \((z_r, \Delta p_r)\) such that the rectification equation can be obtained. In this paper, we use a second order polynomial to quantify the rainfall of ungauged sites. A MATLAB computer program was utilized coupled with several rain gauge positions \((x_r, y_r, z_r)\) and the rainfall volume \(p_r\). The coefficients \(n\) and \(m\), the position \((x_0, y_0)\), and rain volume \(p_0\) of the rainfall center can be obtained by an estimation method which minimizes the sum of squared deviations between model estimates and measurements. The orographic effects on rainfall distribution can be modified by the rectification equation \(p_0\). Then, the spatial distribution of rainfall data can be outputted.

### 3. Application

Santa Catalina Island is just off the coast of Los Angeles, located 33.3450°N, 118.3250°W. It covers an area of 8.15 km² (see Fig. 1). There are seven rain gauges located on the island. The rainfall data of those gauges were obtained from the website of the Western Regional Climate Center (http://www.wrcc.dri.edu/CLIMATEDATA.html). This data resource is overseen by the National Oceanic and Atmospheric Administration and is considered reliable. Rain gauge details as well as data for three rainfall events in 2008 are listed in Table 1. A MATLAB procedure was developed to use the data to estimate the rainfall distribution. During the estimation, the data from one rain gauge is selected to check the model estimation accuracy. Therefore, the data of this rain gauge is separated from the data of other rain gauges which were used to obtain the coefficients in the method. In this paper, Cactus Peak was chosen as the object of study, i.e., Cactus Peak rainfall will be estimated by the mathematical interpolation method which coefficients were decided by other rain gauges. The location and rain volume of the rainfall center will also be estimated. The estimation is then modified via MatLab to account for orographic effects. After the modification to account for orographic effects, the final estimation for rainfall at Cactus Peak is shown in Table 2. The difference between estimated rainfall and observed rainfall at Cactus Peak will indicate the accuracy of the model.

#### Fig. 1. Satellite map of Santa Catalina Island.

**Table 1.** The Rain Gauge Information and Three Rainfall Events in 2008

| Station    | Lat. (N)    | Lon. (W)   | Elevation / m | Rainfall Events / mm     |
|------------|-------------|------------|---------------|--------------------------|
|            | 2008/11/2   | 2008/4/2   | 2008/1/27     |
| Cactus Peak| 33° 20' 09"| 118° 26' 10"| 441.96        | 2.794 9.652 20.066       |
| Dakin Peak | 33° 21' 02"| 118° 21' 07"| 480.06        | 1.270 7.874 12.954       |
| Hayfield   | 33° 20' 55"| 118° 25' 28"| 212.75        | 2.540 9.398 18.796       |
| Parsons Landing | 33° 28' 03"| 118° 32' 49"| 122.83        | 0.508 4.318 27.686       |
| Silver Peak Trail | 33° 26' 46"| 118° 32' 35"| 542.54        | 1.778 10.414 29.210      |
| Wild Boar Gully | 33° 19' 28"| 118° 26' 21"| 221.89        | 2.286 7.874 19.304       |
| Whitleys Peak| 33° 22' 21"| 118° 22' 33"| 441.96        | 0.254 6.604 21.590       |

**Table 2.** MATLAB Estimates for Rainfall at Cactus Peak

| Rain Events | Mathematical Model Coefficients | Cactus Peak Rainfall / mm |
|-------------|--------------------------------|---------------------------|
|             | \(p_0\) | \(n\) | \(m\) | \(x_0\) | \(y_0\) | Estimated | Observed | Absolute error |
| 2008/11/2   | 40.14  | 0.3609 | 3.1214 | 33° 21' 45" | 118° 33' 06" | 2.76 | 2.79 | -0.03 |
| 2008/4/2    | 37.27  | 0.1307 | 7.0501 | 33° 22' 47" | 118° 34' 25" | 9.61 | 9.65 | -0.04 |
| 2008/1/27   | 61.89  | 0.1572 | 1.0075 | 33° 27' 22" | 118° 27' 48" | 19.57 | 20.07 | -0.50 |
4. Cross-Validation

We use cross-validation to assess the predictive ability of the method. Cross-validation is a widespread and useful technique for the evaluation of interpolation results [25, 26]. We utilized Leave-one out (LOO) validation for this process. Using the MATLAB computer program, the gauge site rainfall volume which is left out will be estimated. After every gauge site is left out once, we use Mean Absolute Error (MAE), Mean Relative Error (MRE), and Root Mean Square Error (RMSE) as three cross validation statistics to measure errors in different aspects of the method. They are defined as:

\[ \text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |p_i - \hat{p}_i| \]  

(4)

\[ \text{MRE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{p_i - \hat{p}_i}{p_i} \right| \]  

(5)

\[ \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (p_i - \hat{p}_i)^2} \]  

(6)

where \( N \) is the number of gauge sites, and \( p_i \) and \( \hat{p}_i \) are the ith gauge site rainfall estimated and observed volumes, respectively.

The cross validation results are shown in Table 3. MAE illustrates the overall performance of the interpolation method. According to Table 3, MAE will increase with increases of mean rainfall volume. For example, the rainfall on 2008/1/27 is the largest among the three rainfall events and is also associated with the largest MAE.

5. Analysis and Discussion

The interpolation method provided the rain centers \((x_0, y_0)\) and the center rain value \(p_0\) for the three rainfall events as well. The center rainfall volume is calculated via the rainfall distribution trend which is in turn based on the neighboring rain gauge data in the research area. As the method shows, there will be a relatively larger rainfall value in the rainfall center. As distance increases from the center, the value of rainfall decreases. In actual watersheds, there will typically be several rainfall centers which may occur at different time stages in a rainfall event. According to the concentric ovals theory, this complex situation can be regarded as the overlapping of several groups of concentric ovals. This complexity is compounded in large watersheds. When we focus on small scale watersheds, however, just as we divided the large watershed into several small watershed units, each small watershed will be in the state estimated by our method. Besides, the goal of this study is to produce an interpolation method for small watersheds in order to estimate the convective rainfall and provide the rainfall center and center rainfall volume.

In the method, the absolute altitude is considered the key factor which affects the rainfall distribution over the given watershed area. However, Thomas and Herzfeld (2004) believe that rainfall is influenced more by atmospheric factors rather than by absolute altitude alone. The rainfall distribution is a complex hydrological process which is influenced by many of factors, but the effect of these factors changes at different scales. For example, in a large scale watershed, the differences between the physical characters of air masses will impact the volume of rainfall. However, the rainfall volume difference between two points that result from

| Rainfall Events | Gauge Sites       | Absolute Error mm | Relative Error (%) | MAE  | MRE (%) | RMSE |
|----------------|-------------------|-------------------|-------------------|------|---------|------|
| 2008/11/2      | Cactus Peak       | -0.0381           | -1.36             |
|                | Dakin Peak        | -0.1092           | -8.60             |
|                | Hayfield          | -0.1346           | -5.30             |
|                | Parsons Landing   | 0.3785            | 74.50             |
|                | Silver Peak Trail | -0.2692           | -15.14            |
|                | Wild Boar Gully   | 0.1676            | 7.33              |
|                | Whitleys Peak     | 0.2184            | 86.00             |
| 2008/4/2       | Cactus Peak       | -0.0457           | -0.47             |
|                | Dakin Peak        | 0.5055            | 6.42              |
|                | Hayfield          | 0.3302            | 0.35              |
|                | Parsons Landing   | 0.6680            | 15.47             |
|                | Silver Peak Trail | -0.8712           | -8.37             |
|                | Wild Boar Gully   | -0.7239           | -9.19             |
|                | Whitleys Peak     | 1.6637            | 25.19             |
| 2008/1/27      | Cactus Peak       | -0.5904           | -2.49             |
|                | Dakin Peak        | 1.8390            | 14.20             |
|                | Hayfield          | 1.3919            | 7.41              |
|                | Parsons Landing   | -0.5994           | -2.17             |
|                | Silver Peak Trail | -2.1158           | -7.24             |
|                | Wild Boar Gully   | -1.8517           | -9.59             |
|                | Whitleys Peak     | -3.5077           | -16.26            |
air mass physical factors are so negligible at the small scale that they can be ignored. Thus, these effects can be ignored in small watersheds when using observed rainfall gauge data to estimate the rainfall of ungauged locations.

Statistical data from the seven gauge stations shows that the mean rainfall of the 2008/11/2, 2008/4/2, and 2008/1/27 events are 1.6332 mm, 8.0188 mm, and 21.3716 mm, and the MREs of the estimation method are 28.3%, 9.35% and 8.48%, respectively. As MRE considers deviations from the verifying values, the result indicates that the method will produce a large MRE when it is used to estimate light rain. Smaller amounts of mean rainfall will result in larger MREs of the estimation method.

We found that the method will have large relative error when it is used to predict the rain volume of a gauge site with the minimum observed rainfall volume among the gauge stations. The 2008/11/2, 2008/4/2, and 2008/1/27 rainfall events had the smallest rain volumes of 0.254 mm, 4.318 mm, and 12.954 mm at Whitleys Peak, Parsons Landing, and Dakin Peak, and the method obtained the largest or second-largest relative error in these events with 86%, 15.47%, and 14.20%, respectively. We can conclude that this method will result in large estimation deviation when it is used to predict the minimum value and its distribution within one rainfall event.

The situation will worsen when the method is used to estimate the rainfall of a site with the minimum rainfall volume during a light rain event. For example, Whitleys Peak has the smallest observed rainfall volume in the 2008/11/2 light rain event, and relative error of the method reached 86% in this site. This large relative error also contributes greatly to the overall MRE of the 2008/11/2 light rain event.

Parsons Landing is a site that is far to the northwest relative to the center location of the gauge stations, and its elevation is the lowest among the seven gauge stations. The method generally results in large relative error at this site (74% in the 2008/11/2 event and 15.74% in 2008/4/2 event). In other words, the method will result in high relative error due to extrapolation not only in the plane position but also with respect to elevation. Thus, we must exercise caution in this situation.

The method will have good estimation result if we remove these two categories rainfall sites 1) those with the minimum observed rainfall volume, and 2) external sites with regard to the plane position and elevation. The overall MRS will be 8.17% in the case presented. This is acceptable for small watersheds where the local convective rainfall often causes the flash floods.

The method employed in this study can indicate rainfall center, the center rainfall volume, and the rainfall distribution. The IDW and Kriging methods always adopt the point where the rain gauges are located as the rainfall center, i.e., the observed maximum rainfall is assumed to be the maximum rainfall of the watershed (see Fig. 2 and Fig. 3). This does not reflect reality. In the field, there is little chance that the rainfall center happened exactly at the rain gauge location, especially in watersheds without rain gauges. In general, this flaw will lead to an underestimate of the rainfall. Actually, the popular methods including the graphical methods, topographical methods, and numerical methods have the same limitations as IDW and Kriging. Kriging, which has been used all over the world, also takes the observed maximum rainfall as the maximum rainfall. Of course, the deviation of IDW and Kriging will decrease when the rain gauge number increases.

In contrast to the IDW method, this method takes altitude factors into consideration. In the IDW method, rainfall distribution is only determined by observed rainfall values and observed locations \((x, y)\). Moreover, the method is a continuous mathematical function
which can be easily used in mathematical processes, such as integration and differentiation. For example, the total rainfall volume and mean rainfall can be determined by integration of the mathematical function in a given watershed boundary. The method can also output the grid rainfall data when it is coupled with distributed hydrological models.

6. Conclusions

This paper proposes a mathematical rainfall distribution interpolation method for small watershed. Based on the observed rainfall event data, the method can estimate the continuous distribution of the convective rainfall over the watershed. A MatLab procedure is presented to quickly obtain the coefficient values of every rainfall event. Three rainfall events data from Santa Catalina Island was applied. Cross-validation is used to test the methods. The results show the method will obtain high estimation relative error in two situations: 1) when estimating the rainfall of the site with the smallest observed rainfall volume, and 2) when estimating the rainfall of outside the range of model validity, either plane position or elevation. Except for these two situations, the method has good performance and the estimation results are acceptable for ungauged watersheds.

Unlike the current rainfall models, the model can estimate the rain center and the center rainfall volume. The IDW and Kriging methods always assume the observed maximum rainfall location as the rainfall center. In other words, the maximum rainfall value of rainfall events always takes place at the spot where the rain gauge is located, and the locations of rain gauges are always the rainfall centers. These characteristics do not fit the nature of rainfall and will result in underestimates of rainfall. In this point, the interpolation method is reasonable. At the same time, the method is a continuous mathematical function. Thus, it can be used to calculate the mean rainfall and total rainfall volume if required in future research.

Given the goal of a rainfall distribution interpolation method aimed at small watersheds, the model behavior in large watersheds with a low density of rain gauges requires further study.

Acknowledgements

We would like to thank the National Natural Science Foundation of China (Grant no. 41471025, 40971021), the Major Research and Development Program of Shandong Province (Grant no. 2016GGB01398), and the Natural Science Foundation of Shandong Province (Grant no. ZR2014DM004) for supporting this project. The authors are grateful to two anonymous reviewers for their comments that helped to improve the manuscript.

References

1. Manohar A, Singh P, Goel NK, Singh RD. Spatial distribution and seasonal variability of rainfall in a mountainous basin in the Himalayan region. Water Resour. Manag. 2006;20:489-508.
2. Gan TY, Dlamini EM, Biftu GF. Effects of model complexity and structure, data quality and objective functions on hydrologic modeling. J. Hydrol. 1997;192:81-103.
3. Francesca B, Daniela C, Fedele G, Elena S. Spatial reconstruction of rainfall fields from rain gauge and radar data. Stoch. Env. Res. Risk A. 2014;28:1235-1245.
4. Yeboah GA, Geoffrey P. Interpolation of daily rainfall networks...
using simulated radar fields for realistic hydrological modelling of spatial rain field ensembles. *J. Hydrol.* 2014;519:777-791.
5. Shah SMS, O’Connell PE, Hosking JRM. Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *J. Hydrol.* 1996;175:89-111.
6. Segond ML, Wheater HS, Onof C. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *J. Hydrol.* 2007;347:116-131.
7. Bell VA, Moore RJ. The sensitivity of catchment runoff models to rainfall data at different spatial scales. *Hydrol. Earth Syst. Sc.* 2000;4:653-667.
8. Goovaerts P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* 2000;228:113-129.
9. Mair A, Faro A. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *J. Hydrol. Eng.* 2011;16:371-383.
10. Hesbon O, Yang JX, Liu WW, Han DW. Influence of raingage density on interpolation method selection. *J. Hydrol. Eng.* 2014;19:1-8.
11. Buytaert W, Celleri R, Willems P, Bièvre DB, Wyseure G. Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *J. Hydrol.* 2006;329:413-421.
12. Campling P, Gobin A, Feyen J. Temporal and spatial rainfall analysis across a humid tropical catchment. *Hydrol. Process.* 2001;15:359-375.
13. Drogue G, Humbert J, Deraisme J, Mahr N, Freslonc N. A statistical topographic model using an omni-directional parameterization of the relief for mapping orographic rainfall. *Int. J. Climatol.* 2002;22:599-613.
14. Kong YF, Tong WW. Spatial exploration and interpolation of the surface precipitation data. *Geogr. Res.* 2008;27:1097-1108.
15. Gotway CA, Ferguson RB, Hergert GW, Peterson TA. Comparison of Kriging and inverse-distance methods for mapping soil parameters. *Soil Sci. Soc. Am. J.* 1996;60:1237-1247.
16. Wagner PD, Fiener P, Wilken F, Kumar S, Schneider K. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *J. Hydrol.* 2012;464-465:388-400.
17. Webster R, Oliver M. Geostatistics for environmental scientists. New York: Wiley; 2000. p. 45-60.
18. Hrachowitz M, Weiler M. Uncertainty of precipitation estimates caused by sparse gauging networks in a small, mountainous watershed. *J. Hydrol. Eng.* 2011;16:460-471.
19. Bacchi B, Ranzi R, Boga M. Statistical characterization of spatial patterns of rainfall cells in extratropical cyclones. *J. Geophys. Res.* 1996;101:26277-26286.
20. De Launay GJM, Verhoyst NEC, De Troch FP. Characteristics of rainstorms over a temperate region derived from multiple time series of weather radar images. *J. Hydrol.* 2005;307:126-144.
21. Goudenhooft E, Delobbe L. Statistical characteristics of convective storms in Belgium derived from volumetric weather radar observations. *J. Appl. Meteor. Climatol.* 2013;52:918-934.
22. Linsley RK, Franzini JB. Water Resources Engineering. New York: McGraw-Hill; 1979. p. 11-12.
23. Wood ET, Sivalapan M, Beven K, Band L. Effects of spatial variability and scale with implications to hydrologic modeling. *J. Hydrol.* 1988;102:29-47.
24. Abo-Monasar, Al-Zahrani. Estimation of rainfall distribution for the southwestern region of Saudi Arabia. *Hydrolog. Sci. J.* 2014;59:420-431.
25. Hattermann FF, Wattenbach M, Krystofova V, Wechsung F. Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment-validation and uncertainty analysis. *Hydrol. Process.* 2005;19:693-714.
26. Lloyd CD. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J. Hydrol.* 2005;308:128-150.