Uncertainty Analysis of Rainfall Spatial Interpolation in Urban Small Area

Jie Huang[1], Changfeng Jing[2(✉)], Jiayun Fu[2], and Zejun Huang[1]

[1] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

[2] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil, Engineering and Architecture, Beijing 100044, China

jingcf@bucea.edu.cn

Abstract. Uncertainty analysis have attracted increasing attention of both theory and application over the last decades. Owing to the complex of surrounding, uncertainty analysis of rainfall in urban area is very little. Existing literatures on uncertainty analysis paid less attention on gauge density and rainfall intensity. Therefore, this study focuses on urban area, which a good complement to uncertainty research. In this study, gauge density was investigated with carefully selecting of gauge to covering evenly. Rainfall intensity data were extracted from one rainfall event at begin, summit and ending phases of rainfall process. Three traditional methods (Ordinary Kriging, RBF and IDW) and three machine methods (RF, ANN and SVM) were investigated for the uncertainty analysis. The result shows that (1) gauge density has important influence on the interpolation accuracy, and the higher gauge density means the higher accuracy. (2) The uncertainty is progressively stable with the increasing of rainfall intensity. (3) Geostatistic methods has better result than the IDW and RBF owing to considering spatial variability. The selected machine learning methods have good performance than traditional methods. However, the complex training processing and without spatial variability may reduce its practicability in modern flood management. Therefore, the combining of traditional methods and machine learning will be the good paradigm for spatial interpolation and uncertainty analysis.

Keywords: Rainfall · Spatial interpolation · Ordinary Kriging · Random forest · Machine learning

1 Introduction

Rainfall is one of the most important parameters for the flood management, such as hydrological models. Although some weather radars or satellite can get the short timely precipitation data, the ground rain gauge network is still the precise rainfall measure instruments, especially in the urban area. But at most situation it is with a sparse network [1]. Spatial interpolation had been widely used method to estimate the rainfall based on the gauge network, which has been a hot research issue. The common domains using spatial interpolation include meteorology [2], climate [3] and environment [4]. Three substantial roles of interpolation in meteorology domain include

H. Gao et al. (Eds.): TridentCom 2018, LNICST 270, pp. 79–95, 2019.
https://doi.org/10.1007/978-3-030-12971-2_5

parameter for hydrological model [5], area mean rainfall [6], simulating and mapping the rainfall map [7].

Understanding of the uncertainty of interpolation is vital for hydrological model or flood management. But there are only a very few researches on interpolation uncertainties [1]. Uncertainty is the one of challenge in modern flood management because the non-stationarity data come from multiple sources which raises new challenges for uncertainty analysis [8, 9]. Research on uncertainty focus on catchment such as estimating basin precipitation [10], uncertainty analysis for gauge network design within Folsom Lake watershed [11], assessment of precipitation spatiotemporal distribution for hydrological process simulation in the Three Gorges Basin [12]. Considering uncertainty factors, Elevation based topographic influence on spatial interpolation was validated [2]. Rainfall spatial variability is another important factor in interpolation, and its influence is investigated for the numerical simulation model [7].

In general, most of the existing research on uncertainty of rainfall interpolation suffer from several drawbacks. First, there is seldom researches in urban area. The probable causes is the spare gauge network which can bring bias and non-stationarity error [1]. In fact, it is necessary to estimate the uncertainty in urban areas because the precipitation patterns, climatology and surroundings in urban areas are quite different from those catchments [13, 14]. Furthermore, existing research lacks attention on gauge density and rainfall intensity. Existing research argues that rain gauge density is one of factor for the uncertainty [5]. However, owing to the spare gauge network in urban area, it is insufficient number for gauge density analysis [15, 16]. Rainfall intensity affects the measuring error of rain gauge and shows the spatial and temporal distribution of rainfall. Hence, it has also affected the interpolation uncertainty. As a result, the uncertainty of rainfall in urban area and its relationship to gauge density and rainfall intensity are worth of to further investigation.

In this paper, taking the aspects described above into consideration, we investigated the uncertainty with gauge density and rainfall intensity using various spatial interpolation including machine learning methods in urban area. Rain gauge network has 37 gauges, which is enough in number to validate the relationship of uncertainty and gauge density. The network is thoroughly designed based on rules for flood emergency management in dense population city [17]. Although many factors in uncertainty analysis such as rainfall variability, catchment size, topography, and the spatial interpolation technique [5], the topography and rainfall variability were excluded because on the most plain area in study area and data selected from one rainfall event. A 10% interval was adopted to select rain gauge representing different gauge density, meanwhile ensuring the even covering the whole area. Rainfall intensity data was carefully selection at the beginning, summit and ending phases of rainfall event. As for the interpolation methods, deterministic methods and geostatistic methods were both validated. In order to compare the intelligence of interpolation, three machine learning methods were also investigated.

The rest of our paper is organized as follows. Some interpolation methods and uncertainty analysis works are presented in Sect. 2. In Sect. 3, methods and measurements for uncertainty used in this paper are descripted. Following these methods, results are shown in Sect. 4. Validation on uncertainty with gauge density and rainfall intensity are discussed in Sect. 4. We conclude our work in Sect. 5.

2 Study Area and Related Work

2.1 Study Area

A case study in Xicheng District, Beijing, China is selected to investigate our research. It has a total area of approximately 50.7 km2 with about 1,259,000 inhabitants (in 2000 Census). Now, it has totally 37 rain gauges, which means one gauge evenly covering 1.37 km2. Figure 1 shows the study area and rain gauge network.

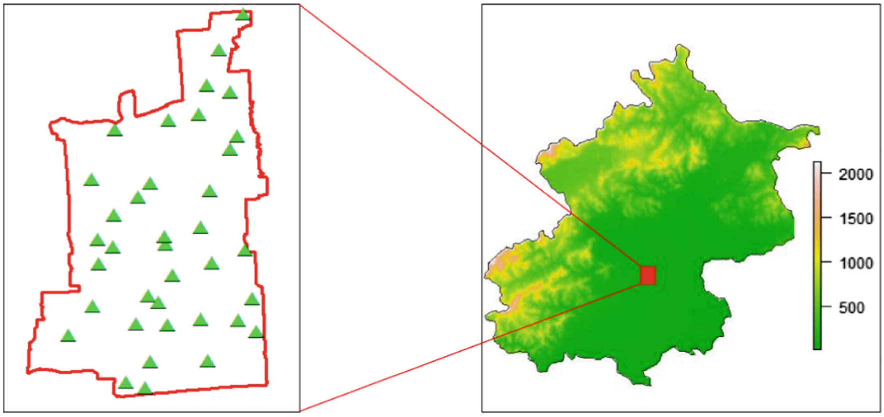

Fig. 1. Study are and rain gauge network.

2.2 Spatial Interpolation Methods

Rain gauge network is generally the first selection for precise measuring rainfall. However, its spare deployment needs various interpolation methods for accurate estimating. One of challenges lies in choosing the right interpolation [16, 18, 19]. The widely used classification of interpolation methods is two categories: deterministic and geostatistic methods [16]. According to complexity of methods, deterministic methods include arithmetic average, Thiessen polygon, Inverse Distance Weight (IDW) and polynomial interpolation [5]. Among them, IDW is the conspicuous method that gives the weight by the inverse of distance. Geostatistic methods such as Ordinary Kriging (OK) and its variants present estimation for un-sample points considering the spatial correlation of sample points [20]. Some literatures argue that geostatistic methods have more performance than deterministic method [21, 22], the other present the interpolation accuracy is case dependent [5, 16, 20]. Although some robust interpolation methods have been developed, such as Gaussian copulas [1], they are rarely used in rainfall interpolation due to their complexity and heavy data requirement.

Machine learning methods, such as random forest (RF) and support vector machine (SVM), have present their capability for accurate estimation at un-sample points. These methods can achieve good estimation even if there are noise data in sample points [23].

SVM has been applied to rainfall data in a previous study [24]. Machine learning methods (RF and SVM) have be investigated for spatial interpolation of environmental variables [25].

Either deterministic method or geostatistic method are found to be more case dependent and no one of these methods can carry perfect interpolation for all rainfall event [20]. Therefore, some combined methods are applied in rainfall prediction. Sometime, the combined method is called the third classification of interpolation methods [21, 25]. The combined schema can strength their advantages and minimized their weakness. For example, regression-kriging (RK) were developed that combines a regression of the dependent variable on auxiliary variables (such as land surface parameters, remote sensing imagery and thematic maps) with simple kriging of the regression residuals [26]. The combination of Random Forest and OK (RFOK) and combination of Random Forest and IDW (RFIDW) were developed and validated on the interpolation of seabed sediments [25].

2.3 Uncertainty Analysis on Rainfall

There are only a very few researches on interpolation uncertainties [1], but it is vital for hydrological model and rainfall estimation. Moulin et al. proposed an error estimated model to interpolate uncertainty of hourly precipitation [18]. Tsintikidis et al. investigated uncertainty analysis for gauge network design [11]. Chen conducted uncertainty assessment of precipitation based on rainfall spatiotemporal distribution for hydrological process simulation [12]. In order to holistic understanding uncertainty in flood management, a framework for uncertainty analysis to support decision making has been established [8]. In view of case study area, catchment and big space had been pay more attention, such as basin precipitation [10], Folsom Lake watershed in US [11], the Three Gorges Basin in China [12], and the upper Loire River in France covering 3234 km2 [18]. On the opposite, there is seldom research in urban area. Elevation and rainfall variability are the widely considered factor for uncertainty analysis. It has been validated that incorporating with elevation can improve the interpolation accuracy [2]. Incorporating the use of spatially-variable precipitation data from a long-range radar in the simulation of the severe flood, spatial variability can influent the total precipitated volumes, water depths and flooded areas [7]. In addition, research argues measurement error from rain gauge network is one of the main sources for uncertainty [5]. Gauge error includes error from device and gauge density in gauge network. In generally, the former had been calibrated in factory. But there is a little research on gauge density. In recent, Otieno conducted similar research in catchment covering 135 km2 with 49 gauges [16].

3 Methodology

3.1 Spatial Interpolation Methods

The general interpolation methods used in this study were Inverse Distance Weighting (IDW), Ordinary Kriging (OK) and Radial Basis Functions (RBF). The main software is the Geostatistical Analyst package of software ArcGIS developed by ESRI Inc. Furthermore, three machine learning methods were also investigated in this study to validate its' accuracy.

Inverse Distance Weighting (IDW). IDW interpolation method is established on the basis of the hypothesis that neighboring point has the more similar properties than the farther one. The principle of IDW methods shows that the estimated value of interpolation points is inversely proportion of the distance from known points [27, 28]. Therefore, it gives greater weights to points closest to the prediction location, and weights diminish as a function of distance. The formula as shown:

$$z(x) = \left[\sum\nolimits_{i=1}^{n} \frac{z_i}{d_i^{\lambda}}\right] \Big/ \left[\sum\nolimits_{i=1}^{n} \frac{1}{d_i^{\lambda}}\right] \tag{1}$$

Where $z(x)$ is the predicted value at an interpolated point, z_i is the i^{th} sample point, n is the total number of sample points, d_i is the distance between the i sample point and the interpolated point, λ is the weighting power which may decide the weight affected by distance.

Ordinary Kriging (OK). Ordinary kriging is one of the most widely used stochastic interpolation methods (Webster and Oliver 2007), which has been engaged for estimating missing rainfall, areal rainfall distribution from point rainfall data, and data fusion of rain gauge and radar data [29, 30]. Kriging is an exact or smooth interpolation method depending on the measurement error model. Ordinary Kriging method is a widely used variant method that is closely relative to autocorrelation. The Kriging estimator is a linear combination of the observed values with weights that are derived from Kriging equations with semi variogram function. The parameters of the semi variogram function and the nugget effect can be estimated by an empirical semi variogram function. The semi variogram function is:

$$y(h) = \frac{1}{2N(h)} \sum\nolimits_{i=1}^{N(h)} \left[z(x_i) - z(x_i + h)\right]^2 \tag{2}$$

Where $y(h)$ is the semi variance value at distance interval h; $N(h)$ is the number of sample pairs within the distance interval h; and $z(x_I + h)$ and $z(x_I)$ are sample values at two points separated by the distance interval h.

Radial Basis Functions (RBF). Radial basis function methods are a series of exact interpolation techniques, that is, the surface must pass through each measured sample point. They are special cases of splines. There are five different basis functions: tine-plate spline, spline with tension, completely regularized spline, multi-quadric function, and inverse multi-quadric function [31]. In this study, we used the inverse multi-quadric function as the radial basis function. The prediction value by RBF can be expressed as the sum of two components [32]. The formula is:

$$z(x) = \sum_{i=1}^{m} a_i f_i(x) + \sum_{i=1}^{n} b_j h_j(d_j) \tag{3}$$

Where $h(d_j)$ shows the radial basis functions and d_j the distance from sample site to prediction point x, $f(x)$ is a trend function, a member of a basis for the space of polynomials of degree is less than m. The coefficients a_i and b_j are calculated by means of the resolution of the linear equations; m and n are the total number of known points used in the interpolation.

Machine Learning Methods. During the past several decades, machine learning methods have been extensively engaged in numerous fields of science and technology. Machine learning methods can train complex data to find a fit model with maximum accuracy and lowest complexity. In this study, three methods were investigated to validate rainfall interpolation accuracy, which include Random Forest (RF), Artificial Neural Network (ANN) and Support Vector Machine (SVM).

Random forest is an ensemble predictor with many tree models. Each tree model depends on the values of a sampled random vector respectively and under the same distribution for all trees in the forest [33]. The RF method can achieve higher predictive accuracy because a set of trees or networks have more robust abilities than one single tree [34], which has been used for data clear applications such as missed data prediction [35]. In addition, RF method has higher efficiency on large dataset with high dimension and easier to use without understanding the data distribution model. In this study, RF adopted the training rainfall data as regressive trees, then forming the forest. Each tree is independent from the others since a random predictor variable is prepared for each node.

The Artificial Neural Network (ANN) is famous on its capability to learn linear predictors from the complex nonlinear data by modeling the target variable using a hidden layer of variable [36]. The general multilayer ANN model is made of three or more neuronal layers: input layer, output layer and one or more intermediate or hidden layer for feature extraction [37]. ANN is a black-box and isn't support the monitoring of model processing. The mean squared error (MSE) is often acted as measure for stopping criterion at each training and validation iteration. If multiple networks are averaged, the approach is comparable to the idea of random forest.

The Support Vector Machine (SVM) is a group of supervised learning method for classification or regression problem based on statistic learning theory [38]. The performance of SVM depends on the kernel function and responding parameters. The radial basis function is one of the popular kernel functions for SVM that had been used for land cover classification [39].

3.2 Accuracy and Uncertainty Evaluation

The performance of rainfall value interpolation is assessed by cross validation with leave-one-out, which has been used in existing rainfall interpolation literatures [16, 30]. According to principles of cross validation with leave-one-out, one sample data of the dataset is temporarily excluded for interpolation and the estimated value of this point is interpolated using the remaining sample points. This step is then repeated until all the points are all "removed – estimated" in turn. The accuracy and uncertainty are evaluated by the measurements of Root Mean Square Error (RMSE) and Coefficient Variation (CV), respectively.

RMSE shows reliable indicator for the spatial interpolation, which has been considered as a preferred evaluation for many applications. RMSE has been widely used in rainfall interpolation and prediction [16, 40]. The equation is shown as follow:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[\hat{z}(s_i) - z(s_i)\right]^2}{n}} \tag{4}$$

Where $\hat{z}(s_i)$ shows the estimated value of point s_i, $z(s_i)$ is the true value of point s_i, n is the total number of points used in the validation.

RMSE is often used to evaluate how far the estimated value are from the true value. RMSE ranges from 0 to infinity, and the smaller of the RMSE means the better estimation of this sample point. RMSE is robust for the evaluation based on the same data, but it would not match the multiple datasets with different variability. In this paper, seven rainfall events with different rainfall intensity were used to evaluate the relationship of rainfall intensity and interpolation accuracy. Considering the variability in intensity, the coefficient of variation (CV) is employed to measure the uncertainty with intensity [5]. The equation is shown as below.

$$CV = \frac{standard\ deviation\ of\ predicted\ rainfall\ value}{predicted\ rainfall\ value} \tag{5}$$

According to the Leave-one-out cross validation, the predicted rainfall value is the mean of values comes from all iterate computations. Therefore, the CV is defined as the standard deviation (SD) divided by the predicted rainfall value. The CV indicates how large differences of estimated rainfall value to its true value tend to be in comparison to their average [41]. The main appeal of the CV is that it takes account on the variability of observe variable by using the mean value, which removes the proportional affection with standard deviation. The CV is therefore a standardization of the SD that allows comparison of variability dataset [42].

4 Result and Discussion

4.1 Gauge Density Experiment

Numerous studies have been done on the comparison of various interpolation methods under different circumstances such as rainfall intensity, spatial and temporal scales. However, rain gauge density is a wide across issue relate to cost, interpolation accuracy and reliability. The influence of rain gauge density on rainfall interpolation is worth of particular interest [16], which need more attention.

In this study, different number of rain gauges were selected representing the different rain gauge density. Then three commonly used interpolation methods (IDW, OK and RBF) were adopted for interpolation, cross validation methods used for accuracy evaluation. Inspiring the sample method for varying density in literature [16], we used 10% interval as the rule to determine the number. It is about 5, 7, 10, 14, 17, 20, 24, 27 and 31 points from all 37 rain gauges, which mean 10.10, 7.21, 5.05, 3.61, 2.97, 2.53, 2.10, 1.87, 1.63 km2 per gauge covering respectively. With the distribution of selected gauges, selection was carefully done ensuring that the selected gauges can still reasonably covered. Figure 2 shows the gauge distribution of partial density.

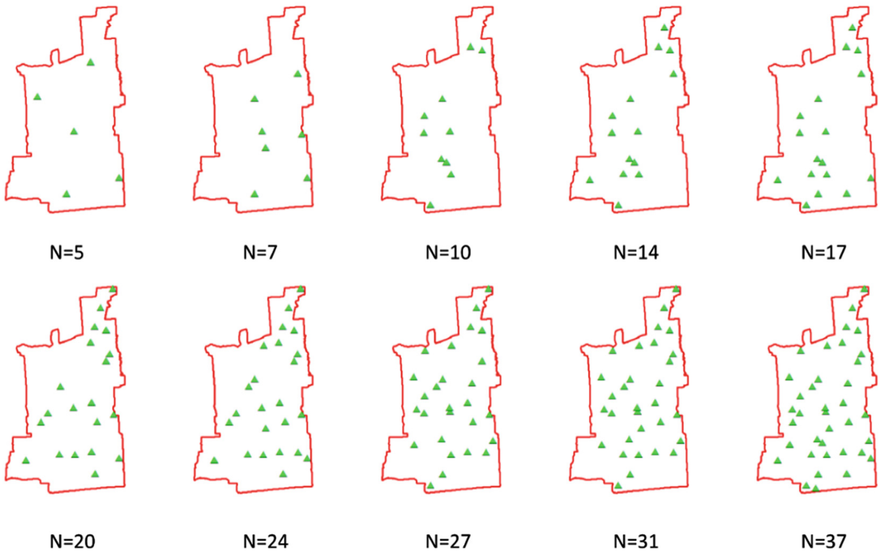

Fig. 2. Rain gauge distribution of various density.

Results were derived from the three interpolation methods, in where, the high quadric surface function is used in RBF methods. Table 1 gives the results. Figure 3 shows the trend graph of the result. Result shows the RMSE is unstable when the sample rate under 50%. Above 50%, it is the overall decreasing trend that is obvious along the increasing of gauge density and number among the three interpolation methods. The RMSE improved about 40%. Therefore, it can conclude that insufficient sample data may cause higher uncertainty and enough measurement is necessary for interpolation. The 50% sample ratio maybe a better threshold for good interpolation accuracy. It can also confirm in Otieno's research work [16]. However, more gauge density samples than Otieno's work (only 3 of 49 points) were tested in this paper. For the high RMSE with less points (such as 5 and 7 points) may has large uncertainty because the unstable error comes from rain gauge measuring error and rainfall variability. Rain gauge data are sensitive to wind and other surrounding environments [43]. The error caused by wind exposure and field is 2–10% for rainfall [44]. Other causes such as water splashing into and out of the collector, evaporation also have influence on measurements. Therefore, enough gauge is necessary for high accuracy interpolation.

Table 1. Rain gauge number and RMSE

Sample rate (rate/number)	RMSE of interpolation methods (mm)		
	IDW	OK	RBF
15/5	5.74	4.04	5.05
20/7	14.67	13.19	15.03
30/10	10.72	8.71	9.65
40/14	10.13	8.92	9.21
50/17	11.16	10.04	10.87
60/20	9.35	8.42	8.72
70/24	9.45	7.57	7.74
80/27	8.45	7.17	7.43
90/31	7.5	6.57	6.92

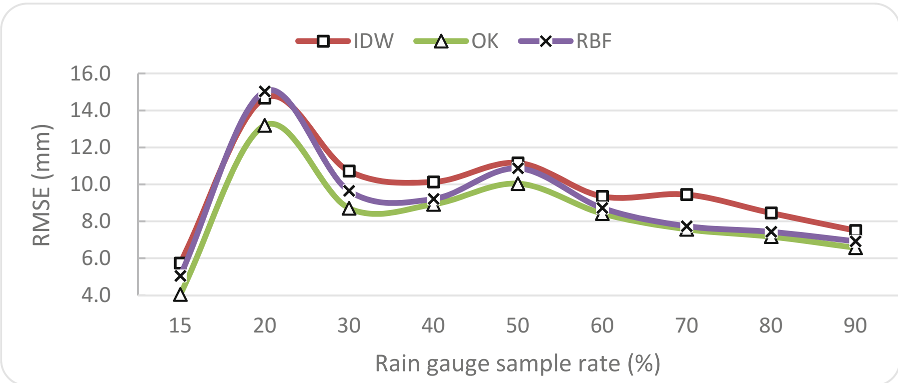

Fig. 3. The RMSE of varying gauge density.

4.2 Interpolation Uncertainty Against Rainfall Intensity

In addition to rainfall gauge network density, rainfall intensity is also of significant interest in urban hydrology as an important parameter for flood model and hydraulic infrastructure planning. Understanding uncertainty level would help for determining these parameters. Hence, the interpolation uncertainty against rainfall intensity is worth of investigation.

In order to reduce the rainfall variability influence, different rainfall intensity data are extracted from the same rainfall event. This study selected seven hourly rainfall data to achieve their rainfall intensity. The instant rainfall of all rain gauges at seven hours are collected to interpolation. The cross validation and CV represent the uncertainty. The description of intensity data is shown in Table 2. And Fig. 3 shows the value distribution of intensity. The intensity of 23rd hour has the maximum magnitude of intensity (Fig. 4).

Table 2. Intensity data of one rainfall event

No.	Time interval[a]	Average hourly intensity (mm/hour)	Instant average rainfall (mm)
1	22:00–23:00	9.41	9.42
2	23:00–24:00	6.37	15.79
3	6:00–7:00	16.82	52.52
4	9:00–10:00	15.08	70.56
5	13:00–14:00	2.63	76.93
6	15:00–16:00	0.77	78.72
7	16:00–17:00	1.17	79.89

[a]Day precipitation is between 20:00 yesterday and 20:00 today.

The scatter plot of CV against rainfall intensity is shown as Fig. 5. With seven intensity datasets and 37 gauges, 259 points are plotted by their intensity. The substantial decreasing trend is observed in this figure, which is similar to Muthusamy's result in literature [5]. In here, a general interpolation uncertainty is discussed, therefore there isn't the classification of rainfall intensity. Meanwhile, something can be drawn: (1) when the intensity <5.0 mm/h, the CVs have bigger deviation with some outliers; (2) when the intensity falling in [5.0, 10.0], the corresponding CVs range between 0.0 and 1.0. And the points are evenly distributed; (3) when the intensity bigger than 1.0 mm/h, the CVs are very stable and close to zero. This pattern is similar to existing research on relationship between intensity and CV [5].

After analyzed the all of 259 sample points, we explored the relationship of uncertainty and the various rainfall intensity at different phases of rainfall event. Three phases with seven datasets are picked up: beginning (23 o'clock and 24 o'clock), summit of rainfall process (7 o'clock and 10 o'clock) and ending phase (14 o'clock, 16 o'clock and 17 o'clock). The result is shown as Fig. 6.

The distribution trend of CV changes from substantial decreasing state (as shown in Fig. 6(a)) to a stable state with clustering distribution (as shown in Fig. 6(e) to (f)).

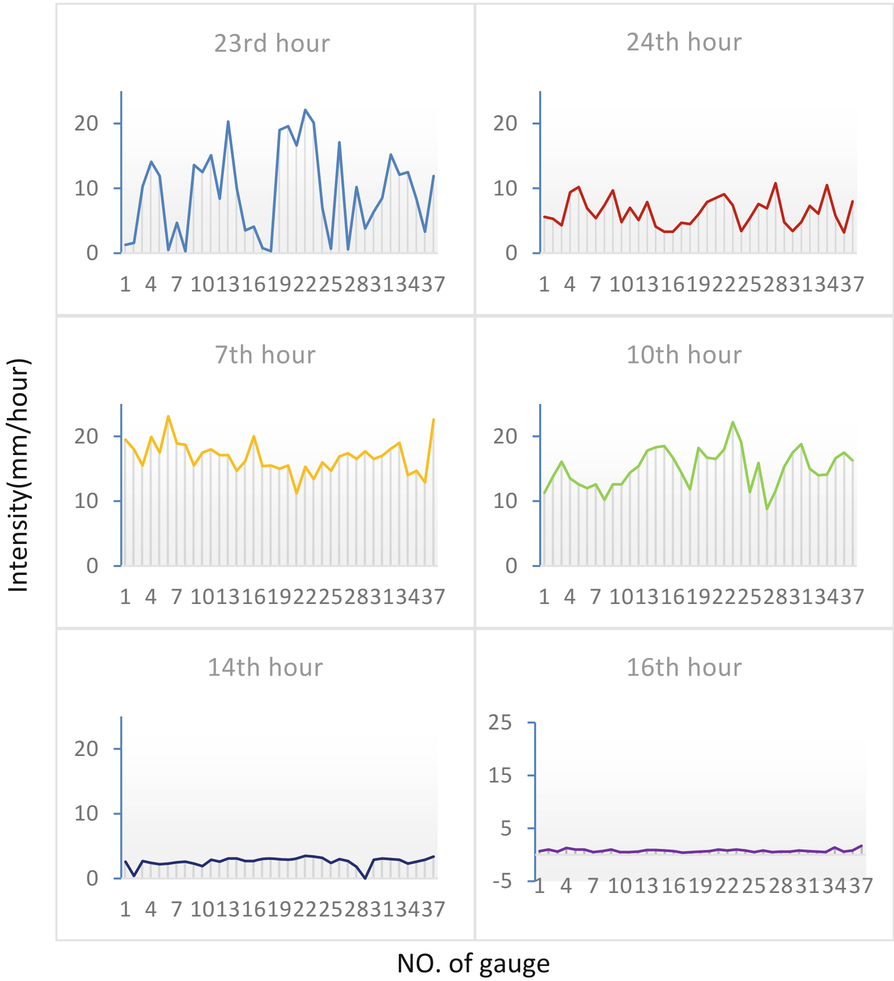

Fig. 4. Intensity value of selected.

In the 23 o’clock, the rain is beginning with large span of intensity (Fig. 6(a)), which is similar to Muthusamy’s result in literature [5]. This decreasing trend is kept in the next hour, but these scatter points show discrete distribution (Fig. 6(b)). In the summit and ending phases of rainfall, the clustering distribution is strengthened. The CV value has changed from a big value to small one, which is from 6.0 of beginning phase (Fig. 6(a)) to 0.2 of ending phase (Fig. 6(e) to (g)).

The different average rainfall value means the bigger gap among CV. At begin, the rainfall value increased from zero and has a lower average rainfall with uneven distribution, then the higher CV maybe occurrence (Fig. 6(a)). On the opposite, at the ending phase, the bigger average rainfall value means the small Standard Deviation of rainfall, which brings a less amplitude of CV (Fig. 6(e) to (g)).

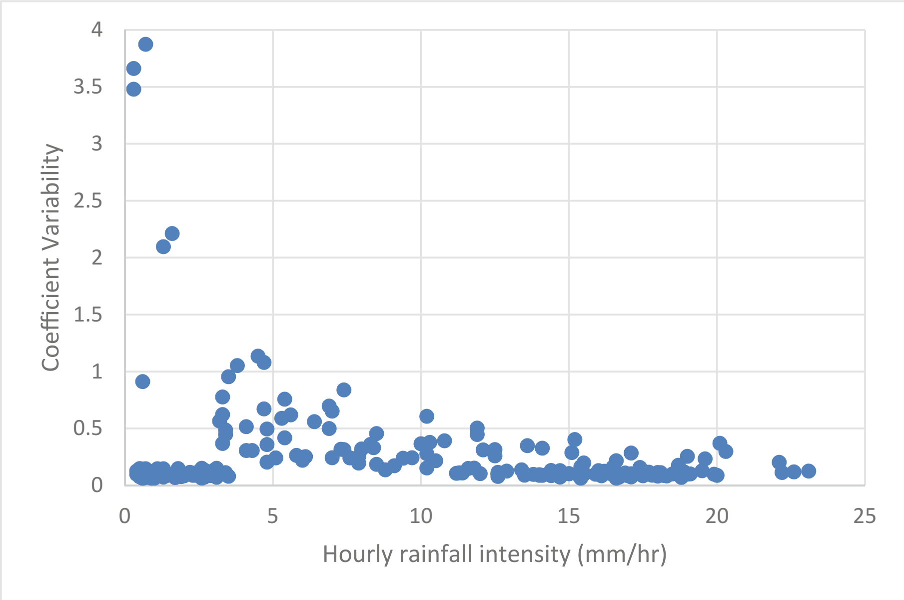

Fig. 5. Interpolation CV against rainfall intensity.

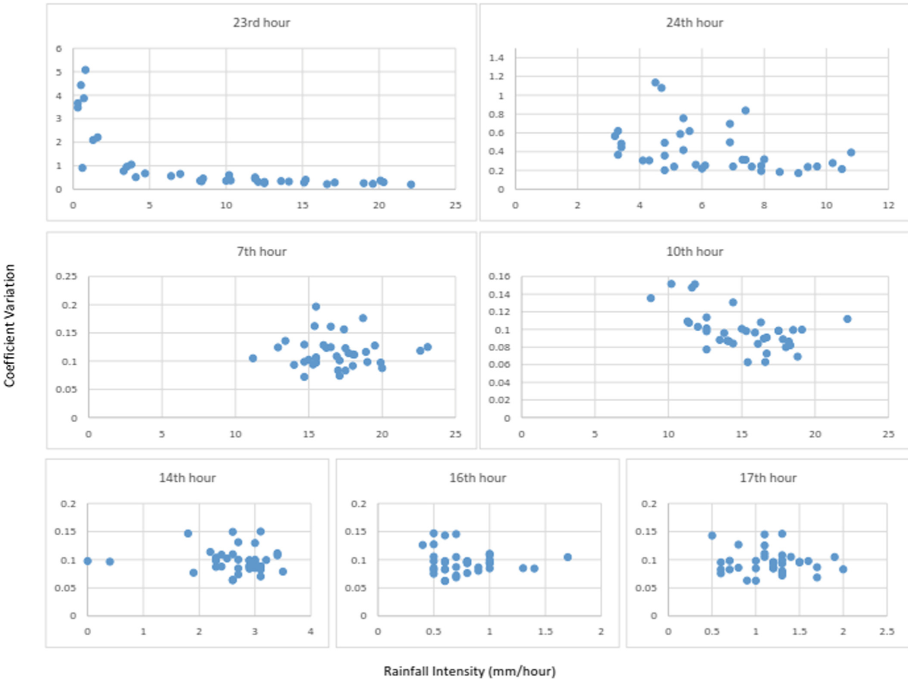

Fig. 6. CV against different phase of rainfall process

4.3 Interpolation Method Comparison

Spatial interpolation is generally estimating or predicting the value at un-sampled points. The mainstream methods can divided into two categories: deterministic and geostatistic methods [16]. The deterministic methods use the similarity and smoothness of surface as measurements for the interpolation to interpolate values. While the geostatistic methods utilize statistic methods and spatial correlation methods with the local variability or global variability theory [45]. Therefore, the geostatistic methods are capable of taking account of the spatial distribution of gauges and spatial variability of data. Therefore, Geostatistic methods has some advantage comparing to deterministic methods in theory. The widely used deterministic methods include IDW, RBF and Thiessen polygon [45]. Meanwhile, the Kriging and other variants such as Simple Kriging and Universal Kriging are the common geostatistic methods. These methods are validated in this paper. In addition, three machine learning methods were validated and compared with the main interpolations in this paper.

Four datasets were selected from one rainfall event at different time according to Table 2. The methods include OK, IDW and RBF with inverse multi-quadric function, RF, ANN and SVM were validated. The RMSE measurement of interpolation result are shown as Table 3.

Table 3. RMSE of different methods

Method	RMSE (mm)			
	23rd hour	7th hour	16th hour	17th hour
OK	**1.92**	**5.09**	**6.81**	**6.99**
IDW	3.35	6.14	7.81	7.97
RBF	2.18	5.32	7.01	7.20
RF	**1.94**	**4.05**	**5.33**	**5.47**
ANN	2.14	4.92	5.73	5.85
SVM	2.05	4.05	5.49	5.52

The results demonstrate that all the interpolation methods can generate different interpolation accuracy because of the existing of uncertainty [6, 46]. According to the RMSE results from Table 3, OK was generally more accurate than IDW and RBF, which is a little different with Otieno's research [16] in which the IDW has better performance than OK method. Therefore, different data under different gauge network may achieve various result. Another probable cause is the input parameters for every interpolation method. Many studies argued that the input parameter of interpolation method is one of the important factors of uncertainty [47]. The validation of optimal power parameter of IDW was conducted by Otieno (2014), then more superior result was achieved. Generally, IDW and RBF need less and simple input parameters comparing to Kriging method. Although the more accurate result can get by carefully calibrating parameters of Kriging method, this is a dynamic parameter for different data.

Meanwhile, the machine learning methods have outstanding performance than traditional methods, which can be concluded from Table 3. The principle of machine learning may be able to explain this result. The training and fitting in machine learning can identify and remove the outlier or noise data for the hypothetic model. Therefore, the higher accuracy may be achieved. However, owing to the spatial variability of rainfall, some noise data in machine learning methods maybe the right value. So, the spatial variability of value should be considered and will be investigated in the future work.

Another criterion for selection interpolation method is the purpose of interpolation [32]. There are two main purposes of rainfall spatial interpolation: assessing the mean area rainfall and mapping the rainfall level. The former focuses on the overall trend of rainfall, and the latter cares of boundary of rainfall level with local max-min rainfall. IDW is very sensitive to weighting power, which is a function of inverse distance. This means that there is a higher influence or weight when closer to the center of the cell being estimated. RBF are based on the degree of smoothing across all sample points. Therefore, IDW methods keep the maximum and minimum value occurring at sample point, but it can generate at un-sample points in RBF method. Kriging method accounts for the spatial autocorrelation overall sample points and keeps the smooth of whole trend. In small area, owing to the weak spatial correlation and strong smoothing effect of kriging, the local maximum was underestimated and the local minimum was overestimated. Therefore, various methods are suitable for different purpose. When considering the overall trend, the Kriging method is more suitable, such as mapping rainfall distribution in study area. In contrast, IDW and RBF have the stronger ability to predict the local maximum and minimum and are suitable for interpolation to generate the rainfall level map.

5 Conclusion

Understanding of the uncertainty of interpolation is vital for flood management, which is the one of challenges in modern flood management. This paper carried out the uncertainty analysis on rainfall in urban area, which is a good complement work with the existing uncertainty analysis on catchment. This paper focus on influence comes from the gauge density and rainfall intensity. We carefully selected the gauge sample points with different number manually to our best to covering the study area. The rainfall intensity data comes from one rainfall event, which weakened the rainfall variability. Three traditional methods and three machine learning methods were selected for rainfall interpolation. We draw conclusions from our work: (1) the accuracy is largely improved with the increasing of gauge density. (2) The uncertainty is progressively stable with the increasing of rainfall intensity. In the hourly intensity, the decreasing of uncertainty with more range of intensity is more obvious than short range of intensity. (3) Interpolation method selection is a crucial thing to achieve accurate estimation according to interpolation purpose and rainfall intensity. Overall, the geostatistic methods has better performance than deterministic methods owing to considering spatial variability. But input parameters of geostatistic methods should carefully calibrate. The machine learning methods have better performance than traditional methods but without considering the spatial variability. In the future work, more combined methods with machine learning will be investigated, meanwhile, the spatial variability must be examined.

Acknowledgments. The authors would like to thank the valuable comments from anonymous reviewers. This study is jointly supported by the National Natural Science Foundation of China (Grant No. 41771412), the Beijing Natural Science Foundation (Grant No. 8182015), Beijing Advanced innovation center for future urban design (Grant No. X18052, X18058, X18158) and the Zhejiang Province Research Program (Grant No. 2015C33064).

References

1. Bárdossy, A., Pegram, G.: Interpolation of precipitation under topographic influence at different time scales. Water Resour. Res. **49**(8), 4545–4565 (2013)
2. Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. **228**(1–2), 113–129 (2000)
3. Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Model Softw. **16**(4), 309–330 (2001)
4. Li, J., Heap, A.D.: Spatial interpolation methods applied in the environmental sciences: a review. Environ. Model Softw. **53**, 173–189 (2014)
5. Muthusamy, M., Schellart, A., Tait, S., Heuvelink, G.B.M.: Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models. Hydrol. Earth Syst. Sci. **21**(2), 1077–1091 (2017)
6. Wagner, P.D., Fiener, P., Wilken, F., Kumar, S., Schneider, K.: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. J. Hydrol. **464–465**, 388–400 (2012)
7. Courty, L., Rico-Ramirez, M., Pedrozo-Acuña, A.: The significance of the spatial variability of rainfall on the numerical simulation of urban floods. Water **10**(2), 207 (2018)
8. Hall, J., Solomatine, D.: A framework for uncertainty analysis in flood risk management decisions. Int. J. River Basin Manag. **6**(2), 85–98 (2008)
9. Hutter, G., Schanze, J.: Learning how to deal with uncertainty of flood risk in long-term planning. Int. J. River Basin Manag. **6**(2), 175–184 (2008)
10. Hrachowitz, M., Weiler, M.: Uncertainty of precipitation estimates caused by sparse gauging networks in a small. Mountainous Watershed. J. Hydrol. Eng. **16**(5), 460–471 (2011)
11. Tsintikidis, D., Georgakakos, K.R., Sperfslage, J.A., Smith, D.E., Carpenter, T.M.: Precipitation uncertainty and raingauge network design within Folsom Lake watershed. J. Hydrol. Eng. **7**(2), 175–184 (2002)
12. Cheng, M., et al.: Performance assessment of spatial interpolation of precipitation for hydrological process simulation in the Three Gorges Basin. Water **9**(11), 838 (2017)
13. Rupa, C., Mujumdar, P.P.: Quantification of uncertainty in spatial return levels of urban precipitation extremes. J. Hydrol. Eng. **23**(1), 04017053(2018)
14. Yang, L., Tian, F., Niyogi, D.: A need to revisit hydrologic responses to urbanization by incorporating the feedback on spatial rainfall patterns. Urban Clim. **12**, 128–140 (2015)
15. Liu, M., Bárdossy, A., Zehe, E.: Interaction of valleys and circulation patterns (CPs) on spatial precipitation patterns in southern Germany. Hydrol. Earth Syst. Sci. **17**(11), 4685–4699 (2013)

16. Otieno, H., Yang, J., Liu, W., Han, D.: Influence of rain gauge density on interpolation method selection. J. Hydrol. Eng. **19**(11), 04014024(2014)
17. Jing, C., Yu, J., Dai, P., Wei, H., Du, M.: Rule-based rain gauge network design in urban areas aided by spatial kernel density. Water Pract. Technol. **11**(1), 166–175 (2016)
18. Moulin, L., Gaume, E., Obled, C.: Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrol. Earth Syst. Sci. **13**(2), 99–114 (2009)
19. Kobold, M., Sušelj, K.: Precipitation forecasts and their uncertainty as input into hydrological models. Hydrol. Earth Syst. Sci. **9**(4), 322–332 (2005)
20. Ly, S., Charles, C., Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: a review. Base **17**(2), 392–406 (2013)
21. Li, J., Heap, A.D.: A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecol. Inform. **6**(3–4), 228–241 (2011)
22. de Amorim Borges, P., Franke, J., da Anunciação, Y.M.T., Weiss, H., Bernhofer, C.: Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. Theor. Appl. Climatol. **123**(1–2), 335–348 (2016)
23. Appelhans, T., Mwangomo, E., Hardy, D.R., Hemp, A., Nauss, T.: Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spat. Stat. **14**, 91–113 (2015)
24. Gilardi, S., Begio, N.: Local machine learning models for spatial data analysis. J. Geogr. Inf. Decis. Anal. **4**(EPFL-ARTICLE-82651), 11–28 (2000)
25. Li, J., Heap, A.D., Potter, A., Daniell, J.J.: Application of machine learning methods to spatial interpolation of environmental variables. Environ. Model Softw. **26**(12), 1647–1659 (2011)
26. Hengl, T., Heuvelink, G.B.M., Rossiter, D.G.: About regression-Kriging: from equations to case studies. Comput. Geosci. **33**(10), 1301–1315 (2007)
27. Bhargava, N., Bhargava, R., Tanwar, P.S., Narooka, P.C.: Comparative study of inverse power of IDW interpolation method in inherent error analysis of aspect variable. In: Mishra, D., Nayak, M., Joshi, A. (eds.) Information and Communication Technology for Sustainable Development, pp. 521–529. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-3920-1_52
28. Maciej, T.: Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW) - cross-validation/Jackknife approach. J. Geogr. Inf. Decis. Anal. **2**(2), 18–30 (1998)
29. Adhikary, S.K., Muttil, N., Yilmaz, A.G.: Genetic programming-based Ordinary Kriging for spatial interpolation of rainfall. J. Hydrol. Eng. **21**(2), 1–14 (2016)
30. Berndt, C., Rabiei, E., Haberlandt, U.: Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. J. Hydrol. **508**, 88–101 (2014)
31. ESRI: How radial basis functions work (2013)
32. Xie, Y., et al.: Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. Chemosphere **82**(3), 468–476 (2011)
33. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
34. Genuer, R., Poggi, J.M., Tuleau-Malot, C., Villa-Vialaneix, N.: Random forests for big data. Big Data Res. **9**, 28–46 (2017)
35. Kühnlein, M., Appelhans, T., Thies, B., Nauss, T.: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning - a random forests-based approach applied to MSG SEVIRI. Remote Sens. Environ. **141**, 129–143 (2014)

36. Basheer, I.A., Hajmeer, M.: Artificial neural networks: fundamentals, computing, design, and application. J. Microbiol. Methods **43**(1), 3–31 (2000)
37. Prasad, R., Deo, R.C., Li, Y., Maraseni, T.: Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos. Res. **197**, 42–63 (2017)
38. Cortes, C., Cortes, C., Vapnik, V., Vapnik, V.: Support vector networks. Mach. Learn. **20**(3), 273–297 (1995)
39. Kavzoglu, T., Colkesen, I.: A kernel functions analysis for support vector machines for land cover classification. Int. J. Appl. Earth Obs. Geoinf. **11**(5), 352–359 (2009)
40. Sadler, J.M., Goodall, J.L., Morsy, M.M.: Effect of rain gauge proximity on rainfall estimation for problematic urban coastal watersheds in Virginia Beach, Virginia. J. Hydrol. Eng. **22**(9), 04017036(2017)
41. Cox, J.C., Sadiraj, V.: On the coefficient of variation as a measure of risk sensitivity. SSRN **3** (3), (2011)
42. Reed, G.F., Lynn, F., Meade, B.D.: Quantitative assays. Clin. Diagn. Lab. Immunol. **9**(6), 1235–1239 (2002)
43. Cristiano, E., Veldhuis, M.-C., Van De Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas-a review. Hydrol. Earth Syst. Sci. **21**, 3859–3878 (2017)
44. WMO: Guide to Meteorological Instruments and Methods of observation (WMO-No.8), Seven edit. Geneva, Switzerland (2008)
45. Goovaerts, P.: Geostatistics in soil science: state-of-the-art and perspectives. Geoderma **89** (1–2), 1–45 (1999)
46. Ma, L., Chi, X., Zuo, C.: Evaluation of interpolation models for rainfall erosivity on a large scale. In: First International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 1–5. IEEE, Shanghai (2012)
47. Zhang, P., Liu, R., Bao, Y., Wang, J., Yu, W., Shen, Z.: Uncertainty of SWAT model at different DEM resolutions in a large mountainous watershed. Water Res. **53**, 132–144 (2014)