Inverse Distance Weighting and Kriging Spatial Interpolation for Data Center Thermal Monitoring

Esi Oktavia, Widyawan, and I Wayan Mustika
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Jalan Grafika No. 2, Yogyakarta, 55281 Indonesia
Email: esi.oktavia@mail.ugm.ac.id, {widyawan, wmustika}@ugm.ac.id

***Abstract*—Studies have shown that data center performance is also influenced by its environmental conditions, one of them is thermal state. If thermal information inside a data center, such as temperature and humidity are not well-monitored, then the data center might experience overheat or overcool state, resulting in downtimes or other performance issues. However, for more accurate thermal monitoring, it is better to collect temperature readings from many sensors, which is not really feasible in reality. Spatial interpolation methods have been adopted as a means to predict spatial information at certain locations without adding sensors continuously. However, most of the previous works utilize spatial interpolation concept inside an environment with medium to large scale sensor nodes. In this study, we first customize a model of real server room with 5 temperature and humidity sensors in order for it to be fitted in spatial interpolation concepts. We then apply and evaluate two most commonly used spatial interpolators, i.e. Inverse Distance Weighting (IDW) and Kriging with regard to our customized server room model. Our results from 30 measurements followed by significance test demonstrate that IDW gives higher accuracy than Kriging when it is implemented inside an environment involving small-scale sensors.**

***Index Terms*—Interpolation, data center, Inverse Distance Weighting (IDW), Kriging.**

I. Introduction

Data center can be defined as a facility used to accommodate computer systems and associated components, such as telecommunications and storage system [1]. It holds a substantial role in enterprise where multiple servers host applications and store critical data. Hence, great data center performance has been believed as something that contribute to better service delivery in an organization or enterprise.

In a 2004 white paper from Cisco Systems [2], it can be derived that there are four objectives in the design of any high performance data center, i.e. security, availability, scalability, and manageability. Power and thermal management have also become an important issue in data center performance due to their relation with physical security aspect of a data center. According to research, more than 60% of the data centers and server rooms experience 1-4 downtimes a year [3] because of changes in environmental conditions, one of them is rise in temperature. The previous statement substantiates the claim that temperature has been a constant concern as one of the most significant parameters in data center monitoring.

In order to monitor data center temperature accurately, it is better to collect temperature information from many devices or sensors. However, by considering the fact that the number of available sensors is limited in practical, adding sensors continuously to the data center environment is not a feasible solution. How to deploy limited number of temperature sensors to accurately monitor data center temperature is another big challenge. It will be necessary to apply interpolation in order to predict temperature readings in between the sensors.

Although the implementation of spatial interpolators other than in geological field, such as in Wireless Sensor Network or WSN-related works has been studied in the past [4]–[7], researches implemented inside small-scale wireless sensor nodes and further analyzed by a means of statistical tests were rarely found. In this paper, we attempt to model spatial interpolation concept, generally implemented in geological field, into data center environment. Our data center model is developed based on the existing condition of a real server room involving only 5 temperature and humidity sensors. Two most commonly used spatial interpolation methods, i.e. IDW and Kriging, are applied to our data center model. The performance evaluation of those two interpolation methods is derived in terms of their performance accuracy and significance.

Specifically, the contributions of this paper are two-fold.

- While most of the previous researches related to spatial interpolation methods are conducted in geological field, two interpolation methods are applied based on a real server room situation involving small-scale temperature sensors. We simplify the server room model in order for it to be fitted in the spatial interpolation concepts and calculations.
- IDW and Kriging spatial interpolation methods are evaluated based on their accuracy in predicting temperature reading at certain locations. Statistical tests are also utilized to obtain whether the two methods have significant difference in terms of accuracy. The results of our research demonstrate a comparable accuracy between the two aforementioned interpolation methods.

The remainder of this paper is organized as follows. Section II highlights the distinction of our research by discussing several related works. Section III presents a brief description of IDW and Kriging interpolation, which are utilized in this research. Our data center model is depicted in Section IV, followed by Section V that elaborates on experimental method in conducting our work. In Section VI, we present results from the interpolated calculations and statistical tests, followed by Section VII that concludes the paper.

II. Related Work

We have found that most researches related to spatial interpolation methods were conducted in geographical or geological science field. However, this section highlights several researches regarding spatial interpolators which were found in WSN-related works.

Despite the basis of spatial interpolation concepts, which is derived from geological science field, several WSN-related researches have also utilized those concepts. Previous researches indicate that Kriging interpolation is one of the most well-known and frequently used method. Reference [8] have applied Kriging interpolation method to measure the performance of two different sampling methods, i.e. grid and gradient-based method. For such a case involving the measurement of indoor air distributions, they recommend gradient-based sampling method alongside with Kriging interpolation for interpolating the measured data to form field distributions. It can be concluded that this research focuses on the performance of compared sampling methods, not the interpolation method itself.

The fact that Kriging is one of the most frequently used method triggers the continuous improvements of it. Those improvements result in different types of Kriging interpolation method. A novel Distributed Kriging (DISK) [4] have been proposed and compared with global Kriging with respect to their accuracy and cost. DISK method offers Quad Suppress (QS) algorithm in its variogram calculation in order to reduce communication costs.

However, several researches have also discussed the performance comparison of some well-known interpolation methods other than Kriging, in terms of WSN deployment. Jedermann et al. [5], [6] have analyzed the performance of Null-model, IDW and several types of Kriging interpolation method in terms of accuracy. They conducted their research in a customized setup environment which represents the condition during food transportation and storage. Those studies conclude that the accuracy of Kriging interpolation method almost overpower the rest of compared methods. Moreover, they have proved that it is possible to run interpolation methods on wireless sensor nodes, such as iMote2 equipped with and ARM processor.

The concept of spatial interpolation can also be combined with temporal interpolation, as in [7]. Such combination is utilized not only for estimating data at missing points, but also for preventing redundant data from being injected to the wireless sensor networks. Their proposed model, called ODAST, results in effective reduction in energy consumption due to its capability for reducing data traffic in an efficient manner.

From several works related to WSN interpolation discussed above, it is shown that most of them were not conducted inside an environment involving small-scale sensor nodes. Furthermore, none of the aforementioned works have utilized statistical tests to compare the performance of various interpolators. Our research focuses on evaluating two most commonly used spatial interpolation methods with regard to a real server room model involving only 5 temperature and humidity sensors. We also refine our analysis by a means of significance test to ensure whether both methods differ significantly in terms of accuracy.

III. Spatial Interpolation Method

Issue of incomplete data is one of the biggest challenges that might be encountered when we are dealing with spatial modeling. Therefore, interpolation should be applied as a means to overcome this issue. The concept of interpolation itself is actually derived from geological sciences. It is defined as a process for predicting value at a point which is not the sample point, based on the values of sample points in its neighborhood [9].

Basically, one interpolation method can be distinguished from others based on how it defines weight of any sample points. Spatial interpolation methods can also be classified into 2 types, i.e. deterministic and stochastic. Deterministic interpolators use mathematical formulas in order to calculate the value of unsampled location, while stochastic methods take into account the statistical information and spatial arrangement of values at sampled points. This section highlights the basic concept of two most commonly used methods from each aforementioned types of spatial interpolators, which are Inverse Distance Weighting (IDW) and Kriging.

A. *Inverse Distance Weighting (IDW)*

IDW is a deterministic interpolation method which utilizes the concept that values at unsampled points are determined by a linear combination of values at known sampled points. By considering distance as weighting parameter, IDW assumes that values closer to the unsampled location are more representative of the value to be estimated than values from samples further away [10]. That assumption concludes that nearby observations will have a heavier weight.

In order to calculate the distance between sampled point and interpolation point d_i, IDW method makes use of Euclidean distance formula, which can be presented as

$$d_i = \sqrt{(x - x_i)^2 + (y_- y_i)^2} \tag{1}$$

where (x,y) and (x_i,y_i) represent the location of the sampled point and interpolation point in (x,y) or Cartesian coordinate, respectively.

IDW interpolators are of the form:

$$f(x, y) = \frac{\sum_{i=1}^{N} d_i^{-d_{exp}} P_i}{\sum_{i=1}^{N} d_i^{-d_{exp}}} \tag{2}$$

From Eq. 2, N is the number of sampled points, P_i is the measured values at the sampled points (x_i,y_i), d_i is the

Euclidean distance between the referred sample point and interpolation point (x,y), while d_{exp} is the distance exponent or power index. The value of d_{exp} is generally 2 in order to ease the calculation and also to give good empirical results for purposes of general surface mapping and description [11], [12].

B. Kriging

Kriging is one of the most common interpolators used in geological field. It is classified into stochastic interpolation method where the weight for each sampled points is defined by means of both distance and spatial arrangement among those points.

There are various kinds of Kriging method, for example Simple Kriging, Ordinary Kriging, and Universal Kriging. Due to its characteristic which employs a statistical model, several assumptions must be met in order to apply Kriging method. Different types of Kriging interpolator may have different assumptions. However, this research utilizes the concept of Ordinary Kriging, which assumes that there is an unknown constant mean that must be estimated based on the data [13].

To predict the value at an unsampled point using Kriging method, we have to conduct a series of steps [14] as follows.

1) Calculate empirical semivariogam: The empirical semivariogram is used to determine spatial relationship between sampled points. As in IDW method, we first compute the Euclidean distance and squared difference between each pair of sampled points. Then, the empirical semivariogram γ can be calculated as

$$\gamma = \frac{1}{2} \times [(P_i - P_j)^2] \tag{3}$$

where P_i and P_j are the values (in this case, temperature values) at two sampled locations.

2) Fit a model: This step is conducted by first transforming the distance versus empirical semivariance on to a scatter plot - the empirical semivariogram. We then fit a certain model to that scatter plot in order to determine semivariogram values for various distances.

There are several models that could be used to fit the empirical semivariogram, such as linear, spherical, exponential, and Gaussian model [13]. However, this research utilizes linear model due to its simplicity and ease for manual calculation. It is also known as least-squares regression line.

The general form of linear regression line can be represented as

$$\hat{y} = a + bx \tag{4}$$

where $\hat{y}$ is the estimated value of dependent variable, a is the value of $\hat{y}$ when x = 0, x is the value of independent variable, while b is the slope of the fitted model which can be computed as

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \tag{5}$$

where n is the number of observations, x is the value of independent variable, and y is the value of dependent variable. The value of this slope is used to create the matrices in the next step.

Having computed the value of b, we can calculate the value of a as

$$a = \bar{y} - b\bar{x} \tag{6}$$

where $\bar{x}$ and $\bar{y}$ are the mean of x and y, respectively.

3) Create the matrices: In order to determine the weights that are assigned to each sampled points, Ordinary Kriging makes use of matrices and vectors derived from the spatial autocorrelation between sampled points and unsampled or interpolation points. The elements of those matrices and vectors can be calculated as

$$\tilde{C}_{ij} = b \times d_{ij} \tag{7}$$

$$\tilde{C}_i = b \times d_i \tag{8}$$

where b represents the slope of the fitted model, d_{ij} is the distance between sampled points, and d_i is the distance between sampled point and interpolation point. The covariance values from Eq. 7 and Eq. 8 are then used to construct Γ and g, which can be presented as follows.

$$\Gamma = \begin{pmatrix} \tilde{C}_{11} & \tilde{C}_{12} & \cdots & \tilde{C}_{1N} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \tilde{C}_{N1} & \tilde{C}_{N2} & \cdots & \tilde{C}_{NN} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \tag{9}$$

$$g = \begin{pmatrix} \tilde{C}_i \\ \vdots \\ \tilde{C}_N \\ 1 \end{pmatrix} \tag{10}$$

Then, the weights for each sampled points can be computed as

$$\lambda = \Gamma^{-1} g \tag{11}$$

where Γ^{-1} is the inverse of matrices Γ defined in Eq. 9.

4) Make a prediction: From Kriging weights computed with Eq. 11, we can predict the value at various unsampled locations. The interpolated value can be determined by multiplying the weight for each sampled points by the measured value at those points, then add the products together, as described in Eq. 12.

$$f(x, y) = \sum_{i=1}^{N} \lambda_i P_i \tag{12}$$

IV. Modeling the Data Center

As mentioned in Section I, this research aims to evaluate the performance of IDW and Kriging interpolators inside a real data center environment. The data center model refers to a 10.8 meters × 9.6 meters server room located at Directorate of System and Information Resource, Universitas Gadjah Mada, which is divided into 3 compartments or separate areas. Five temperature and humidity sensors, as depicted in Fig. 1, are located inside the server room. However, in this research we

Figure 1. Temperature and humidity sensor

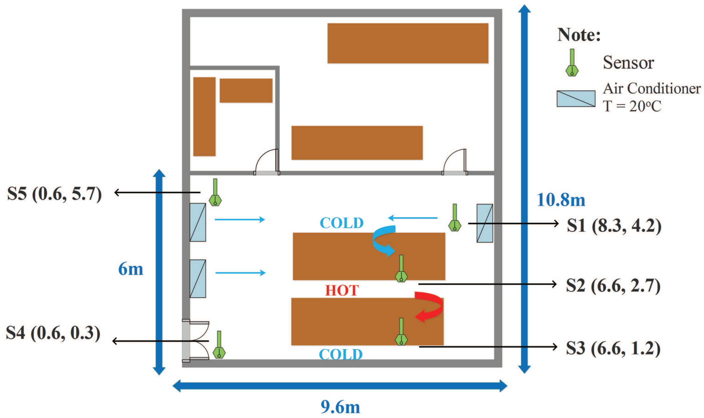

Figure 2. 2D model - top view of the server room

only focus on the temperature readings from those sensors. Fig. 2 shows an overview of the 2D data center model which is used to conduct this research. In order to make the data center model to be fitted in spatial interpolation concept, as to make it easier for calculating interpolated temperatures, several assumptions were made, which are as follows.

- We only consider the largest compartment in the data center, which is 6 meters $\times$ 9.6 meters in size.
- The data center model is simplified into 2D model with (x,y) or Cartesian coordinates representing the location of each sensors, as shown in Fig. 2.

V. Experimental Method

In this section, we introduce the steps conducted in this research. We define our experimental method into three main tasks, including data extraction, calculation of interpolated temperature, and performance evaluation of the two spatial interpolators.

A. Data Extraction

We first record the temperature readings from 5 sensors located on the server room every hour, twice a day. This step is conducted for 3 days. Therefore, at the end of this step we have 30 temperature readings in total. We then utilize those data as an input for predicting temperature at certain unsampled location.

B. Calculation of Interpolated Temperature

Having extracted the temperature data, we calculate the interpolated temperature value at an unsampled point based on the temperature readings at 4 sampled points. Due to the differences between IDW and Kriging calculation, we attempt to define several parameters as follows.

- For IDW interpolator, we define the value of power index equal to 2, as suggested in Section II.
- For Kriging interpolator, we apply the concept of linear regression (least-square method) into fitted semivariogram model due to its simplicity and ease for manual calculation.

C. Performance Evaluation

The accuracy of the two methods can be measured by using the concept of root-mean-square error or RMSE which can be described as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_r - T_i)^2} \quad (13)$$

where T_r represents the real temperature value derived from sensors, while T_i is the predicted temperature value resulted from spatial interpolation.

In order to statistically compare the performance between IDW and Kriging, we then attempt to conduct significance test for determining whether the two interpolation methods have significant difference in their accuracy. The initial step that should be done in conducting significance test is normality test, which is used to find out the data distribution, whether it is normal or non-normal. In this research, we choose Shapiro-Wilk test which is recommended for normality test involving small sample sizes < 50. If the data are distributed normally (parametric), then we can continue our significance test by using t-test. The accuracy of both compared methods can be stated to have significant difference if $p\text{−}value < \alpha$. However, if the data have non-normal distribution (non-parametric), then it is recommended to run Mann-Whitney U Test, where the value of $U_{critical}$ must be less than $U_{calculated}$ in order for both methods to have a significant difference in terms of accuracy.

VI. Results and Discussions

In this section, we show the experimental results from both interpolation methods based on their accuracy for predicting temperature value at certain location. We divide this section into several subsections as described below.

A. IDW and Kriging Results

From all five sensors located on the server room, one is treated as an unsampled point, which its temperature is

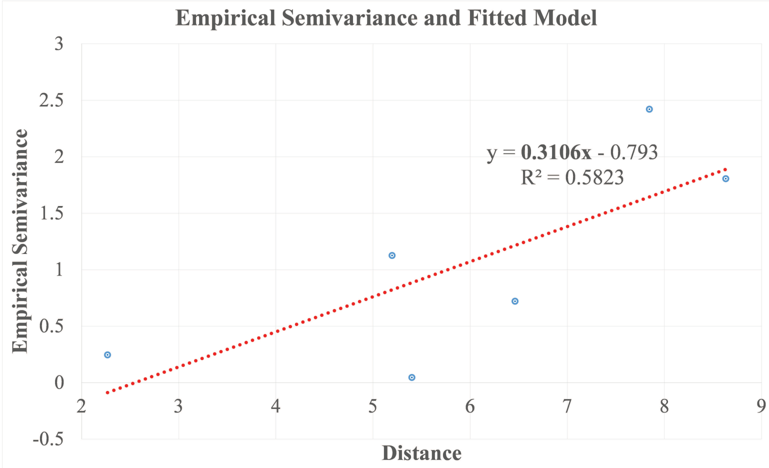

Figure 3. Empirical semivariance and fitted model of Kriging method

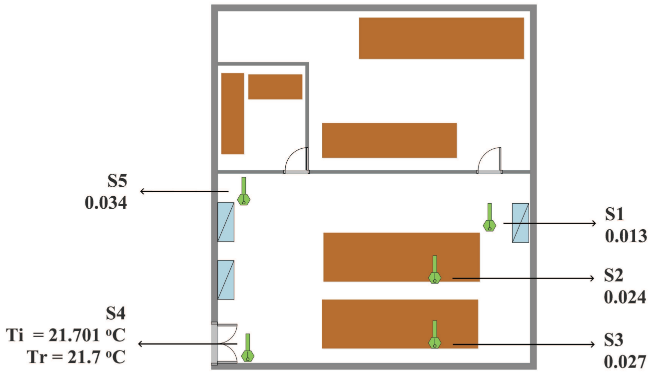

Figure 4. IDW weights

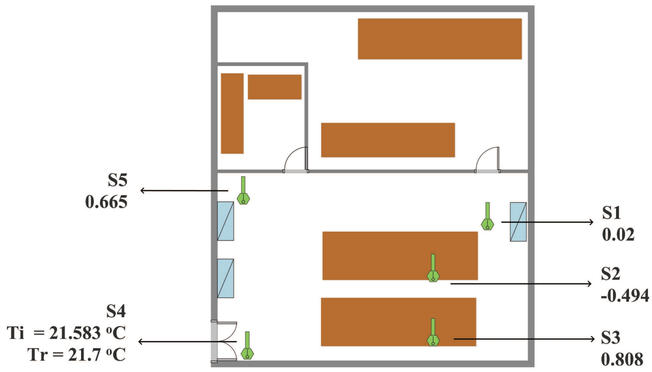

Figure 5. Kriging weights

predicted by considering the location and temperature value of the remaining four sensors. We then compare the results of interpolated temperature value with the real value shown by the available sensors.

As discussed in the beginning of Section 2, various interpolation methods can be distinguished by the way they define weights for each sampled points. For IDW interpolator, we simply compute those weights by a means of inverse distance formulas described in Eq. 2, while in Kriging we fit a simple linear regression line into the empirical semivariance to determine weights at various locations, as depicted in Fig. 3. The weights resulted from both interpolators demonstrate the concept that the weights decrease as the distances increase. However, they still can be differentiated due to Kriging characteristic which takes the spatial arrangement of the data into account. The comparison between IDW and Kriging weights is illustrated in Fig. 4 and Fig. 5.

Having calculated the interpolated temperature values at all five locations, we first evaluate the performance of both interpolators by comparing their RMSE. In our works involving 30 measurements for each interpolators, we find that the RMSE of IDW is 0.5512, while the RMSE of Kriging equals to 0.7424. Therefore, our result concludes that IDW spatial interpolation method gives better accuracy theoretically due to their smaller value of RMSE compared to Kriging.

B. Normality Test

We conduct normality test to the RMSE values using Shapiro-Wilk method to determine the data distribution. The results of this test are presented in Table I. In order to derive a conclusion about whether the data are distributed normally, we compare the Sig. value for both IDW and Kriging. We use 95 percents of confidence interval for this test, so it can be stated that the data have normal distribution if the value of $Sig.$ for each method is less than α (or $Sig. > 0.05$). From the results presented in Table I, it is shown that the data from both interpolators have normal distribution. Therefore, we continue our statistical test by using t-test.

C. Significance Test

t-test is conducted in this step due to normal distribution shown by the dataset. This step aims to demonstrate whether the accuracy of both interpolators shows a significant difference or not. In this step, two initial hypotheses must be defined, which are as follows.

Table I
RESULTS OF SHAPIRO-WILK TEST

Method	Statistic	df	$Sig.$
IDW (MRSE)	0.915	5	0.500
Kriging (MRSE)	0.975	5	0.907

Table II
RESULTS OF T-TEST

Assumption	$Sig.$	Mean difference	Std. error difference
Equal variances	0.017	-0.24382	0.081083
Unequal variances	0.020	-0.24382	0.081083

- H_0: $\mu_1 = \mu_2$ (Mean accuracy for both methods are the same)
- H_1: $\mu_1 \neq \mu_2$ (Mean accuracy for both methods are different)

With 95 percents of confidence interval (α = 0.05), the null hypothesis (H_0) will be rejected if the value of $Sig.$ is less than α ($Sig. <$ 0.05). The results of this significance test, as shown in Table II present that the value of $Sig.$ is 0.017, which is less than 0.05. Therefore, H_0 is rejected. In case of our study, it can be concluded that, IDW gives higher accuracy than Kriging interpolators, both theoretically and statistically.

VII. CONCLUSION

Thermal monitoring is one of critical issues in data center performance due to its relation with physical aspects of data center and other network resources, which requires many sensors to obtain thermal data accurately. As the number of sensors is limited in practical, spatial interpolation has been identified as an ideal solution to predict thermal data, such as temperature without adding sensors continuously. However, most previous works related to performance evaluation of spatial interpolators were not conducted inside an environment involving small-scale sensors. In this paper, we have presented a performance evaluation of two most commonly used spatial interpolators, i.e. IDW and Kriging, based on a real server room model consisting of 5 temperature sensors. We also apply significance test to determine whether the mean accuracy of the two methods differ statistically. Different from the results of several related works which stated that the accuracy of Kriging interpolators outperforms that of IDW, our study demonstrates that, theoretically and statistically, IDW gives higher accuracy than Kriging when it is implemented inside an environment involving small-scale sensors.

REFERENCES

[1] N. M. S. Hassan, M. M. K. Khan, and M. G. Rasul, Temperature monitoring and CFD analysis of data centre, Procedia Eng., vol. 56, pp. 551559, 2013.

[2] Cisco Systems, Data Centers : Best Practices for Security and Performance, 2004. [Online]. Available: http://www.echomountain.com/pdfs/CiscoBestPractices.pdf

[3] Avtech, Protect Your IT Facility, 2009. [Online]. Available: http://www.avtech.com/About/Articles/AVT/NA/25/-/AA-ND-DD-TN/Protecting_IT_Equipment_09.htm

[4] M. Umer, L. Kulik, and E. Tanin, Kriging for localized spatial interpolation in sensor networks, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5069 LNCS, pp. 525532, 2008.

[5] R. Jedermann, J. Palafox-Albarrn, P. Barreiro, L. Ruiz-Garca, J. Ignacio Robla, and W. Lang, Interpolation of spatial temperature profiles by sensor networks, Proc. IEEE Sensors, vol. 10001, no. 01, pp. 778781, 2011.

[6] R. Jedermann, J. P.- Albarrn, P. Barreiro, L. Ruiz-garca, and J. Ignacio, Modelling and Interpolation of Spatial Temperature During Food Transportation and Storage by the Variogram, unpublished.

[7] A. Kanzaki, Y. Iima, T. Hara, and S. Nishio, Overhearing-based data transmission reduction using data interpolation in wireless sensor networks, Proc. Intl. Conf. Mob. Comput. Ubiquitous Netw. (ICMU 2010), pp. 16, 2010.

[8] Y. Huang, X. Shen, J. Li, B. Li, R. Duan, C. H. Lin, J. Liu, and Q. Chen, A method to optimize sampling locations for measuring indoor air distributions, Atmos. Environ., vol. 102, pp. 355365, 2015.

[9] P. A Burrough and R. A McDonnell, Principles of Geographical Information Systems, Econ. Geogr., vol. 75, no. 4, p. 422, 1998.

[10] A. D. Hartkamp, K. De Beurs, A. Stein, and J. W. White, Interpolation Techniques for Climate Variables, 1999.

[11] D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, 23rd ACM Natl. Conf., pp. 517524, 1968.

[12] Z. Xu and J. Guan, A Distributed Inverse Distance Weighted Interpolation Algorithm Based on the Cloud Computing Platform of Hadoop and Its Implementation, pp. 24122416, 2015.

[13] S. Eberly, J. Swall, D. Holland, B. Cox, and E. Baldridge, Developing Spatially Interpolated Surfaces and Estimating Uncertainty, Quality, p. 169, 2004.

[14] K. Johnston, J. M. Ver Hoef, K. Krivoruchko, and N. Lucas, The principles of geostatistical analysis, Using ArcGIS geostatistical Anal., pp. 4980, 2003.