International Journal of Parallel Programming (2020) 48:534–548
https://doi.org/10.1007/s10766-018-0595-5

Comparison of Estimating Missing Values in IoT Time Series Data Using Different Interpolation Algorithms

Zengyu Ding[1] · Gang Mei[1] · Salvatore Cuomo[2] · Yixuan Li[1] · Nengxiong Xu[1]

Received: 31 May 2018 / Accepted: 10 August 2018 / Published online: 17 August 2018

Abstract
When collecting the Internet of Things data using various sensors or other devices, it may be possible to miss several kinds of values of interest. In this paper, we focus on estimating the missing values in IoT time series data using three interpolation algorithms, including (1) Radial Basis Functions, (2) Moving Least Squares (MLS), and (3) Adaptive Inverse Distance Weighted. To evaluate the performance of estimating missing values, we estimate the missing values in eight selected sets of IoT time series data, and compare with those imputed by the standard kNN estimator. Our experiments indicate that in most experiments the estimation based on the Lancaster's MLS is the best. It is also found that the number of nearest observed values for reference and the distribution of missing values could strongly affect the accuracy of imputation.

Keywords Internet of Things (IoT) · Time series · Missing values · Interpolation · Data imputation

1 Introduction

Nowadays, Internet of Things (IoT) [5,18,26] is one of the promising technologies that have attracted lots of attention in both industrial and academic fields [11]. IoT aims to integrate seamlessly both physical and digital worlds in one single ecosystem that makes up a new intelligent era of Internet [15], which has been widely applied in the fields of smart agriculture [28], smart cities [1,21], smart home [2,9,24], and personalised healthcare [17]. In short, IoT can be considered as the extension and expansion of the Internet, which has and will affect many fields in our daily life both on personal and business levels.

✉ Gang Mei
gang.mei@cugb.edu.cn

1 China University of Geosciences (Beijing), 100083 Beijing, China

2 Department of Mathematics and Applications "R. Caccioppoli", University of Naples Federico II, Naples, Italy

In the IoT, data represents the bridge that connects digital and physical worlds. Data is quite critical in the IoT due to its utility from the need of ways to represent and manipulate the huge amount of raw data expected to be generated from the "things". A major characteristic of the IoT data is its large amount. This is due to the fact that: technological advances have impressively enhanced the "data harvesting" capabilities of embedded sensor devices resulting in more generated data and more continuous data streams from the real world.

When analyzing the IoT data, a critical issue needed to be carefully dealt with is the data quality. Data quality is crucial to gain user engagement and acceptance of the IoT paradigm and services. If data are of poor quality, decisions are likely to be unreasonable [10]. In general, there are two categories of values in the IoT data that can strongly reduce the quality of data gained in the IoT, i.e., the abnormal values and the missing values. The existence of the abnormal values and missing values is mainly due to incorrect response or nonresponse.

To improve the data quality reduced by missing values, many approaches have been proposed for dealing with the missing values in the IoT data. These approaches can be broadly classified into statistical and machine learning techniques [27]. The statistical techniques for estimating missing values are usually based on mean/mode and regression, which have been widely used for a long time. In the recent years, several powerful machine learning techniques occur and have been frequently used to impute missing values in various Big Data. For example, the k Nearest Neighbor (kNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) techniques have been employed in missing data imputation [16,22,25].

When dealing with the missing values in IoT data, two critical issues need to be carefully considered, i.e., the estimation accuracy and computational efficiency. Typically, more complex algorithms for imputing missing values might be able to produce better imputation results but will generally require a higher computational cost. Most machine learning techniques are usually more computationally expensive than many statistical techniques due to the (off-line) model training and construction process they entail. However, one exception is that the kNN technique is the most computationally efficient in most cases because it is the *Lazy* Learning [23].

The standard kNN estimator [8] for imputing missing values is in fact a straightforward interpolant. It can be considered as a specific version of the Shepard's method [19] since it has the same principle as that of the basic form of the Shepard's method. Motivated by estimating missing values using the Shepard's interpolant / kNN estimator, we are quite interested in assessing the performance (i.e., the estimation accuracy and computational efficiency) of imputing missing values using several of other interpolation algorithms.

There have been several research work focusing on imputing missing values using various interpolation algorithms. For example, Beveridge [3] extended the minimum mean square error linear interpolator to handle missing values in time series for any pattern of nonconsecutive observations. Bhattacharjee et al. [4] proposed semantic Kriging to blend the semantics of spatial features of surrounding data points with ordinary Kriging (OK) method for prediction of the missing attribute in Geographic Information System(GIS). Shtiliyanova et al. [20] explored the properties of Kriging to impute missing data in air temperature series.

In this paper, we focus on estimating the missing values in IoT data using various interpolation algorithms. Specifically, we employ the interpolants based on (1) Radial Basis Functions (RBF), (2) Moving Least Squares (MLS), and (3) Adaptive Inverse Distance Weighted (AIDW) to predict the missing values occurring in the IoT time-series data. To the best of the authors' knowledge, there is currently no such research work.

The rest of this paper is organized as follows. Section 2 introduces the method for estimating the missing values in IoT data using several different interpolation algorithms. Section 3 presents the benchmark experiments and Sect. 4 discusses the experimental results. Finally, Sect. 5 draws several conclusions.

2 Material and Methods

In this paper, our objective is to compare the estimation of missing values in IoT time series data when using the MLS, RBF, and AIDW interpolation algorithms. In this section, we will first briefly introduce the principle of the above mentioned three interpolation algorithms, and then describe the process of comparing the estimation of missing values in IoT time series data.

2.1 Brief Introduction to the MLS, RBF, and AIDW Interpolation

2.1.1 The Orthogonal MLS Interpolation Algorithm

For a given polynomial basis function $p_i(x)$, there is an orthonormal basis function $q_i(x, \bar{x}), i = 1, 2, \ldots, m$ (m is the number of the basis functions), which satisfies:

$$\begin{aligned} q_1(x, \bar{x}) &= p_1(x), \\ q_i(x, \bar{x}) &= p_i(x) - \sum_{j=1}^{i-1} \alpha_{ij}(x, \bar{x}) q_j(x, \bar{x}), \end{aligned} \tag{1}$$

where

$$\alpha_{ij}(\bar{x}) = \frac{\sum_{k=1}^{n} w_k(\bar{x})\, p_i(x_k)\, q_j(x_k, \bar{x})}{\sum_{k=1}^{n} w_k(\bar{x})\, q_j^2(x_k, \bar{x})}. \tag{2}$$

Because the coefficient matrix is a diagonal matrix, the solving for $a_i(x)$ does not require the inversion of matrix, i.e.,

$$a_i(\bar{x}) = \frac{\sum_{k=1}^{n} w_k(\bar{x})\, q_i(x_k, \bar{x})\, f_k}{\sum_{k=1}^{n} w_k(\bar{x})\, q_i^2(x_k, \bar{x})}, \tag{3}$$

where n is the number of data points.

When the number or order of basis functions increases, it only needs to calculate a_{m+1} and α_{m+1} in Schmidt's orthogonalization. It is not needed to recalculate all entries

in the coefficient matrix. This could reduce the computational cost and probably also reduce the computational error.

2.1.2 The Lancaster's MLS Interpolation Algorithm

To make the approximation function $f^h(x)$ constructed by the interpolation type moving least square method satisfy the properties of the Kronecker δ function, a singular weight function is adopted:

$$\omega(x, x_k) = \begin{cases} \|(x - x_k)/\rho_k\|^{-\alpha}, & \|x - x_k\| \le \rho_k \\ 0, & \|x - x_k\| > \rho_k \end{cases}. \tag{4}$$

Let $p_0(x) \equiv 1, p_1(x), \ldots, p_{\bar{m}}(x)$ denote the basis functions used to construct the approximation function, where the number of basis functions is $\bar{m} + 1$. To be able to implement interpolation properties, a new set of basis functions is constructed for a given basis function. First, $p_0(x)$ is standardized, i.e.,

$$\tilde{p}_0(x, \bar{x}) = \frac{1}{\left[\sum_{k=1}^{n} \omega(x, x_k)\right]^{1/2}}, \tag{5}$$

Moreover, we construct a new basis function of the following form:

$$\tilde{p}_i(x, \bar{x}) = p_i(\bar{x}) - \sum_{k=1}^{n} \frac{\omega(x, x_k)}{\sum_{l=1}^{n} \omega(x, x_l)} P_i(x_k). \tag{6}$$

2.1.3 The RBF Interpolation Algorithm

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of pair wise distinct points in a domain $\Omega \subseteq R^d$ with associated data values $f_i, i = 1, 2, \ldots, n$. We consider the problem of constructing a d-variety function $F \in C^k\left(R^d\right)$ that interpolates the known data. Namely, we require $F(x_i) = f_i, i = 1, 2, \ldots, n$. If we denote F in the form

$$F(x) = \sum_{j=1}^{n} w_j \varphi\left(\left\|x_i - x_j\right\|_2\right), \tag{7}$$

where $\varphi : [0, \infty] \to R$ is a suitable continuous function, then the interpolation conditions become:

$$\sum_{j=1}^{n} w_j \varphi\left(\left\|x_i - x_j\right\|_2\right) = f_i, i = 1, 2, \ldots, n. \tag{8}$$

2.1.4 The AIDW Interpolation Algorithm

The AIDW is an improved version of the standard IDW [19], which is originated by Lu and Wong [12]. The distance-decay parameter α is no longer a pre-specified constant value but adaptively adjusted for a specific unknown interpolated point according to the distribution of the nearest neighboring data points.

2.2 Comparison of Estimating Missing Values Using Different Interpolation Algorithms

2.2.1 The Baseline Estimator for Comparison

We employ the standard kNN method [8] as the baseline to examine the estimation of missing values. In the kNN method, the estimates of the missing observations are calculated as weighted averages of the k nearest neighboring observations:

$$\hat{y}_j = \frac{\sum_{i=1}^{k} w_{ij} y_i}{\sum_{i=1}^{k} w_{ij}}, \tag{9}$$

where, k is a number of nearest neighboring observations, y_i is the observed value of dependent variable, $\hat{y}_j$ is the respective prediction for missing observation j, and w_{ij} is the weight of a neighboring observation i for the missing observation j.

The weight is calculated as follows:

$$w_{ij} = \frac{\left(\frac{1}{d_{ij}}\right)^{pm}}{\sum_{i=1}^{k} \left(\frac{1}{d_{ij}}\right)^{pm}}, \tag{10}$$

where d_{ij} is the similarity distance between i and j, and pm is the weighting parameter ($i \neq j$).

The distance d_{ij} is defined as

$$d_{ij} = \sum_{l=1}^{L} c_l \left| x_{il} - x_{jl} \right|, \tag{11}$$

where L is the number of independent variables, and c is their respective weights. In this work, we aim at estimating a single independent variable in each imputation. Thus, L is 1, and c is ignored in the imputation.

2.2.2 Measures for Evaluating the Estimated Missing Values

The estimation accuracy and computational efficiency are critical in any methods for imputing missing values. In this work, we will compare both the accuracy and

Table 1 Employed IoT time series data in our experiments

Dataset	City	Selected attribute	Number of all observations	Number of missing observations	Missing rate (%)
City air quality in China	Beijing	PM2.5	52560	2174	4
	Shanghai	PM2.5	30816	2062	7
	Chengdu	PM2.5	35136	1103	3
	Shenyang	PM2.5	23592	1913	8
City air quality in Italy	Milan	CO	9351	1683	18
	Milan	C_6H_6	9351	366	4
	Milan	NO_X	9351	1639	18
	Milan	NO_2	9351	1642	18

efficiency of the MLS, RBF, AIDW, and *k*NN estimators in the estimation of missing values in several time series datasets.

We use two measures, i.e., the Root Mean Square Error (RMSE) and the Standard Deviation (SD), to evaluate the estimation of missing values in IoT time series data. The RMSE is used to measure the accuracy of the estimated missing values by comparing to the observed values. The SD is used to quantify the amount of variation or dispersion of (1) each dataset without the missing values and (2) each dataset with the estimated missing values.

We also record the imputation time to compare the efficiency. Note that when the size of the employed dataset is not large, the computational time of imputation cannot be counted apparently. Thus, we will repeat the same imputation several times and then obtain the average computational time.

3 Results

3.1 Employed IoT Time Series Data

In our experiments, we use two datasets of city air quality obtained from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/index.php). Note that there are multiattributes in each dataset, but only several of those attributes are selected for the estimation of missing values. Basic information of those employed datasets is listed in Table 1.

3.2 Accuracy of Estimating Missing Values in IoT Time Series Data

For each set of selected variables, we sequentially select 90% observed values as the training / known observations, and assume the rest 10% values as the testing / missing observations; see Fig. 1. It should be noted that in fact the 10% missing values have the

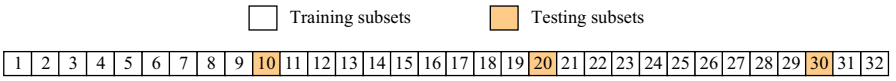

Fig. 1 The sequentially selected training subset (90%) and the testing subset (10%)

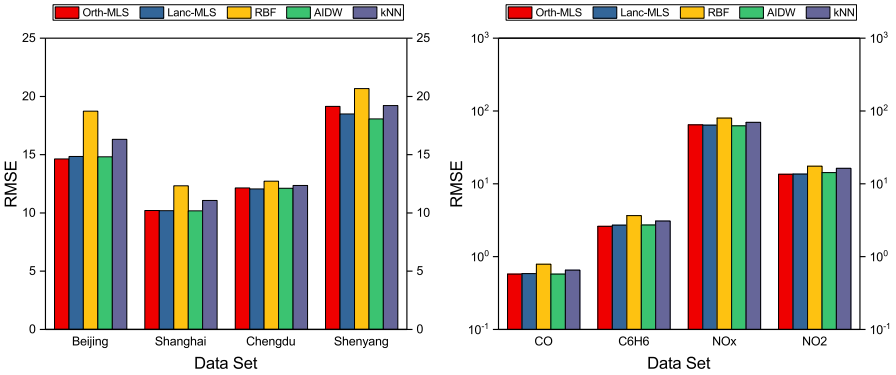

Fig. 2 Comparison of the RMSE when $k = 10$. **a** The concentration of PM 2.5 in four Chinese cities. **b** The concentration of four gases in Milan

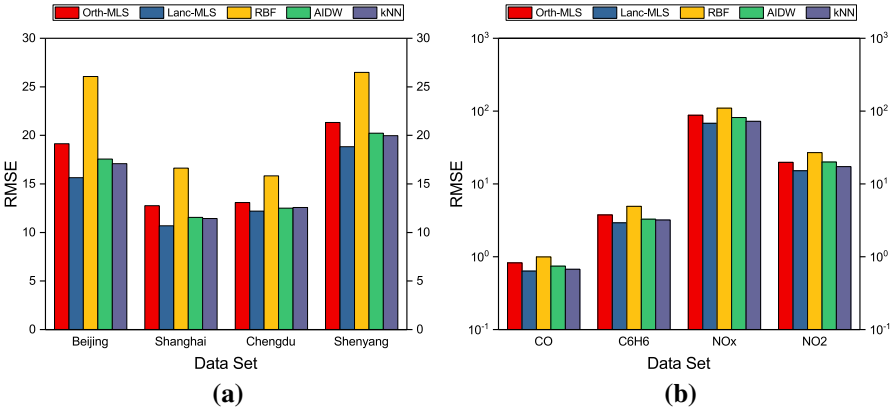

Fig. 3 Comparison of the RMSE when $k = 20$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

really observed values, after imputing, the estimated values of the assumed missing observations will be compared to the predicted values.

To compare the performance of the interpolation algorithms to that of the kNN estimator, we also select the k nearest neighboring observations to impute those missing observations. That is, all the estimators are on the basis of local rather than global reference observations. We have configured the k as 10, 20, 40, 80, and 160, and then obtained the imputation accuracy. The imputation accuracy is measured with the RMSE; see Figs. 2, 3, 4, 5 and 6.

The results illustrated in Figs. 2, 3, 4, 5 and 6 show that:

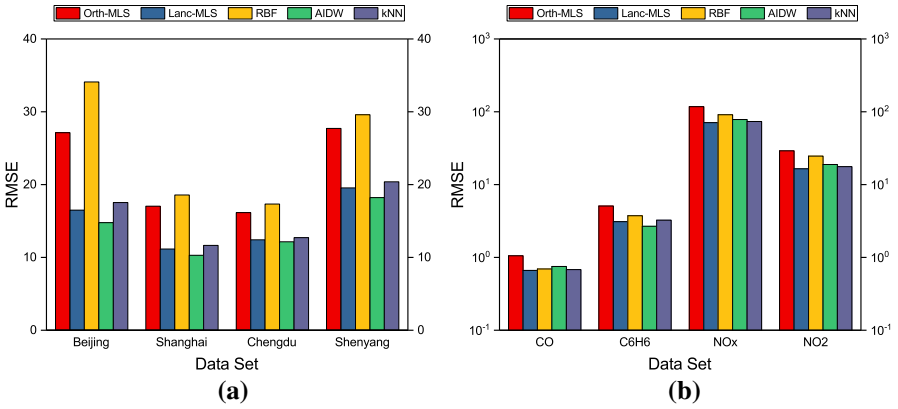

Fig. 4 Comparison of the RMSE when $k = 40$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

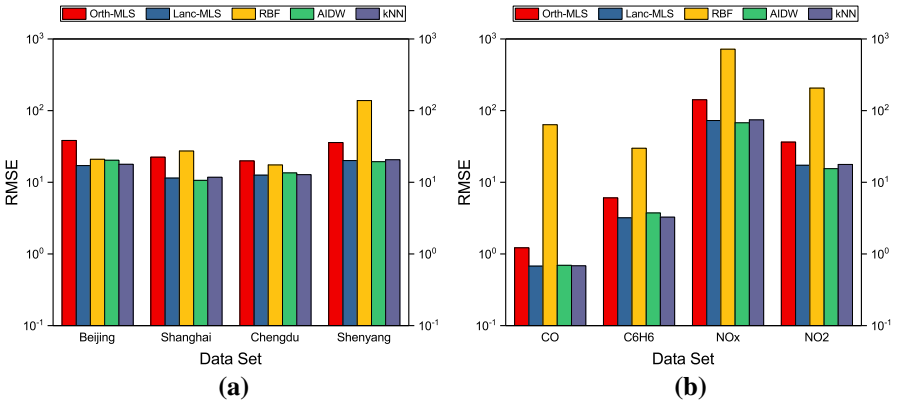

Fig. 5 Comparison of the RMSE when $k = 80$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

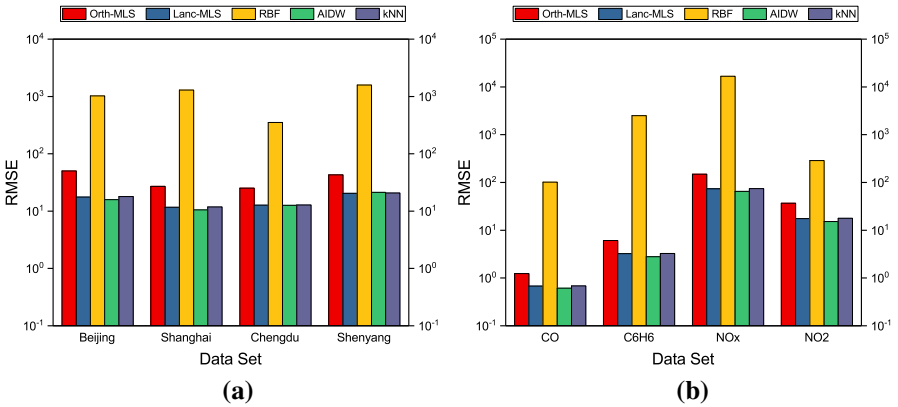

Fig. 6 Comparison of the RMSE when $k = 160$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

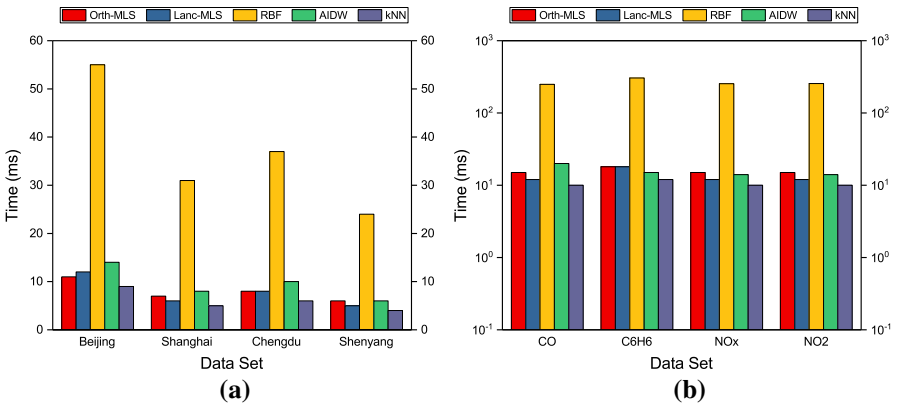

Fig. 7 Computational efficiency of imputing missing values when using different estimators. **a** Imputing the concentration of PM2.5 when $k = 80$. **b** Imputing the concentration of four gases when $k = 20$

When $k = 10$, the accuracy of the RBF estimator is the lowest, the kNN and AIDW estimators can achieve better accuracy, and those three MLS estimators are the best.

When $k = 20$, the accuracy of the RBF estimator is still the lowest, the kNN and AIDW estimators can achieve better accuracy, and the Lancaster's MLS estimator is the best.

When $k = 40$, for the estimation of PM2.5 concentration, the accuracy of the RBF and the Orthogonal MLS estimators are the lowest, the kNN and the Lancaster's MLS estimators can achieve better accuracy, and the AIDW estimator is the best. For the estimations of other gas concentrations, the accuracy of the RBF and the Orthogonal MLS estimators are also the lowest, the kNN and AIDW estimators can achieve better accuracy, and the Lancaster's MLS estimator is the best.

When $k = 80$, the kNN and Lancaster's MLS estimators are the best.

When $k = 160$, the kNN and Lancaster's MLS estimators are still the best.

In general, it could be summarized that the Lancaster's MLS estimator performs well in most cases.

3.3 Efficiency of Estimating Missing Values in IoT Time Series Data

The computational efficiency of those five estimators in the imputation of air quality data is illustrated in Fig. 7. It can be observed that: for the same variable and the fixed value of k, the RBF estimator is the slowest and the kNN estimator is the fastest.

4 Discussion

4.1 Effect of Values of *k* on the Estimation Accuracy

When using the MLS, RBF, AIDW, and kNN estimators to impute missing values in IoT time series data, a number of k nearest observations are selected. The use of

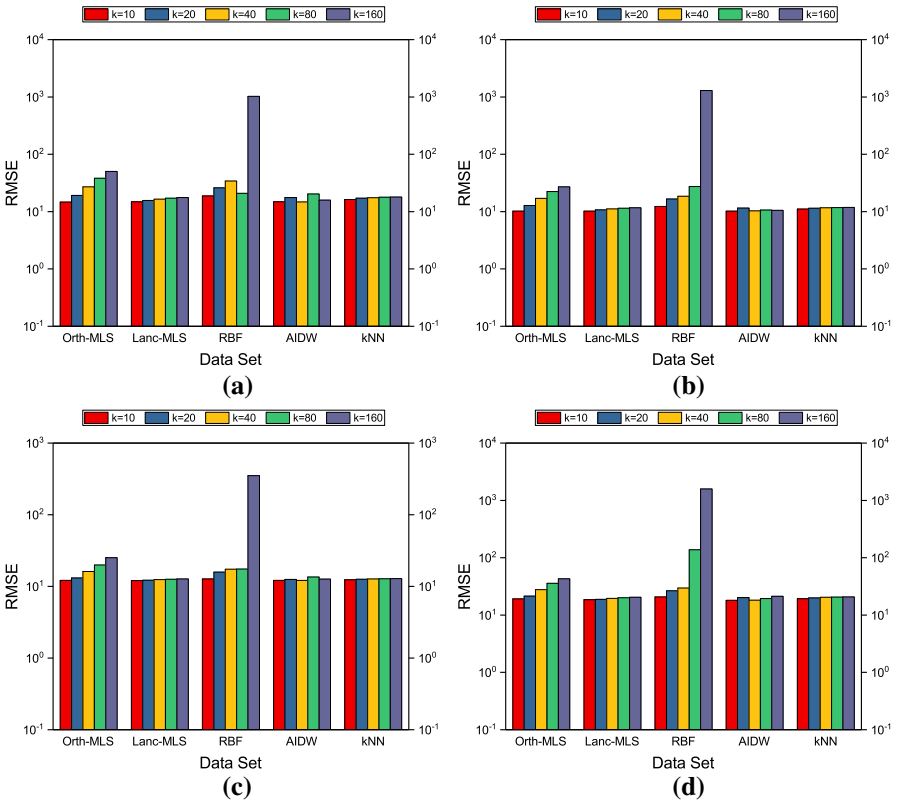

Fig. 8 Impact of values of k on the RMSE when imputing the concentration of PM2.5. **a** Beijing. **b** Shanghai. **c** Chengdu. **d** Shenyang

different values of k might stongly affect the accuracy of estimation. In our experiments, we configured the values of k as 10, 20, 40, 80, and 160, and then compared the estimation accuracy; see Figs. 8 and 9.

The experimental results indicate that: for any of the five estimators, the RMSE of imputation in general become larger with the increase of values of k. Noticabely, the accuracy when using the RBF estimator is the most significantly reduced, and the accuracy of the AIDW estimator fluctuates.

The above behavior is probably because of the following reasons. When the value of k is too small, the local trend of the time series data would be magnified, and the global trend would be masked. This may lead to large deviations or even extreme predicted values (see Fig. 10a). In contrast, when the value of k is too large, the local trend of the time series data cannot be preserved or reflected, and the global trend would become too smooth to produce inaccurate predictions (see Fig. 10b).

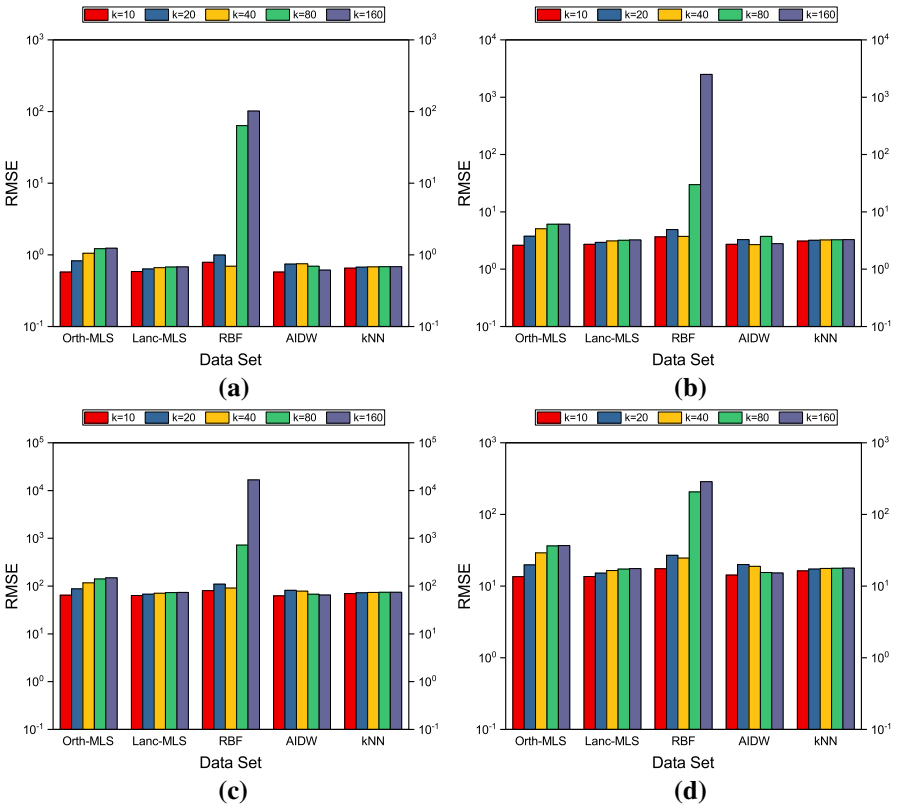

Fig. 9 Impact of values of k on the RMSE when imputing the concentration of four gases in Milan. **a** CO. **b** C_6H_6. **c** NO_X. **d** NO_2

4.2 Influence of Distribution of Missing Values on the Estimation Accuracy

In our experiments, we have sequentially selected 90% of the observed values as the training subset and the rest 10% as the testing subset. After imputing the assumed missing values in the testing subset, we have calculated the RMSE of the predicted 10% subset. According to this group of experiments, we have found that: the Lancaster's MLS estimator can achieve the highest accuracy in most cases, while the RBF is the worst.

Furthermore, we have also predicted the real rather than assumed missing values in the time series data. More specifically, we have used all the really observed values to interpolate the really missing values (see Table 1). After the interpolating, we calculate the SD rather than RMSE of (1) the original dataset without the predicted missing values and (2) the entire dataset including the predicted missing values; see Figs. 11 and 12.

If the SD calculated in the above two cases differs significantly, then it indicates that the prediction is not well performed. The main reason is that: if the size of successive missing values is large, the value near the middle of the missing segment may be

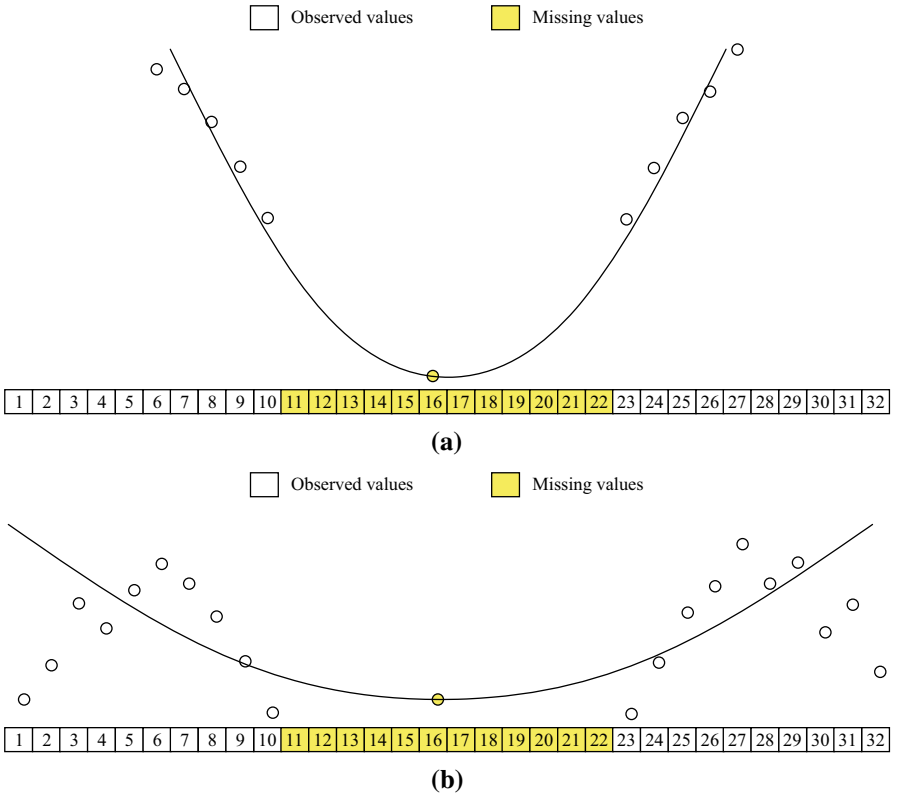

Fig. 10 Illustration of the impact of values of k on the imputed missing values. **a** Imputing when $k = 10$. **b** Imputing when $k = 20$

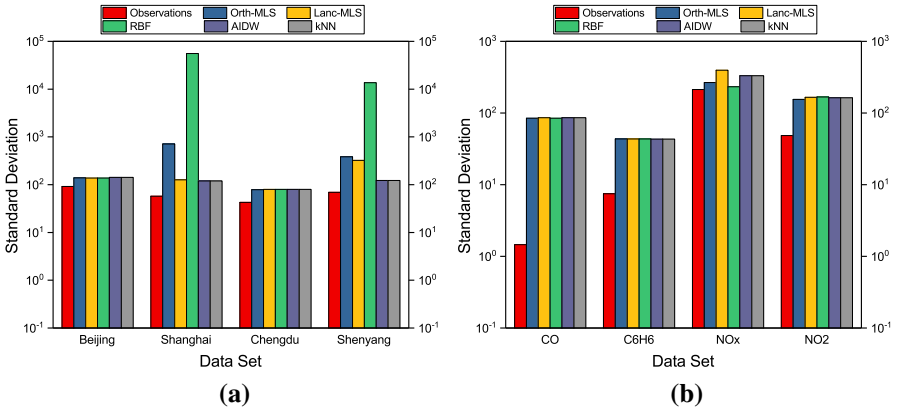

Fig. 11 Comparison of the SD when $k = 20$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

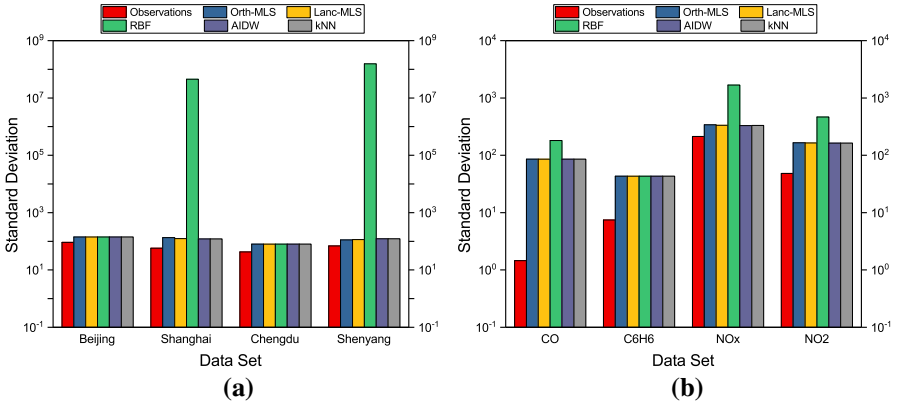

Fig. 12 Comparison of the SD when $k = 80$. **a** The concentration of PM2.5 in four Chinese cities. **b** The concentration of four gases in Milan

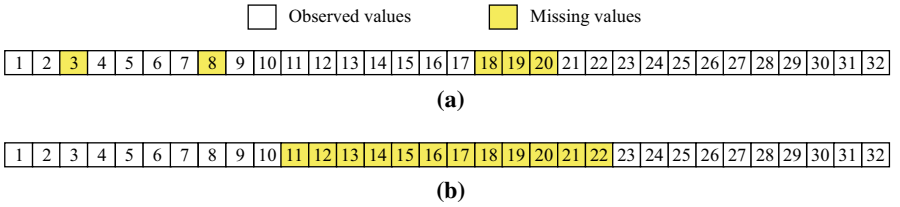

Fig. 13 Distribution of missing values. **a** The size of successive missing values is small; **b** The size of successive missing values is large

calculated as an extreme value when using the MLS or RBF predictor; see Fig. 13. In this case, the calculated SD of the entire dataset with the predicted missing values would be quite large. In contrast, if the size of successive missing values is small, the SD would not be quite large.

However, in the above two cases, the SD calculated when using the AIDW and *k*NN do not differ apparently. This is probably because that the predicted values calculated by the two methods are not larger than the maximum value of the original data set and are also not smaller than the minimum value of the original dataset.

5 Conclusions

In this paper, we have compared the performance of estimating missing values in IoT time series data by using different interpolation algorithms. Specifically, we have imputed the missing values in eight selected sets of time series data using three categories of interpolation algorithms (i.e., the MLS, RBF, and AIDW). We have found that in most experiments the estimation based on the Lancaster's MLS is the best. We have also found that the value of k and the distribution of missing values could strongly affect the accuracy of imputation.

The computational efficiency is one of the critical issues in estimating missing values, in particular, for the large-scale IoT time series data. In the current work, we only consider the computational efficiency of estimating missing values for small-scale IoT time series data. In the future, we will focus on large-scale IoT data and examine the computational efficiency of parallel interpolation algorithms such as those GPU-accelerated interpolation algorithms [6,7,13,14].

Acknowledgements This work was supported by the Natural Science Foundation of China (Grant Numbers 11602235 and 41772326), the College Students Innovation and Entrepreneurship Training Program (201811415014), the Fundamental Research Funds for the Central Universities (2652017086).

References

1. Ahmed, S.H., Rani, S.: A hybrid approach, smart street use case and future aspects for internet of things in smart cities. Future Gener. Comput. Syst. **79**, 941–951 (2018). https://doi.org/10.1016/j.future.2017.08.054
2. Alaa, M., Zaidan, A.A., Zaidan, B.B., Talal, M., Kiah, M.L.M.: A review of smart home applications based on Internet of Things. J. Netw. Comput. Appl. **97**, 48–65 (2017). https://doi.org/10.1016/j.jnca.2017.08.017
3. Beveridge, S.: Least squates estimation of missing values in time series. Commun. Stat.–Theory Methods **21**(12), 3479–3496 (1992). https://doi.org/10.1080/03610929208830990
4. Bhattacharjee, S., Mitra, P., Ghosh, S.K.: Spatial interpolation to predict missing attributes in GIS using Semantic Kriging. IEEE Trans. Geosci. Remote Sens. **52**(8), 4771–4780 (2014). https://doi.org/10.1109/TGRS.2013.2284489
5. Borgia, E.: The internet of things vision: Key features, applications and open issues. Comput. Commun. **54**, 1–31 (2014). https://doi.org/10.1016/j.comcom.2014.09.008
6. Cuomo, S., Galletti, A., Giunta, G., Marcellino, L.: Reconstruction of implicit curves and surfaces via RBF interpolation. Appl. Numer. Math. **116**(SI), 157–171 (2017). https://doi.org/10.1016/j.apnum.2016.10.016
7. Ding, Z., Mei, G., Cuomo, S., Xu, N., Tian, H.: Performance evaluation of gpu-accelerated spatial interpolation using radial basis functions for building explicit surfaces. Int. J. Parallel Program. (2017). https://doi.org/10.1007/s10766-017-0538-6
8. Haara, A., Maltamo, M., Tokola, T.: The k-nearest-neighbour method for estimating basal-area diameter distribution. Scand. J. Forest Res. **12**(2), 200–208 (1997). https://doi.org/10.1080/02827589709355401
9. Hui, T.K., Sherratt, R.S., Snchez, D.D.: Major requirements for building smart homes in smart cities based on internet of things technologies. Future Gener. Comput. Syst. **76**, 358–369 (2017). https://doi.org/10.1016/j.future.2016.10.026
10. Karkouch, A., Mousannif, H., Moatassime, H.A., Noel, T.: Data quality in internet of things: a state-of-the-art survey. J. Netw. Comput. Appl. **73**, 57–81 (2016). https://doi.org/10.1016/j.jnca.2016.08.002
11. Kouicem, D.E., Bouabdallah, A., Lakhlef, H.: Internet of things security: a top-down survey. Comput. Netw. (2018). https://doi.org/10.1016/j.comnet.2018.03.012
12. Lu, G.Y., Wong, D.W.: An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. **34**(9), 1044–1055 (2008). https://doi.org/10.1016/j.cageo.2007.07.010
13. Mei, G.: Evaluating the power of GPU acceleration for IDW interpolation algorithm. Sci. World J. (2014). https://doi.org/10.1155/2014/171574
14. Mei, G., Xu, L., Xu, N.: Accelerating adaptive inverse distance weighting interpolation algorithm on a graphics processing unit. R. Soc. Open Sci. (2017). https://doi.org/10.1098/rsos.170436
15. Ouaddah, A., Mousannif, H., Elkalam, A.A., Ouahman, A.A.: Access control in the internet of things: big challenges and new opportunities. Comput. Netw. **112**, 237–262 (2017). https://doi.org/10.1016/j.comnet.2016.11.007
16. Poulos, J., Valle, R.: Missing data imputation for supervised learning. Appl. Artifi. Intell. **32**(2), 186–196 (2018). https://doi.org/10.1080/08839514.2018.1448143

17. Qi, J., Yang, P., Min, G., Amft, O., Dong, F., Xu, L.: Advanced internet of things for personalised healthcare systems: a survey. Pervasive Mobile Comput. **41**, 132–149 (2017). https://doi.org/10.1016/j.pmcj.2017.06.018
18. Ray, P.: A survey on internet of things architectures. J. King Saud Univ. Comput. Inf. Sci. (2016). https://doi.org/10.1016/j.jksuci.2016.10.003
19. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference, pp. 517–524 (1968)
20. Shtiliyanova, A., Bellocchi, G., Borras, D., Eza, U., Martin, R., Carrere, P.: Kriging-based approach to predict missing air temperature data. Comput. Electron. Agric. **142**(A), 440–449 (2017). https://doi.org/10.1016/j.compag.2017.09.033
21. Silva, B.N., Khan, M., Han, K.: Towards sustainable smart cities: a review of trends, architectures, components, and open challenges in smart cities. Sustain. Cities Soc. **38**, 697–713 (2018). https://doi.org/10.1016/j.scs.2018.01.053
22. Sovilj, D., Eirola, E., Miche, Y., Bjrk, K.M., Nian, R., Akusok, A., Lendasse, A.: Extreme learning machine for missing data using multiple imputations. Neurocomputing **174**, 220–231 (2016). https://doi.org/10.1016/j.neucom.2015.03.108
23. Stekhoven, D.J., Buehlmann, P.: MissForest-non-parametric missing value imputation for mixed-type data. Bioinformatics **28**(1), 112–118 (2012). https://doi.org/10.1093/bioinformatics/btr597
24. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of internet of things for smart home: Challenges and solutions. J. Clean. Prod. **140**, 1454–1464 (2017). https://doi.org/10.1016/j.jclepro.2016.10.006
25. Tang, F., Ishwaran, H.: Random forest missing data algorithms. Stat. Anal. Data Min. **10**(6), 363–377 (2017). https://doi.org/10.1002/sam.11348
26. Trappey, A.J.C., Trappey, C.V., Govindarajan, U.H., Chuang, A.C., Sun, J.J.: A review of essential standards and patent landscapes for the Internet of Things: a key enabler for Industry 4.0. Adv. Eng. Inf. **33**, 208–229 (2017). https://doi.org/10.1016/j.aei.2016.11.007
27. Tsai, C.F., Li, M.L., Lin, W.C.: A class center based approach for missing value imputation. Knowl. Based Syst. **151**, 124–135 (2018). https://doi.org/10.1016/j.knosys.2018.03.026
28. Tzounis, A., Katsoulas, N., Bartzanas, T., Kittas, C.: Internet of things in agriculture, recent advances and future challenges. Biosyst. Eng. **164**, 31–48 (2017). https://doi.org/10.1016/j.biosystemseng.2017.09.007