

Article

A Two-Step Method for Missing Spatio-Temporal Data Reconstruction

Shifen Cheng ^{1,2} and Feng Lu ^{1,2,3,*}

¹ State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; chengsf@lreis.ac.cn

² Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou 350003, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: luf@lreis.ac.cn; Tel.: +86-10-6488-8966

Academic Editor: Wolfgang Kainz

Received: 5 April 2017; Accepted: 19 June 2017; Published: 23 June 2017

Abstract: Missing data reconstruction is a critical step in the analysis and mining of spatio-temporal data; however, few studies comprehensively consider missing data patterns, sample selection and spatio-temporal relationships. As a result, traditional methods often fail to obtain satisfactory accuracy or address high levels of complexity. To combat these problems, this study developed an effective two-step method for spatio-temporal missing data reconstruction (ST-2SMR). This approach includes a coarse-grained interpolation method for considering missing patterns, which can successfully eliminate the influence of continuous missing data on the overall results. Based on the results of coarse-grained interpolation, a dynamic sliding window selection algorithm was implemented to determine the most relevant sample data for fine-grained interpolation, considering both spatial and temporal heterogeneity. Finally, spatio-temporal interpolation results were integrated by using a neural network model. We validated our approach using Beijing air quality data and found that the proposed method outperforms existing solutions in term of estimation accuracy and reconstruction rate.

Keywords: spatio-temporal interpolation; spatio-temporal heterogeneity; dynamic sliding window; neural network

1. Introduction

Following both the rapid development and popularization of geographic information and the enhancement of data collection, data with temporal and spatial attributes are quickly accumulated and form large numbers of spatio-temporal datasets [1]; however, missing data are extremely common; for example, missing data on air quality monitoring sensor readings, missing data on floating car track points or the absence of mobile phone signaling records. If these gaps in data cannot be accurately filled, subsequent analysis and modeling of the data can lead to inaccurate results and unreasonable inference [2]. Simply deleting records containing missing data would lead to significant loss of original information and would be a waste of data resources [3]; therefore, methods to accurately and efficiently interpolate missing data are urgently needed.

In past decades, a large number of interpolation methods has been proposed to solve the problem of spatio-temporal missing data [4–10]. These methods can be roughly divided into three categories: spatial interpolation, temporal interpolation and spatio-temporal interpolation.

Spatial interpolation methods mainly use spatial correlation among data to interpolate missing data. Traditional methods (e.g., inverse distance weighting (IDW)) simply assume that the data distribution obeys the first law of geography, namely the closer data are in spatial distribution, the

greater the contribution they make to missing data interpolation. During interpolation, each site is assumed to be independent of the others, and weights are calculated by computing the distance between missing data and the surrounding site [11,12]. Most approaches use kriging, a linear regression method that utilizes minimum mean square error and does not treat each site independently. The method assumes consistent sample mean and variance to meet second order stationarity and assumes that the covariance between any two spatio-temporal points is only associated with the distance (i.e., the absolute position of time and space is irrelevant) [13,14]. However, due to the existence of spatial and temporal heterogeneity, the data distribution can show uneven characteristics and relationships according to different regions [15]; therefore, the accuracy of interpolation results obtained by existing methods remains unsatisfactory if data are not homogeneously distributed. To solve this problem [16] considered spatial autocorrelation and heterogeneity in a study area and proposed a point estimation model of biased hospital-based area disease estimation (P-BSHADE). The P-BSHADE method calculates the covariance and correlation coefficient of historical observational data and uses the expectation between surrounding stations and interpolation sites to get an optimal linear unbiased estimator. However, in cases of continuous missing data, the method leads to a singular value of the missing data matrix, which results in large interpolation error. At the same time, this method does not consider the heterogeneity of the time dimension [2].

Time series prediction methods typically use historical data for a given location to build a prediction framework for predicting the values of missing data points at the same site. The autoregressive integrated moving-average (ARIMA) model [17] and simple exponential smoothing (SES) [18] are two representative examples of this approach. However, this approach fails to address two major problems. First, many prediction models do not fully utilize the essential characteristics of spatio-temporal data, which can degrade performance; second, if a consecutive series of data is all lost, prediction methods often cannot achieve complete reconstruction [19].

Given that single dimension interpolation methods only consider spatial or temporal dimensions, achieving satisfactory interpolation results is challenging. In recent years, a number of studies have extended single dimension interpolation methods to consider both space and time; for example, spatio-temporal probabilistic principal component regression (ST-PCR), spatio-temporal IDW (ST-IDW), spatio-temporal kriging (ST-kriging) and the spatio-temporal heterogeneous covariance method (ST-HC) [2,3,7,9,10,20–22]. ST-PCR [9] is a statistical learning-based method, which takes advantage of the statistical feature of observed data. However, it often needs a strong hypothesis over the data. ST-IDW [23] defined a three-dimensional space-time distance, which then applied IDW to estimate missing values; however, due to the existing problems with the IDW method, application of ST-IDW remains limited and fails to achieve unbiased estimation. The ST-kriging method [14] estimates the interpolation function by adopting a spatio-temporal covariance function, but it does not take into account the effects of spatial and temporal heterogeneity on interpolation results. To address this issue, the ST-HC method, which is an extension of P-BSHADE, estimates missing data by considering temporal and spatial heterogeneity. Missing datasets are firstly partitioned into homogenous spatial regions, and then, the most correlated spatial sampling and time sampling series are selected for the partition of missing data. According to the P-BSHADE algorithm, both spatial and temporal contribution weights are calculated to obtain the best linear unbiased estimates in spatial and temporal dimensions. Finally, using the correlation coefficient to determine the spatial and temporal weights, estimated values in spatial and temporal dimensions are integrated to obtain overall estimated values of missing data [2]. However, this method requires the whole dataset to participate in computation, which leads to high computational complexity and a large volume of redundant data. For example, when the time span of the dataset is large, including the n dimensional space sequence, both n^2 covariance and the correlation coefficient need to be calculated for each missing data point. In addition, when data are missing continuously, interpolation accuracy is low, and it may even be impossible to obtain a final interpolation result.

Missing data reconstruction methods are further challenged by the variation in patterns of missing data [24,25]; for example, missing data completely at random, non-random missing data [26], or whole blocks of missing data [27] (Figure 1). The work in [10] formulates the spatio-temporal sensory data as a high-dimensional tensor, using the tensor completion method to recover missing values. However, when a whole temporal or spatial dimension of the data is missing, this method may fail. The same problem exists for other spatio-temporal interpolation methods (e.g., IDW, kriging, P-BSHADE, ST-IDW, ST-kriging, ST-HC). Furthermore, most existing methods assume a linear relationship between spatial and temporal data (e.g., ST-HC, ST-IDW, ST-kriging); however, the relationship between spatial and temporal data may be linear or nonlinear. To address these issues, in this study, we developed a new two-step method to reconstruct missing spatio-temporal data (ST-2SMR).

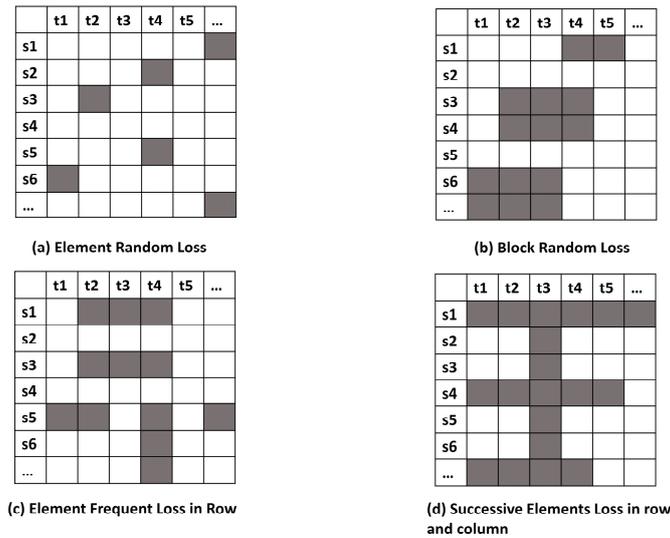


Figure 1. Patterns of missing spatio-temporal data. Black squares represent missing data.

2. Materials and Methods

2.1. Problem Definitions

Definition 1. Suppose that $y(S, T)$ represents a spatio-temporal missing dataset, where S and T are the spatial and temporal dimension, respectively; $S = \{s_1, s_2, \dots, s_m\}$, $T = \{t_1, t_2, \dots, t_n\}$, m is the total number of spatial objects and n is the total number of timestamps. An entry $V_{ij} = y(s_i, t_j)$ refers to the value of the i -th spatial object at the j -th timestamp ($1 \leq i \leq m, 1 \leq j \leq n$). The definition of the size of the sliding window is w ($1 \leq w \leq n$). If there exist $\{V_{ij} \neq \emptyset \mid \forall (j - \frac{w-1}{2}) \leq j \leq (j + \frac{w-1}{2})\}$, then s_i is a complete temporal sequence. If there exist $\{V_{ij} \neq \emptyset \mid \exists (j - \frac{w-1}{2}) \leq j \leq (j + \frac{w-1}{2})\}$, then s_i exists, missing at the j -th timestamp. Similarly, if there exist $\{V_{ij} \neq \emptyset \mid \forall 1 \leq i \leq m\}$, then t_j is a complete spatial sequence. If there exist $\{V_{ij} \neq \emptyset \mid \exists 1 \leq i \leq m\}$, it means that t_j exists, missing at the i -th spatial object.

Definition 2. Suppose that $\hat{z}_i = \hat{y}(s_i, t_i)$ is the estimated value of $z_i = y(s_i, t_i)$, then the problem can be defined as:

$$\begin{cases} \min_{MSE} = E(\hat{z}_i - z_i) \\ \text{s.t. } E(\hat{z}_i) = z_i \end{cases} \quad (1)$$

where MSE is the minimum mean square error and E is the statistical expectation. Meet $E(\hat{z}_i) = z_i$ to ensure that the process of interpolation results in an unbiased estimate.

2.2. Method Framework

The method developed in this study (ST-2SMR) takes into account the problems of existing missing data interpolation methods and was constructed using 5 main steps (Figure 2): the partition of dataset into training and testing parts, coarse-grained interpolation, fine-grained interpolation, integration of temporal and spatial interpolation results and model performance evaluation.

1. First, because the ST-2SMR method uses a neural network model to combine the interpolation results in spatial and temporal dimensions, the space-time missing dataset was divided, with 80% selected for parameter training and 20% used as a test dataset to evaluate model performance, to avoid overfitting and to improve generalization ability.
2. Second, to avoid the influence of continuous missing data on fine-grained interpolation, coarse-grained interpolation for missing datasets was applied.
3. Based on the results of coarse-grained interpolation, spatial and temporal heterogeneity was used to perform fine-grained interpolation. During this process, it was necessary to calculate the correlation coefficient and covariance between time and space sequences to fit the parameters. If whole time series or spatial sequences are involved in interpolation, redundant sample data increase computational complexity; therefore, to improve the accuracy and speed of interpolation, a suitable sliding window was introduced to ensure the strongest correlation between sample data and missing data. Next, a heterogeneous covariance function was constructed for the space-time dimension. The best unbiased estimate of missing data can then be obtained by maximizing the objective function.
4. After interpolation of temporal and spatial dimensions, estimated values without missing data were chosen as training samples for the neural network, which is used to mine nonlinear relationships in spatio-temporal data. Estimated values of missing data were then obtained by inputting the spatio-temporal interpolation results of missing data into the trained neural network model.
5. Finally, the performance of the model was evaluated using the test dataset.

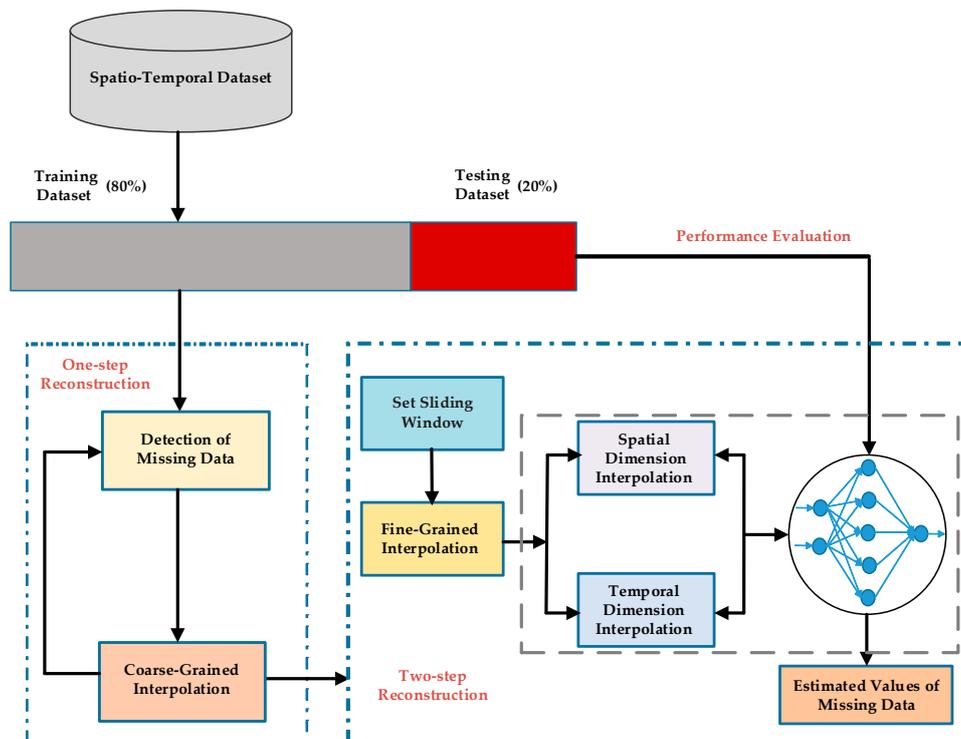


Figure 2. Framework of model development.

3. Detailed Design of ST-2SMR

3.1. Coarse-Grained Interpolation

Using existing spatial and temporal interpolation methods, it is difficult to obtain accurate interpolation estimates where there is a lack of sample points around missing points (Figure 1b,c). In some cases, it may even be impossible to obtain interpolation estimates (Figure 1d). We overcame this issue by introducing coarse-grained interpolation before fine-grained interpolation in order to eliminate the influence of continuous missing data. Using this approach, the ST-2SMR method is able to locate missing patterns in any combination, which ensures that the method is suitable for any serious loss situation.

In the spatial dimension, we used a classical multivariate statistical model (IDW) to interpolate missing data. IDW adopts the observed data of adjacent space points to estimate unknown data [11,28]. When the distance of the adjacent space points is closer to the point of interpolation, the spatial contribution value is larger. Interpolation estimates for missing data $v_spatial_0$ are calculated as follows:

$$v_spatial_0 = \sum_{i=1}^m \chi_i v_spatial_i \quad (2)$$

$$\chi_i = \frac{d_i^{-\alpha}}{\sum_{i=1}^m d_i^{-\alpha}} \quad (3)$$

where d_i denotes the distance between interpolation points and observation points and α is the decay weight rate, where a larger α denotes a faster decay by the distance.

In the time dimension, an exponential moving average model (SES) was used to estimate missing data [18]. SES assumes that there is strong temporal correlation between data, and the closer the sampling point is to the missing point, the bigger the weight that it is given. Traditional SES uses only the sampling data that are before the interpolation time point; however, when the time span is large, this can lead to too many irrelevant data involved in the calculation, which reduces interpolation accuracy. Here, we extended it to set a sliding window wc , which takes only before the wc time slice of the missing data and the last wc time slice of the missing data as the sampling point for the interpolation calculation. The model can be expressed as:

$$v_temporal_0 = \frac{\sum_{j=1}^{wc} v_temporal_j * \beta * (1 - \beta)^{th_j - 1}}{\sum_{j=1}^{wc} \beta * (1 - \beta)^{th_j - 1}} \quad (4)$$

where $v_temporal_0$ is the estimated value for missing data, th_j is a time interval between the sampling data and the missing data and β is a smoothing parameter with a range of (0, 1).

Assuming that $V_{4,6}$ are the data to be interpolated, we first selected all no missing data at t_6 , using IDW interpolation estimation (Figure 3). In the temporal dimension, we set the sliding window to be $wc = 4$ and selected $\{V_{4,2}, V_{4,3}, V_{4,5}, V_{4,7}, V_{4,8}, V_{4,10}\}$ as the sampling data for interpolation. If both the IDW and SES methods can obtain an estimated value, the mean of the two is taken as the interpolation result of $V_{4,6}$. If SES or IDW fail to get an interpolation result, the other estimated value is taken as the estimated value. If both SES and IDW fail to obtain an interpolation result, then $V_{4,6}$ is still missing, and the fine-grained interpolation algorithm is needed to estimate its value. The pseudocode of this process is shown in Algorithm 1.

Algorithm 1: Coarse-Grained interpolation.

Input: Original Missing Matrix $V_{m \times n}$
 Temporal Threshold wc
 Parameter of IDW α
 Parameter of SES β

Output: Coarse-Grained Imputing Matrix $C_M_{m \times n}$

```

1  $C\_M_{m \times n} \leftarrow Initialization(V_{m \times n})$ 
2 For  $i = 1$  to  $m$ 
3   For  $j = 1$  to  $n$ 
4     If  $V_{ij}$  is missing value then
5        $v_{c\_spatial} \leftarrow 0$ 
6        $v_{c\_temporal} \leftarrow 0$ 
7        $v_{c\_spatial} \leftarrow IDW(V_{ij}, \alpha)$ 
8        $v_{c\_temporal} \leftarrow SES(V_{ij}, \beta, wc)$ 
9     If  $v_{c\_spatial}, v_{c\_temporal}$  are not missing value then
10       $C\_M_{ij} \leftarrow (v_{c\_spatial} + v_{c\_temporal}) / 2$ 
11     Else if  $v_{c\_spatial}$  is not missing value then
12       $C\_M_{ij} \leftarrow v_{c\_spatial}$ 
13     Else if  $v_{c\_temporal}$  is not missing value then
14       $C\_M_{ij} \leftarrow v_{c\_temporal}$ 
15     Else
16       $C\_M_{ij} \leftarrow \emptyset$ 
17   End for
18 End for
    
```

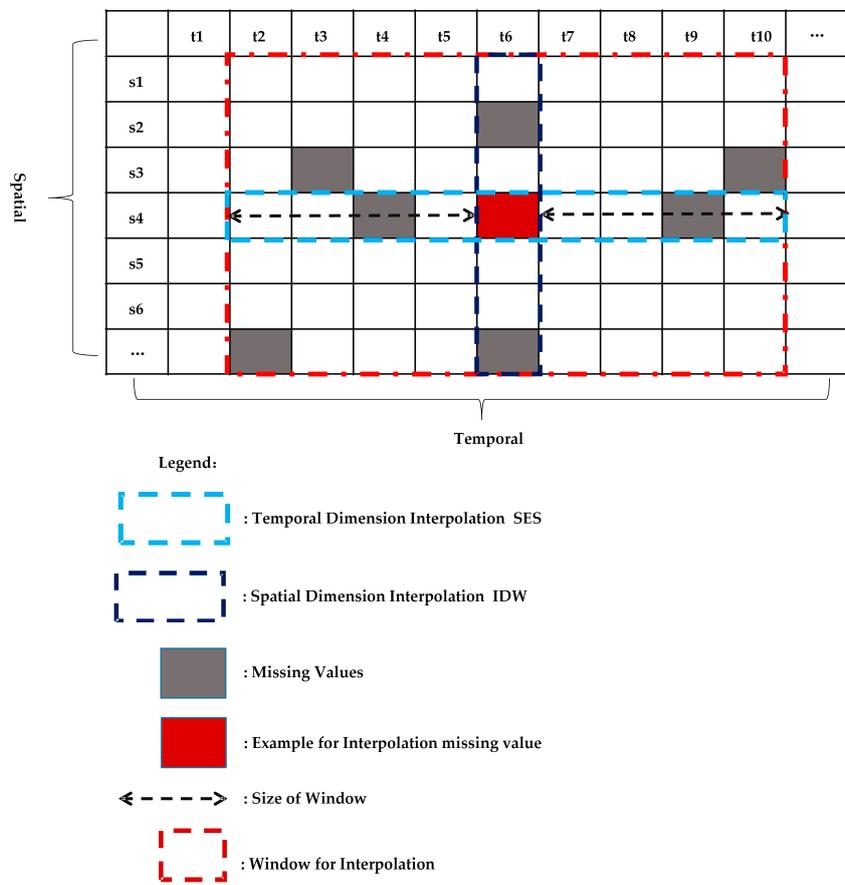


Figure 3. Coarse-grained interpolation method.

3.2. Fine-Grained Interpolation

3.2.1. Sliding Window

Before fine-grained interpolation, the ST-2SMR model needs to set up a dynamic sliding window to determine sample data. Owing to the strong temporal dependence of spatial and temporal data, selecting different numbers of data for interpolation estimation can lead to different results. If the window is set too small, sample data cannot fully reflect spatio-temporal relationships; if the window is too large (i.e., too many data are used), significant redundant data are introduced to the calculation, and the computational complexity increases.

Considering the fact that spatio-temporal data from a short period of time remain within the approximate correlation, take the average correlation of the missing data sequence and their adjacent spatial sequences to determine the optimal sample data using the expressions:

$$R_{begin} = \frac{1}{n-w_{begin}} \sum_{j=n-1}^{w_{begin}} Corr(t_n, t_j)$$

objective : $min w_{begin}$

subject to : $R_{begin} = \max(R_{begin})$

(5)

$$R_{end} = \frac{1}{w_{end}-n} \sum_{i=n+1}^{w_{end}} Corr(t_n, t_i)$$

objective : $min w_{end}$

subject to : $R_{end} = \max(R_{end})$

(6)

where n is the timestamp for missing data, j are the first j moments of missing data, i are the last i moments of missing data, $Corr(t_n, t_j)$ is the correlation coefficient between the spatial sequence of missing data and the first j spatial sequences, $Corr(t_n, t_i)$ is the correlation coefficient between the spatial sequence of missing data and the last i spatial sequences, w_{begin} is the beginning of the window and w_{end} is the end of the window. w_{begin} and w_{end} are determined heuristically and are initially set to $n - 1$ and $n + 1$. $Corr(t_n, t_j)$ and $Corr(t_n, t_i)$ are first calculated, then w_{begin} moves forward, and w_{end} moves backwards until the mean correlation coefficient reaches the maximum. Assuming that $V_{4,6}$ is the missing datum to be interpolated, t_6 is the spatial sequence of missing data, and $t_2 \sim t_{10}$ is the sliding window (Figure 4). The pseudocode of this process is shown in Algorithm 2.

Algorithm 2: Selected optimal window (SOM).

Input: Missing Spatial Series t_n

Output: Start of Window w_{begin}

End of Window w_{end}

1 $R_{begin} \leftarrow 0$

2 $R_{end} \leftarrow 0$

3 **For** $j = n - 1$ to 1

4 $R_{last} \leftarrow R_{begin}$

5 $R_{begin} \leftarrow (R_{begin} + Corr(t_n, t_j)) / (n - j)$

6 **If** $R_{begin} < R_{last}$

7 **Return** j

8 **End if**

9 **End for**

10 **For** $i = n + 1$ to end of the timestamp

11 $R_{last} \leftarrow R_{end}$

12 $R_{end} \leftarrow (R_{end} + Corr(t_n, t_i)) / (i - n)$

13 **If** $R_{end} < R_{last}$

14 **Return** i

15 **End if**

16 **End for**

17 $w_{begin} \leftarrow j$

18 $w_{end} \leftarrow i$

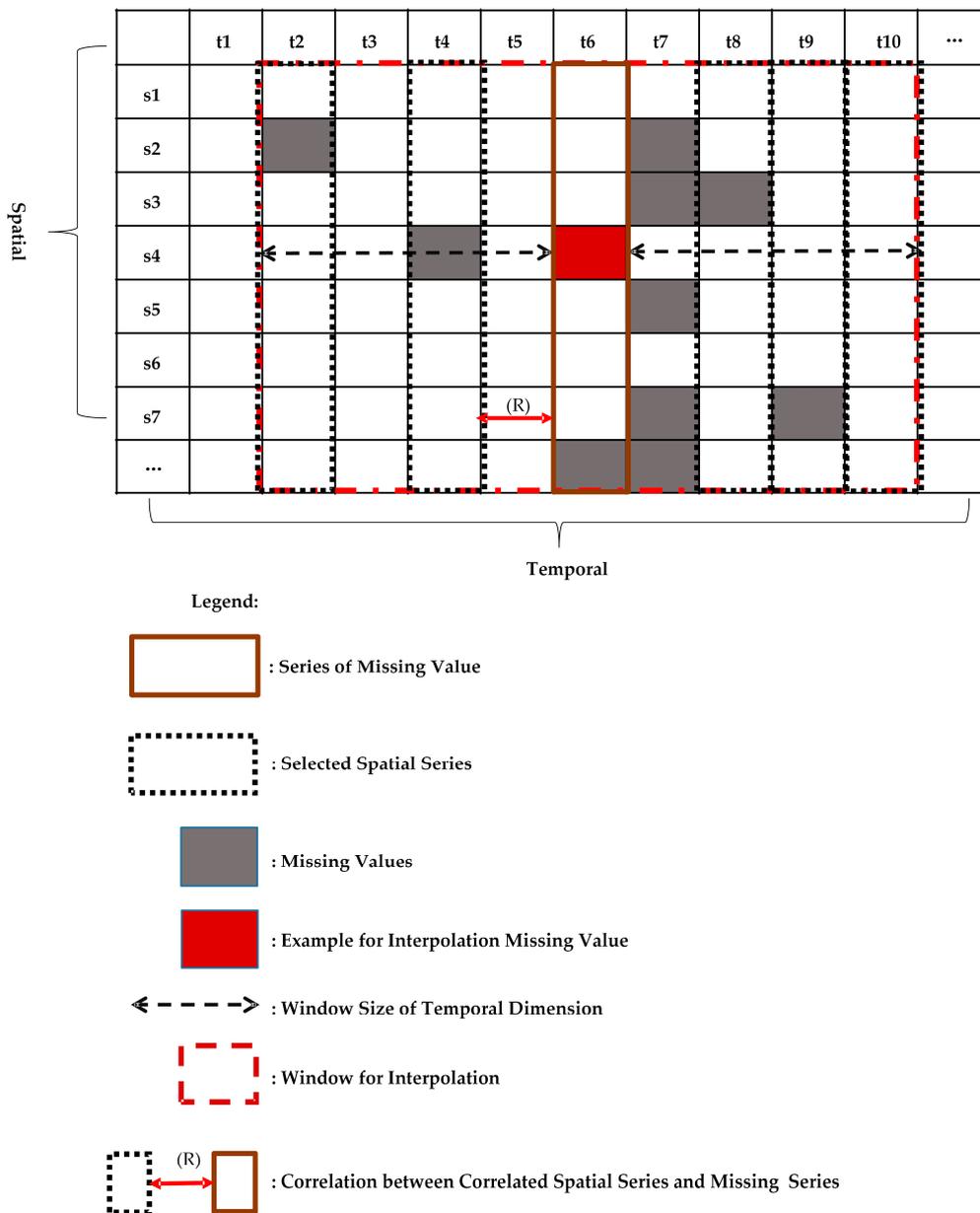


Figure 4. Sliding window selection.

3.2.2. Fine-Grained Spatial Dimension Interpolation

In the spatial dimension, we first selected a sliding window based on the optimal window selection algorithm (SOM), whose size is ws . Assuming that $V_{4,6}$ is the missing data to be interpolated, the start and end positions of the selected window are centered on the $V_{4,6}$ in the first 4 columns and the last 4 columns of Figure 5. In this window, we chose the ms time series with the largest correlation of missing data. In detail, we adopted a pair-wise approach for calculating the correlation between the time series of missing data and its spatial neighbor sampling data, and then, ms spatial sampling data were selected to calculate the estimated value using the expression:

$$\hat{y}_0 = \sum_{i=1}^{ms} w_i y_i \tag{7}$$

where y_i denotes the i -th spatial neighbor sampling data of missing data and w_i denotes the corresponding contribution weight of y_i . As shown in Figure 5, suppose $ms = 3$, if $\{s_2, s_6, s_8\}$ is the most relevant time series with missing data, then we take $\{V_{2,6}, V_{6,6}, V_{8,6}\}$ as the sampled data for interpolation.

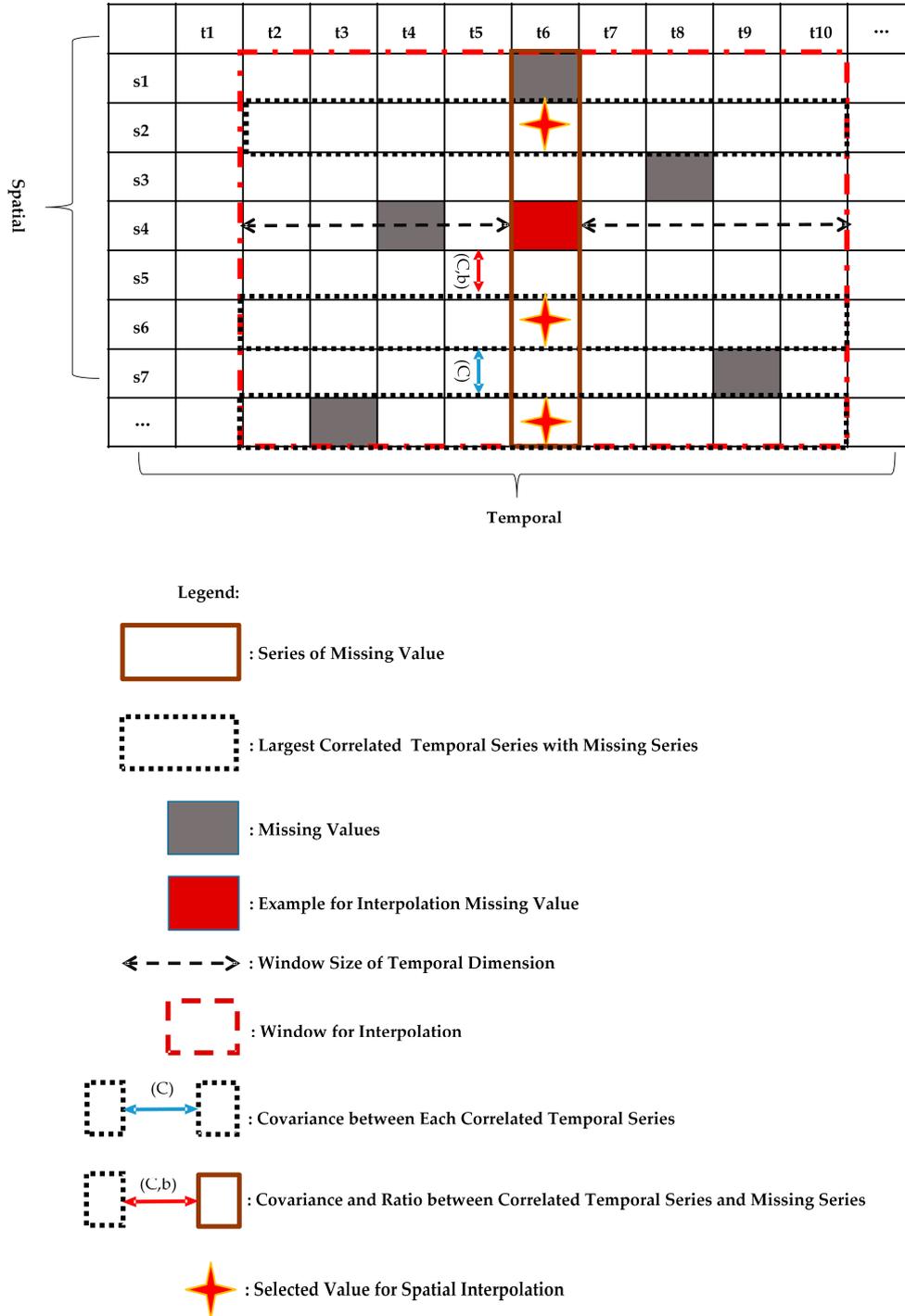


Figure 5. Fine-grained interpolation in the spatial dimension.

In order to ensure \hat{y}_0 is an unbiased estimator of missing data, the following conditions must be satisfied:

$$E(\hat{y}_0) = E(y_0) \tag{8}$$

where $E(\cdot)$ represents the statistical expectation. Considering the heterogeneity of the spatial dimension, we introduced the parameter b_i to calculate the ratio between the statistical expectation of \hat{y}_0 and y_0 to characterize spatial heterogeneity.

$$b_i = E(y_i)/E(y_0) \quad (9)$$

Combining Formulae (7)–(9), the constraint condition of w_i is as follows:

$$\sum_{i=1}^{ms} w_i b_i = 1 \quad (10)$$

In order to obtain the parameter w_i , the objective function was constructed to minimize the variance between the missing and true data.

$$\min_w [\sigma^2_{\hat{y}_0} = E(\hat{y}_0 - y_0)^2] \quad (11)$$

Among them, Formula (11) can be resolved as follows:

$$\sigma^2_{\hat{y}_0} = C(\hat{y}_0, \hat{y}_0) + C(y_0, y_0) - 2C(\hat{y}_0, y_0) = \sigma^2_{y_0} + \sum_{i=1}^{ms} \sum_{j=1}^{ms} w_i w_j C(y_i, y_j) - 2 \sum_{i=1}^{ms} w_i C(y_i, y_0) \quad (12)$$

where C represents the covariance between different spatial points. Considering Formula (10), Formulae (11) and (12) can be written as:

$$\begin{cases} \arg_w \min \sigma^2_{\hat{y}_0} = \arg \min E(\hat{y}_0 - y_0)^2 \\ \text{s.t. } \sum_{i=1}^{ms} w_i b_i = 1 \end{cases} \quad (13)$$

Then, the problem of solving the parameter w_i was transformed into a standard Lagrange constrained optimization problem, and Formula (13) was re-written as:

$$\sigma^2_{\hat{y}_0} = \sigma^2_{y_0} + \sum_{i=1}^{ms} \sum_{j=1}^{ms} w_i w_j C(y_i, y_j) - 2 \sum_{i=1}^{ms} w_i C(y_i, y_0) + 2\mu \left(\sum_{i=1}^{ms} w_i b_i - 1 \right) \quad (14)$$

where μ is a Lagrange multiplier. The partial derivatives of $\sigma^2_{\hat{y}_0}$ produces the equations:

$$\begin{aligned} \frac{\delta \sigma^2_{\hat{y}_0}}{\delta w_i} = 0 & \Rightarrow 2 \sum_{j=1}^{ms} w_j C(y_i, y_j) - 2C(y_i, y_0) + 2\mu b_i = 0 \\ & \Rightarrow \sum_{j=1}^{ms} w_j C(y_i, y_j) + \mu b_i = C(y_i, y_0) \end{aligned} \quad (15)$$

Formula (15) can be written in the matrix form:

$$\begin{bmatrix} C(y_1, y_1) & \cdots & C(y_1, y_{ms}) & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ C(y_{ms}, y_1) & \cdots & C(y_{ms}, y_{ms}) & b_{ms} \\ b_1 & \cdots & b_{ms} & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_{ms} \\ \mu \end{bmatrix} = \begin{bmatrix} C(y_1, y_0) \\ \vdots \\ C(y_{ms}, y_0) \\ 1 \end{bmatrix} \quad (16)$$

In order to obtain the parameter w_i in Formula (16), we first calculated the covariance matrix C_S between the most relevant temporal series and the ratio of statistical expectation b_i and covariance C_i between the most relevant temporal series and the missing temporal series (Lines 6–9 of Algorithm 3). Then, we joined these values into a matrix like Formula (16) and solved this matrix to get parameter w_i (Lines 10–12 of Algorithm 3). Finally, through Formula (7), we obtained estimated values of missing data. As shown in Figure 5, parameter w_i of $V_{4,6}$ can be calculated by: (1) obtaining the covariance

of $\{s_2, s_6, s_8\}$ and the covariance $C(s_2, s_4)$, $C(s_6, s_4)$ and $C(s_8, s_4)$; (2) calculating the ratio of statistical expectation $b_1 = E(s_2)/E(s_4)$, $b_2 = E(s_6)/E(s_4)$ and $b_3 = E(s_8)/E(s_4)$; (3) obtaining weights w_1 , w_2 and w_3 by solving the matrix; and finally, (4) interpolating $V_{4,6}$ from $w_1 \times V_{2,6} + w_2 \times V_{6,6} + w_3 \times V_{8,6}$.

Algorithm 3: Fine-Grained spatial interpolation.

Input: Coarse-Grained Matrix $C_{M \times n}$
 Number of Spatial Neighbors ms
Output: Fine-Grained Spatial Matrix $F_{S \times n}$

- 1 For $i = 1$ to m
- 2 For $j = 1$ to n
- 3 $ws \leftarrow SOM(C_{M \times n}, C_{M_{ij}})$
- 4 $Rss \leftarrow Corrcoeff(M_{W_{m \times ws}}, column)$
- 5 $S_Correlate \leftarrow Max\ Ms\ Correlate(Rss, S_Missing_i, ms)$
- 6 $C_S \leftarrow Cov(S_Correlate)$
- 7 For each $S_k \in S_Correlate$
- 8 $C_k \leftarrow Cov(S_k, S_Missing_i)$
- 9 $b_k \leftarrow Mean(S_k) / Mean(S_Missing_i)$
- 10 $C_Matrix_Left_{(ms+1) \times (ms+1)} \leftarrow Combine(C_S, b)$
- 11 $C_Matrix_Right_{(ms+1) \times 1} \leftarrow Combine(C, 1)$
- 12 $w \leftarrow C_Matrix_Left_{(ms+1) \times (ms+1)} C_Matrix_Right_{(ms+1) \times 1}$
- 13 $F_{S_{ij}} \leftarrow Dot\ Product(S_Correlate_i, w)$
- 14 End for
- 15 End for

3.2.3. Fine-Grained Temporal Dimension Interpolation

In the temporal dimension, we used the SOM algorithm to select an optimal window as the data matrix for fine-grained temporal dimension interpolation. In this sliding window, the nt sample data with the largest correlation of missing data were chosen. The estimated value of missing data \hat{t}_0 was calculated as follows:

$$\hat{t}_0 = \sum_{j=1}^{nt} \varphi_j t_j \quad (17)$$

where t_j denotes the j -th temporal neighbor sampling data of missing data and φ_j denotes the corresponding contribution weight of t_j . Similar to Formulae (12)–(14), to ensure \hat{t}_0 is an unbiased estimator of missing data and to calculate the weight φ_j , the following conditions must be satisfied:

$$\sigma^2_{\hat{t}_0} = \sigma^2_{t_0} \sum_{j=1}^{nt} \sum_{g=1}^{nt} \varphi_j \varphi_g C(t_j, t_g) - 2 \sum_{j=1}^{nt} \varphi_j C(\hat{t}_0, t_0) + 2v \left(\sum_{j=1}^{nt} \varphi_j a_j - 1 \right) \quad (18)$$

where v is a Lagrange multiplier, t_0 is the true value of missing data and a_j is the ratio of statistical expectation between the most relevant spatial series and the spatial series of missing data. The partial derivatives of $\sigma^2_{\hat{t}_0}$ can be written as:

$$\frac{\delta \sigma^2_{\hat{t}_0}}{\delta \varphi_i} = 0 \Rightarrow \sum_{j=1}^{nt} \varphi_j C(t_i, t_j) + v a_j = C(t_i, t_0) \quad (19)$$

Formula (19) can be written in the form of a matrix:

$$\begin{bmatrix} C(t_1, t_1) & \cdots & C(t_1, t_{nt}) & a_1 \\ \vdots & \ddots & \vdots & \vdots \\ C(t_{nt}, t_1) & \cdots & C(t_{nt}, t_{nt}) & a_{nt} \\ a_1 & \cdots & a_{nt} & 0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_{nt} \\ v \end{bmatrix} = \begin{bmatrix} C(t_1, t_0) \\ \vdots \\ C(t_{nt}, t_0) \\ 1 \end{bmatrix} \quad (20)$$

As shown in Figure 6, in order to estimate the value of missing data $V_{4,6}$, we first adopted the SOM algorithm to select $t_2 - t_{10}$ as the sliding window. In this window, we choose the nt spatial series with the largest correlation of missing data. For example, when $nt = 4$, the spatial series would be $\{t_2, t_4, t_8, t_9\}$, and we would take $\{V_{4,2}, V_{4,4}, V_{4,8}, V_{4,9}\}$ as the sampled data for interpolation (Lines 3–5 of Algorithm 4). The covariance matrix C_T was calculated from the most relevant spatial series and the ratio of statistical expectation a_j and covariance C_j by using the most relevant spatial series and the missing spatial series (Lines 6–9 of Algorithm 4). Then, we joined these values into matrix Formula (20) and solved this matrix to get parameter φ_j (Lines 10–12 of Algorithm 4). Finally, through Formula (17), interpolation result $V_{4,6}$ was calculated as $\varphi_1 \times V_{4,2} + \varphi_2 \times V_{4,4} + \varphi_3 \times V_{4,8} + \varphi_4 \times V_{4,9}$.

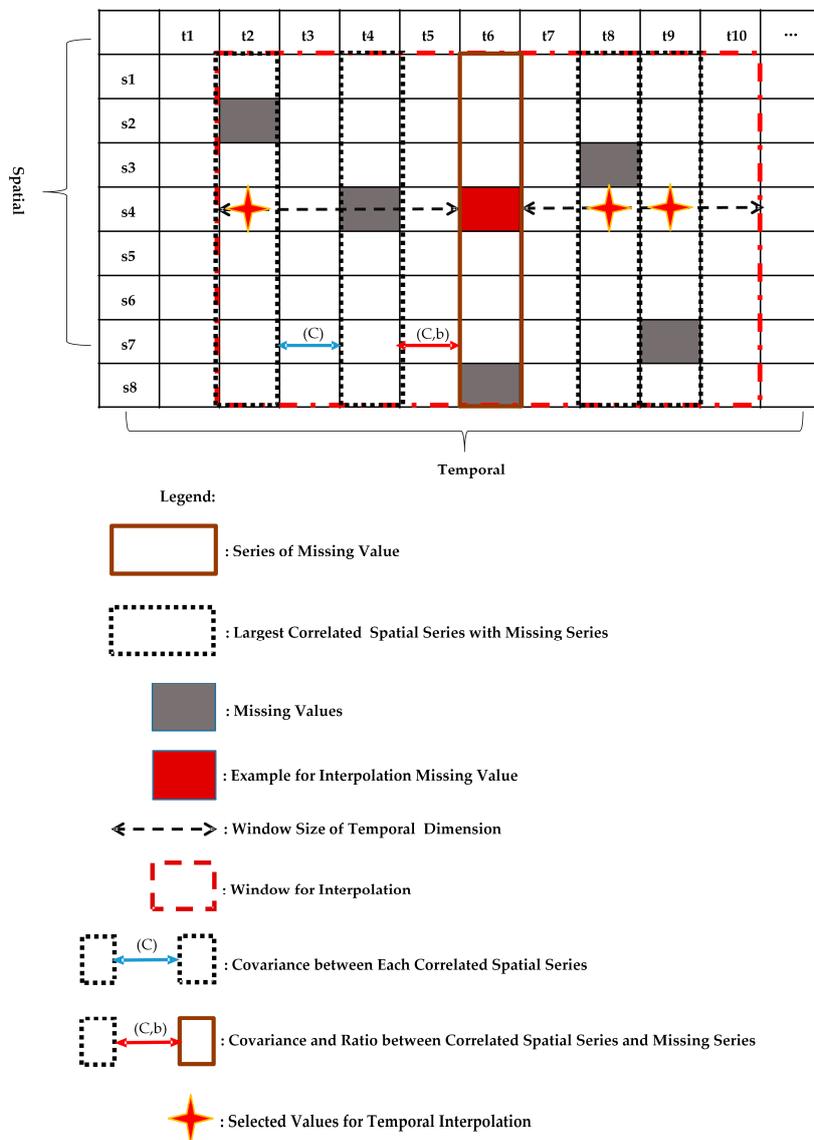


Figure 6. Fine-grained interpolation in the temporal dimension.

Algorithm 4: Fine-Grained temporal interpolation.

Input: Coarse-Grained Matrix $C_{M_{m \times n}}$
Number of Temporal Neighbors nt
Output: Fine-Grained Temporal Matrix $F_{T_{m \times n}}$

- 1 For $i = 1$ to m
- 2 For $j = 1$ to n
- 3 $wt \leftarrow \text{SOM}(C_{M_{m \times n}}, C_{M_{ij}})$
- 4 $Rtt \leftarrow \text{Corrcoef}(M_{W_{m \times wt}}, row)$
- 5 $T_Correlate \leftarrow \text{Max Nt Correlate}(Rtt, T_Missing_j, nt)$
- 6 $C_T \leftarrow \text{Cov}(T_Correlate)$
- 7 For each $T_k \in T_Correlate$
- 8 $C_k \leftarrow \text{Cov}(T_k, T_Missing_j)$
- 9 $b_k \leftarrow \text{Mean}(T_k) / \text{Mean}(T_Missing_j)$
- 10 $C_Matrix_Left_{(nt+1) \times (nt+1)} \leftarrow \text{Combine}(C_T, b)$
- 11 $C_Matrix_Right_{(nt+1) \times 1} \leftarrow \text{Combine}(C, 1)$
- 12 $\varphi \leftarrow C_Matrix_Left_{(nt+1) \times (nt+1)} C_Matrix_Right_{(nt+1) \times 1}$
- 13 $F_{T_{ij}} \leftarrow \text{Dot Product}(T_Correlate_j, \varphi)$
- 14 End for
- 15 End for

3.3. Spatio-Temporal Integration

After obtaining interpolation results for the time and space dimensions, the BP (back propagation) neural network was trained to integrate spatial and temporal interpolation results to obtain final missing data estimation values [5]. The BP neural network can be regarded as a nonlinear function. When the number of input nodes is n , the output node is m , and the BP neural network expresses the mapping function from n independent variables to m dependent variables [29].

Appropriate samples are needed for training and testing neural networks. In this study, we first detect no missing data in the fine-grained temporal matrix $F_{T_{m \times n}}$, fine-grained spatial matrix $F_{S_{m \times n}}$ and coarse-grained matrix $C_{M_{m \times n}}$ to construct the samples (Lines 1–7 of Algorithm 5). Then, the samples were divided into three parts: 80% as a training dataset, 10% as a test dataset and 10% as a cross-validation dataset to control early stopping (Lines 8–10 of Algorithm 5). Next, the error back-propagation algorithm was used to train the neural network model (Figure 7). Assuming that the input variables of the neural network model are $X = \{F_S, F_T\}$, the connection weights between the input layer and the hidden layer are γ_{ij} and the hidden layer bias value is $bias1$, then the output of the hidden layer is:

$$H_j = f\left(\sum_{i=1}^n \gamma_{ij} x_i - bias1\right) \quad j = 1, 2, \dots, l \quad (21)$$

where l is the number of hidden layer nodes and f is the activation function. Activation function f has many forms, from which we selected the sigmoid, which can be expressed as:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (22)$$

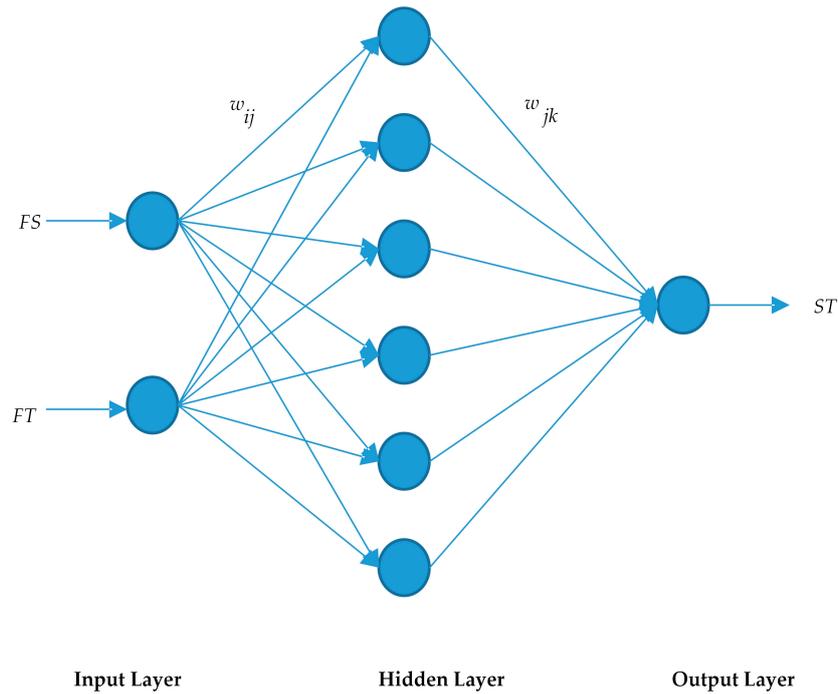


Figure 7. Neural network training.

The interpolation estimate of missing data ST was calculated as follows:

$$ST = \sum_{j=1}^l H_j \gamma_{j1} - bias2 \quad (23)$$

where γ_{j1} is the connection weight between the hidden layer and the output layer and $bias2$ is the bias of the output layer. The weights and bias of Formula (23) can be calculated from:

$$\gamma_{ij} = \gamma_{ij} + \eta H_j (1 - H_j) x(i) \gamma_{jl} e \quad i = 1, 2, \dots, n; j = 1, 2, \dots, l \quad \gamma_{jl} = \gamma_{jl} + \eta H_j e \quad j = 1, 2, \dots, l \quad (24)$$

$$bias1 = bias1 + \eta H_j (1 - H_j) w_{jl} \quad j = 1, 2, \dots, l \quad bias2 = bias2 + e \quad (25)$$

where η is the learning rate and e is the prediction error of the neural network (i.e., the difference between the predicted and expected outputs). The training process of the neural network was completed when the algorithm reached the set of training objectives (i.e., the number of iterations and the minimum error).

After completing the neural network model, we were then able to predict the missing data. We first detected missing data in the coarse-grained matrix $C_{M_m \times n}$. Then, the results of fine-grained spatial interpolation (FGSI) and fine-grained temporal interpolation (FGTI) algorithms were input to the neural network to calculate the estimated values (Lines 12–18 of Algorithm 5). Meanwhile, the results of FGSI and FGTI in the testing dataset were input to the neural network to evaluate the performance of the model (Lines 19–25 of Algorithm 5).

Algorithm 5: Combining spatial and temporal.

Input: Fine-Grained Spatial Matrix $F_S_{m \times n}$
 Fine-Grained Temporal Matrix $F_T_{m \times n}$
 Coarse-Grained Matrix $C_M_{m \times n}$
 Number of Spatial Neighbors ms
 Number of Temporal neighbors nt

Output: Test Estimated Matrix $ST_{m \times n}$
 Missing Estimated Matrix $M_ST_{m \times n}$

- 1 For $i = 1$ to m
- 2 For $j = 1$ to n
- 3 If F_S_{ij} F_T_{ij} C_M_{ij} are not missing values then
- 4 $Sample \leftarrow Sample(F_T_{ij}, F_S_{ij}, C_M_{ij})$
- 5 End if
- 6 End for
- 7 End for
- 8 $Training_Spl \leftarrow Divide(Sample, 0.8)$
- 9 $Testing_Spl \leftarrow Divide(Sample, 0.1)$
- 10 $CrossValidation_Spl \leftarrow Divide(Sample, 0.1)$
- 11 $Net \leftarrow Train(Training_Spl) \nabla$ Neural Network Training
- 12 For $i = 1$ to m
- 13 For $j = 1$ to n
- 14 If C_M_{ij} is missing value then
- 15 $M_ST_{ij} \leftarrow Sim(Net, F_T_{ij}, F_S_{ij})$
- 16 End if
- 17 End for
- 18 End for
- 19 For $i = 1$ to m
- 20 For $j = 1$ to n
- 21 If $\{F_S_{ij}, F_T_{ij}\} \in Testing_Spl$ then
- 22 $ST_{ij} \leftarrow Sim(Net, F_T_{ij}, F_S_{ij})$
- 23 End if
- 24 End for
- 25 End for

4. Results

4.1. Datasets

We evaluated our model based on real air quality datasets from Beijing, which were collected between 5 January 2014 and 30 April 2015. Data included PM_{2.5}, CO, SO₃, O₃, NO₂ and other attributes, each of which was collected at 36 air quality monitoring stations at hourly intervals, as depicted in Figure 8. The dataset contained a total of 8759 records [30] (Table 1).

Table 1. Experimental dataset.

Data	Ratio of Missing	Complete Case	Number of Missing
PM2.5	13.25%	29.11%	41,771
CO	15.10%	0.00%	47,604
SO ₃	15.24%	0.00%	48,041
O ₃	15.43%	0.00%	48,667
NO ₂	16.01%	0.00%	50,470

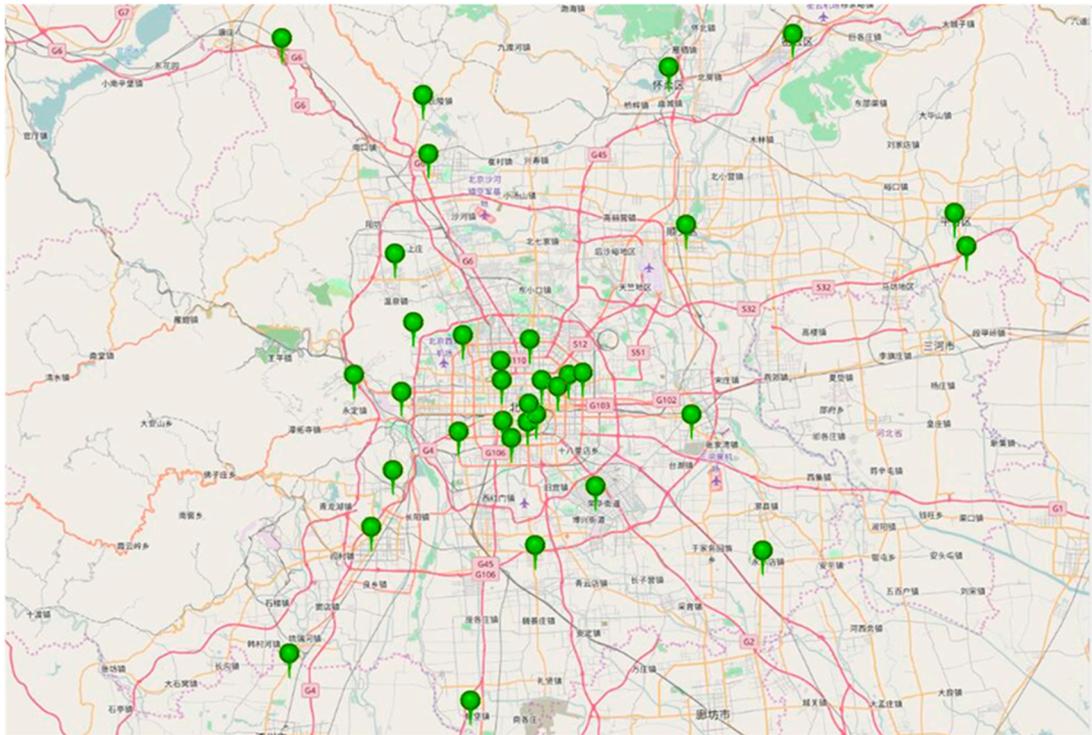


Figure 8. Air quality monitoring stations.

The different air quality variables have different degrees of missing data (Table 1). Among them, only PM_{2.5} contains a complete case. For the other attributes, we used combination analysis to explore the patterns of missing spatio-temporal missing data (Figure 9).



Figure 9. Pattern of missing spatio-temporal PM_{2.5} data. Red squares represent missing data, and values on the right ordinate are the number of missing data combinations.

Using the PM_{2.5} dataset as an example, we set $wt = 48$ as the time sliding window, which took 48 h of data to explore the missing pattern. The missing numbers of st_{19} , st_{27} , st_{35} and st_{36} at the same time were eight (i.e., the missing pattern in Figure 1c). Data for st_{19} , st_{27} , st_{36} were completely missing (Figure 1d). Data for $\{st_{18}, st_{19}\}$, $\{st_{35}, st_{36}\}$ showed random block loss (Figure 1b). Finally, a large number of patterns showed random missing data (Figure 1a). These patterns show that if the interpolation process was performed directly on the original dataset (i.e., without coarse-grained interpolation to eliminate the effect of successive missing data), it would be difficult to obtain accurate evaluation.

4.2. Evaluation Criteria

In order to evaluate the ST-2SMR method, we compared three existing methods (ST-kriging [31], P-BSHADE [16,32] and ST-HC [2,33]), each constrained in three different ways (i.e., increasing coarse-grained interpolation, increasing the sliding window or both; Table 2). We adopted mean absolute error (MAE), mean relative error (MRE) and the ratio of construction (RC) as the evaluation criteria to verify the performances of the proposed method.

Table 2. Combination of different methods. P-BSHADE, point estimation model of biased hospital-based area disease estimation; ST, spatio-temporal; HC, heterogeneous covariance.

Method	Coarse-Grained	Window	Coarse-Grained + Window
ST-kriging	ST-kriging-C	ST-kriging-W	ST-kriging-C-W
P-BSHADE	P-BSHADE-C	P-BSHADE-W	P-BSHADE-C-W
ST-HC	ST-HC-C	ST-HC-W	ST-HC-C-W

4.3. Experimental Results

The proposed method was implemented in MATLAB 2016b. The PM2.5 dataset was selected to validate the proposed method. Through the comparative analysis of different experiments, the relevant parameters were set as follows: $\alpha = 4$, $\beta = 0.85$, $wc = 14$, $ms = 10$, $nt = 10$, $\eta = 0.01$.

4.3.1. Overall Results

Among the first group of experiments (i.e., those using unaltered spatio-temporal interpolation methods), the ST-HC and ST-2SMR have the same reconstruction ratio, but different accuracy. The ST-2SMR method had the highest accuracy, reflecting in the effect of the nonlinear combination on the integration of spatial and temporal results (Table 3). The ST-HC and ST-2SMR have the same reconstruction ratio because they adopt the same spatio-temporal interpolation algorithm; however, the reconstruction rate was the lowest, which is owed to the introduction of heterogeneity in the time dimension. When we calculate the correlation coefficient and covariance between the spatial sequence of the missing data and the time slice, these sequences may seriously be missing. If the data sequence in the calculation is completely missing or using the pare-wise method makes the data sequence be completely missing, which may lead to the covariance matrix still having missing data, therefore we cannot get the final estimates, and this results in lower rate reconstruction.

In the second group of experiments (i.e., those with the original algorithms, but with the sliding window increased), interpolation precision was improved because the increased sliding window ensured that the sample data have the strongest correlation with missing data. In the third group of experiments (i.e., algorithms with coarse-grained interpolation added before the original interpolation method; here, we use IDW + SES as the coarse-grained interpolation method), the interpolation accuracy was also improved, and the complete reconstruction results were obtained because of the influence of continuous missing data has been eliminated. This result also validates the point of view in [34]: the accuracy and reliability of spatio-temporal interpolation methods depend on the pattern of missing data. However, interpolation accuracy improved the most when both constraints were applied at the same time, as demonstrated by the significant improvements in MAE and MSE (Table 3).

4.3.2. Effect of Coarse-Grained Interpolation

Through coarse-grained interpolation, continuous missing data were eliminated, significantly improving accuracy. Our experiments demonstrated that, regardless of the coarse-grained interpolation method chosen, interpolation accuracy improves when coarse-grained interpolation is first used (Table 4; here, we compare with the accuracy of those using unaltered spatio-temporal interpolation methods in Table 3). Among the methods tested, ST-2SMR showed the most significant improvement in accuracy, reflecting the nonlinear integration of spatial and temporal interpolation results, an approach

that is particularly suitable for describing complex relationships between spatial and temporal data. However, overall, the accuracy of the IDW + SES method was found to be the best, so this method was chosen for coarse-grained interpolation in subsequent experiments.

Table 3. Performance comparison of different methods. RC, ratio of construction; ST-2SMR, spatio-temporal missing data reconstruction.

Condition	Method	MAE	MSE	RC
	ST-Kriging	18.3796	0.2211	96.20%
	P-BSHADE	18.2085	0.2190	96.20%
	ST-HC	26.4273	0.3066	65.94%
	ST-2SMR	15.5247	0.1319	65.94%
Window	ST-Kriging-W	14.3281	0.1724	99.60%
	P-BSHADE-W	14.6172	0.1758	99.29%
	ST-HC-W	11.2211	0.1349	93.69%
	ST-2SMR	9.5920	0.0822	93.69%
Coarse-Grained	ST-Kriging-C	13.1726	0.1585	100%
	P-BSHADE-C	12.9178	0.1554	100%
	ST-HC-C	8.7650	0.1054	100%
	ST-2SMR	7.4292	0.0470	100%
Coarse-Grained + Window	ST-Kriging-C-W	12.9717	0.1560	100%
	P-BSHADE-C-W	12.6669	0.1524	100%
	ST-HC-C-W	7.9196	0.0953	100%
	ST-2SMR	7.2285	0.0623	100%

Table 4. Performance of different coarse-grained interpolation methods ¹. SES, simple exponential smoothing.

Method	IDW + SES		ST-Kriging		P-BSHADE	
	MAE	MRE	MAE	MRE	MAE	MRE
ST-kriging	13.1726	0.1585	13.6027	0.1636	13.6010	0.1636
P-BSHADE	12.9178	0.1554	13.5883	0.1635	13.6081	0.1637
ST-HC	8.7650	0.1054	9.0637	0.1089	9.1601	0.1101
ST-2SMR	7.4292	0.0470	7.4826	0.0475	7.6002	0.0484

¹ The abscissa represents fine-grained interpolation; the ordinate represents coarse-grained interpolation.

4.3.3. Effect of the Coarse-Grained Missing Data Rate

According to Algorithm 1, the results of coarse-grained interpolation are mainly affected by the decay rate of weight α , smoothing parameter β and by time threshold wc . According to the experimental results of [27], performed using the same PM2.5 dataset, IDW and SES achieve a minimum MAE value when $\alpha = 4$ and $\beta = 0.85$. For the time threshold, different values exert a significant influence on the coarse-grained interpolation results. In this study, we determined wc heuristically, with the value initially set to one (i.e., the center of missing data, taking the first hour and last hour as the sample data). With the increase in wc , the reconstruction rate of the missing data increased until complete reconstruction was achieved (Figure 10). We found that MAE was smallest when $wc = 1$ and stable when $wc > 3$. At $wc > 3$, the contribution weight of the sample data for missing data was nearly equal to zero, resulting in a small effect on the interpolation results. Furthermore, the termination condition of wc was set to ensure no continuous deletion of the whole space sequence and time series after coarse-grained interpolation, so as to obtain a complete interpolation result in the fine-grained interpolation. When $wc < 14$, the dataset from the coarse-grained interpolation still exhibited complete missing data in both the time and space sequences (Figure 11); therefore, fine-grained interpolation

was required to achieve reconstruction. When $wc = 14$, coarse-grained interpolation eliminated the influence of successive missing data, so the reconstruction rate was 100%.

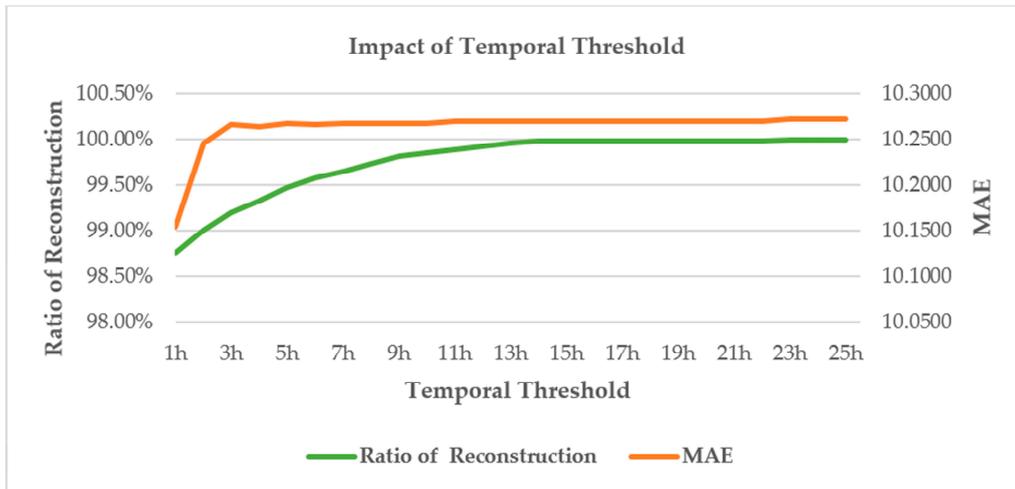


Figure 10. Impact of the temporal threshold on coarse-grained interpolation.

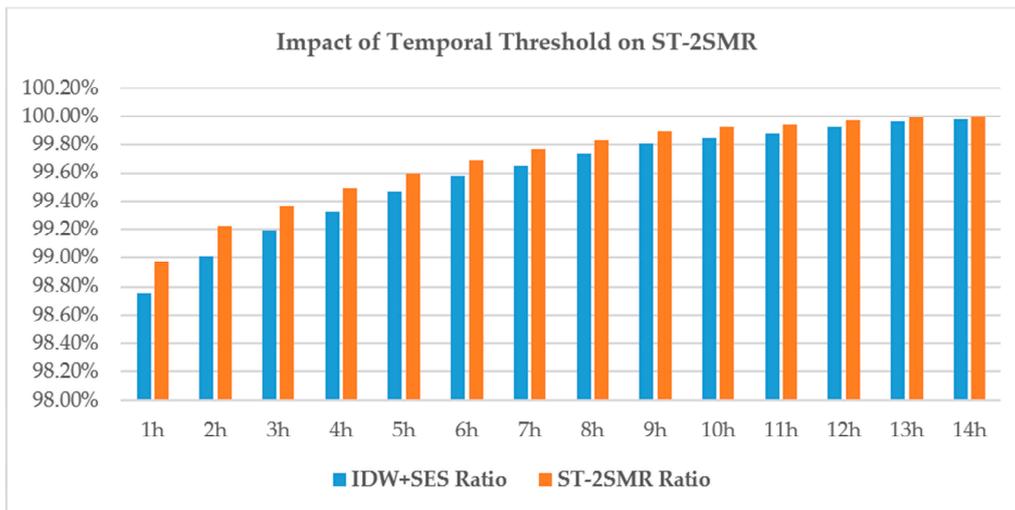


Figure 11. Impact of the temporal threshold on ST-2SMR.

4.3.4. Effect Sample Point Number

During fine-grained interpolation, we were required to select ms spatial neighbors and nt temporal neighbors for missing data. The number of sample points impacted the assessment results. Too few sample data points fail to reflect the correlation of spatial and temporal data, while excessive sample data increase computational complexity and also reduce the accuracy of assessment because of redundant data. According to the experimental results of [16], when the number of sampling points is set to between five and 15, the interpolation results are perfect; therefore, we set three sets of adjacent point selection patterns (5, 10 and 15) in both the spatial and temporal dimensions (i.e., a total of nine sets of experiments) and performed experiments to determine the most suitable number of samples. When $ms = 10$ and $nt = 10$, the ST-2SMR method achieved its best performance (Table 5); however, the results show that the number of neighbor points had no effect on the reconstruction rate. This result reflects the elimination of continuous missing data during coarse-grained interpolation, which

meant that all missing data had access to observed data for interpolation, allowing the whole dataset to obtain corresponding estimates.

Table 5. Influence sample point number on interpolation results.

Neighbor Station Number		MAE	MRE	RC
Spatial	Temporal			
5	5	7.3300	0.0625	100%
5	10	7.3276	0.0635	100%
5	15	7.4736	0.0643	100%
10	5	7.2787	0.0630	100%
10	10	7.2285	0.0623	100%
10	15	7.2761	0.0630	100%
15	5	7.2952	0.0631	100%
15	10	7.2892	0.0631	100%
15	15	7.3332	0.0650	100%

4.3.5. Effect of Sliding Window

Our experimental results show that the SOM algorithm can dynamically select the size of each window through the interaction information for time and space (Figure 12). As a result, interpolation accuracy was greatly improved.

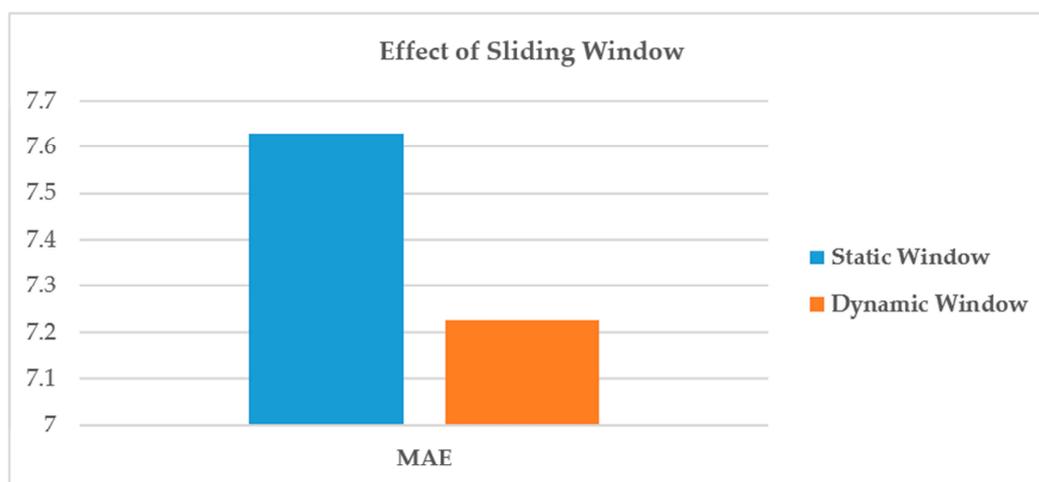


Figure 12. Influence of the sliding window on the interpolation results. The static window refers to the selection of fixed size sliding window (i.e., the center of missing data, taking the first 24 h and last 24 h as the sample data for interpolation).

4.3.6. Performance of Two- and Three-Step Interpolation

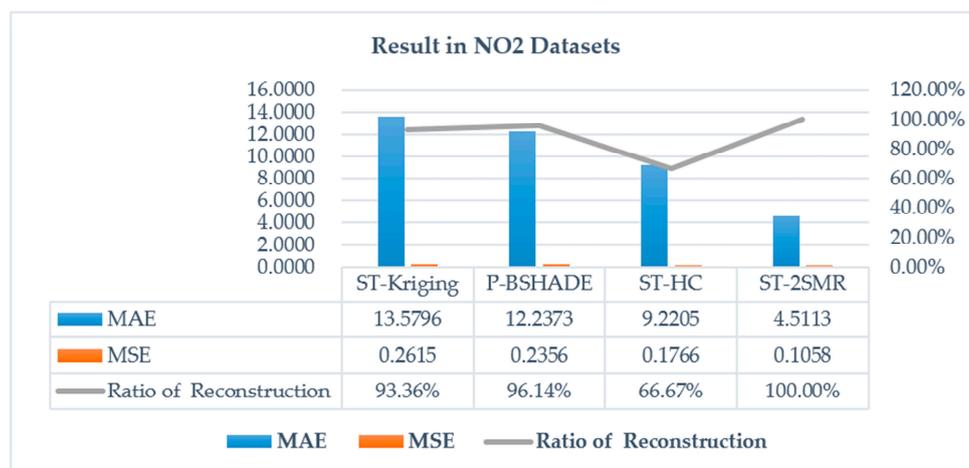
In order to explore the convergence of the two-step interpolation, we further introduced a third-step interpolation (Table 6). The two-step interpolation used IDW + SES for coarse interpolation, while the three-step interpolation was based on the results of two-step interpolation using ST-kriging, P-BSHADE and ST-HC. The results demonstrate that three-step interpolation slightly improved accuracy, but the change was marginal. In addition, regardless of the method used for the third interpolation, the overall results tended to be stable; therefore, we concluded that the additional computational complexity of the three-step method was not justified by the minor improvement in performance.

Table 6. Two- and three-step interpolation results.

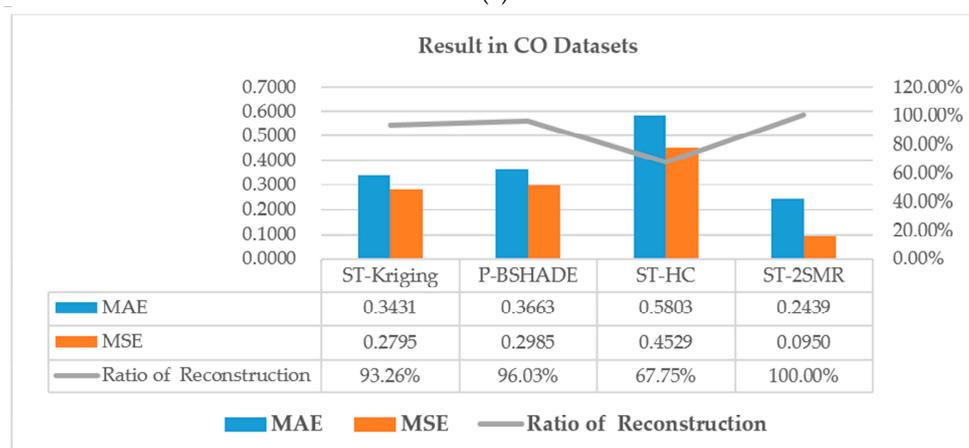
Method	Three-Step						Two-Step	
	ST-Kriging		P-BSHADE		ST-HC		IDW + SES	
	MAE	MRE	MAE	MRE	MAE	MRE	MAE	MRE
ST-Kriging	13.0498	0.1570	13.0235	0.1567	13.0741	0.1573	13.1726	0.1585
P-BSHADE	12.7898	0.1539	12.7964	0.1539	12.8304	0.1543	12.9178	0.1554
ST-HC	8.7075	0.1048	8.7206	0.1049	8.6618	0.1042	8.7650	0.1054
ST-2SMR	7.4143	0.0651	7.4215	0.0652	7.4080	0.0641	7.4292	0.0470

4.3.7. Performance Comparison for Different Datasets

To verify the universality of the proposed method, the approach was tested using the NO₂, CO, SO₃ and O₃ datasets (Figure 13). The results confirmed that the proposed method is superior to the other three methods in terms of accuracy. We found that only our new method can guarantee a complete reconstruction result and is able to maintain consistent stability across different datasets. For example, the P-BSHADE method performed better on the SO₃ dataset, but worse for other datasets. The ST-HC method performed better on the NO₂ dataset, but worse on the other datasets. This variable performance reflects the fact that different datasets have completely different missing data patterns, from which existing methods directly interpolate results (i.e., they do not eliminate the influence of missing patterns before interpolation).

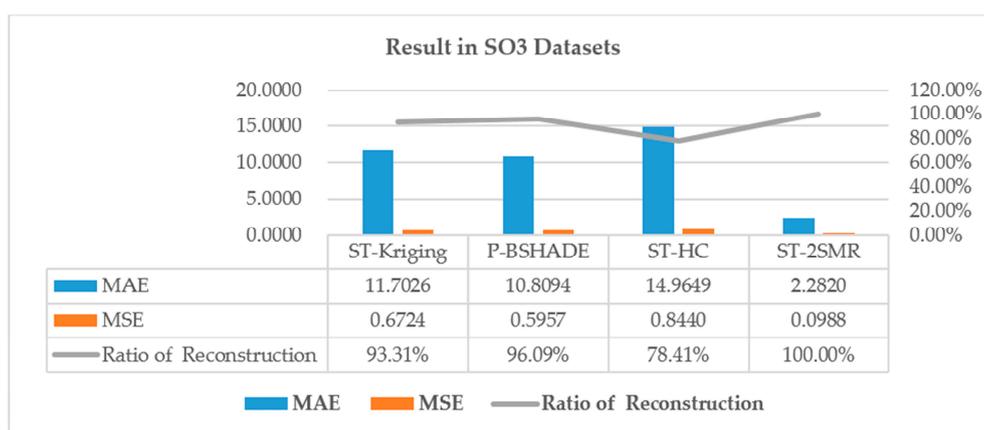


(a)

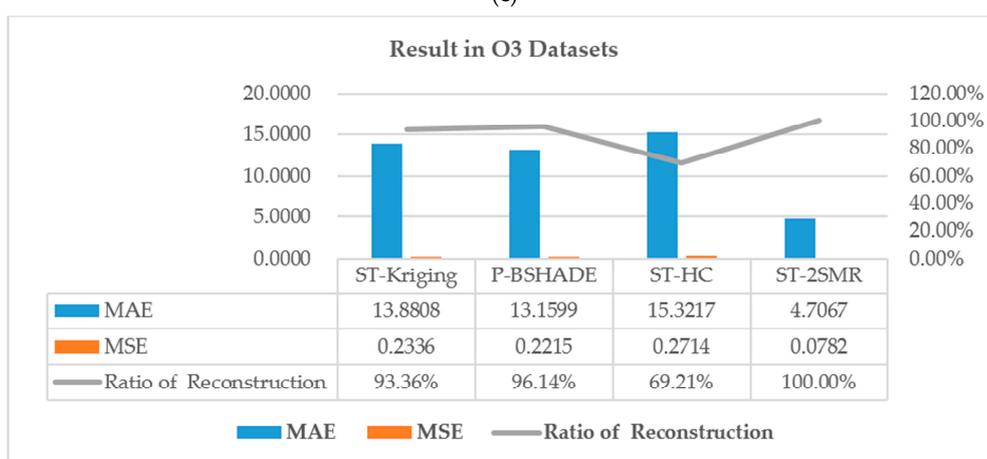


(b)

Figure 13. Cont.



(c)



(d)

Figure 13. Performance of the ST-2SMR using different spatio-temporal datasets. (a) The result of the experiment in NO₂ datasets; (b) The result of the experiment in CO datasets; (c) The result of the experiment in SO₃ datasets; (d) The result of the experiment in O₃ datasets .

4.3.8. Evaluation of Computational Efficiency

The computational efficiency is also an important factor worth evaluation for missing data reconstruction. We conducted a comparison of the computational efficiency on a 3.4-GHz Intel i7 CPU, with a 64-bit operating system, and 16.0 GM RAM. The CPU time costs of each interpolation method in the forecasting stages (i.e., we select 10% of samples as a test dataset) are shown in Figure 14. The computational efficiency of all of the four methods has no distinct difference. Obviously, ST-kriging is the fastest one among the interpolation methods because it does not take into account the effects of spatial and temporal heterogeneity on interpolation results. ST-HC and ST-2SMR consume only a little more time than ST-kriging and P-BSHADE because they are the extensions of P-BSHADE and consider the temporal and spatial heterogeneity. In addition, the linear combination of ST-HC to integrate spatio-temporal interpolation results almost spends the same time as nonlinear ways using a trained neural network, so the time complexity of ST-HC and ST-2SMR is nearly the same. However, the ST-2SMR method makes a significant trade-off between the efficiency and accuracy because the MSEs resulting from other methods are far larger than that of ST-2SMR (see Table 3). Therefore, when both efficiency and accuracy are considered, the proposed ST-2SMR outperforms the other methods.

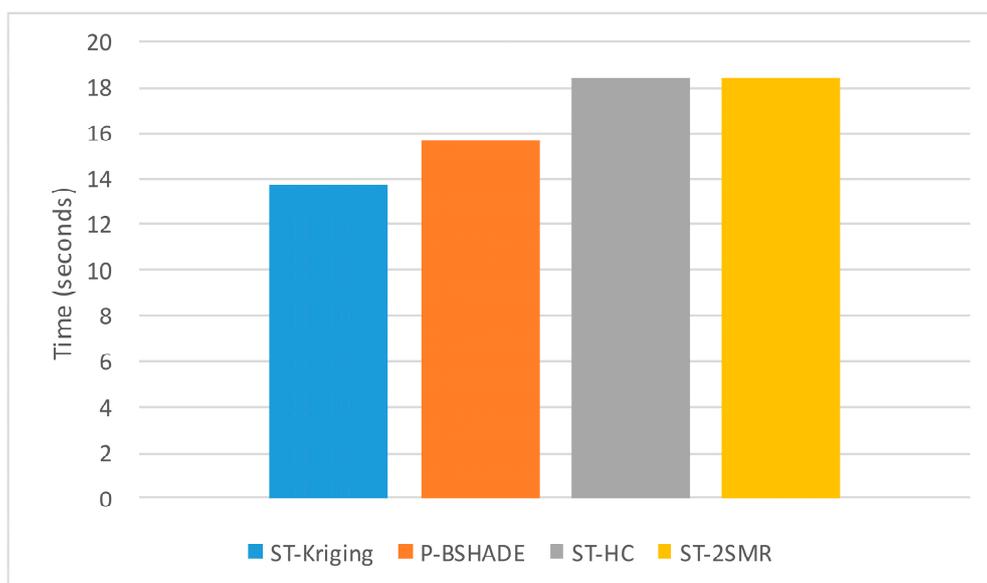


Figure 14. Comparison of the computational time costs for different interpolation methods.

5. Summary

Given the problem with existing missing data interpolation methods, this study developed a novel method called ST-2SMR. In the ST-2SMR method, missing data patterns in spatio-temporal datasets are first identified, and information on temporal and spatial dimensions is integrated to obtain a partial reconstruction. Using the output of this partial reconstruction, spatial and temporal heterogeneity is taken into account and a sliding window is set to both remove redundant sample data (i.e., to reduce computational complexity) and to ensure that the strongest correlation with missing data is selected (i.e., the most suitable data are chosen to improve the accuracy of the analysis). Finally, spatio-temporal interpolation results are integrated through a neural network model. We evaluated ST-2SMR using a real and open air quality dataset collected from Beijing. It is argued that the proposed method performs better than other existing methods. Providing the characteristics of the black box neural network models, the best way to integrate the results of spatio-temporal interpolation and to depict the nonlinear relationships between space and time requires further consideration.

Acknowledgments: This research is supported by the State Key Research Development Program of China (Grant No. 2016YFB0502104), the National Natural Science Foundation of China (41631177), and the Key Research Program of the Chinese Academy of Sciences (Grant No. ZDRW-ZS-2016-6-3). Their supports are gratefully acknowledged. And we also thank the anonymous referees for their helpful comments and suggestions.

Author Contributions: Shifen Cheng, Feng Lu conceived the idea for the research and wrote the paper; Shifen Cheng implemented the ST-2SMR model and carried out the experimental validation. Feng Lu interpreted the results and made the important comments and suggestions for this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, J.; Yong, G.E.; Lianfa, L.I. Spatiotemporal data analysis in geography. *Acta Geogr. Sin.* **2014**, *69*, 1326–1345.
2. Deng, M.; Fan, Z.; Liu, Q. A Hybrid Method for Interpolating Missing Data in Heterogeneous Spatio-Temporal Datasets. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 13. [[CrossRef](#)]
3. Gao, Z.; Cheng, W.; Qiu, X. A Missing Sensor Data Estimation Algorithm Based on Temporal and Spatial Correlation. *Int. J. Distrib. Sens. Netw.* **2015**, *2015*, 1–10. [[CrossRef](#)]
4. Galán, C.O.; Lasheras, F.S.; Juez, F.J.D.C.; Sánchez, A.B. Missing data imputation of questionnaires by means of genetic algorithms with different fitness functions. *J. Comput. Appl. Math.* **2017**, *311*, 704–717. [[CrossRef](#)]

5. Durán-Rosal, A.M.; Hervás-Martínez, C.; Tallón-Ballesteros, A.J. Massive missing data reconstruction in ocean buoys with evolutionary product unit neural networks. *Ocean Eng.* **2016**, *117*, 292–301. [[CrossRef](#)]
6. Tak, S.; Woo, S.; Yeo, H. Data-Driven Imputation Method for Traffic Data in Sectional Units of Road Links. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1762–1771. [[CrossRef](#)]
7. Tonini, F.; Dillon, W.W.; Money, E.S. Spatio-temporal reconstruction of missing forest microclimate measurements. *Agric. For. Meteorol.* **2016**, *218–219*, 1–10. [[CrossRef](#)]
8. Londhe, S.; Dixit, P.; Shah, S. Infilling of missing daily rainfall records using artificial neural network. *ISH J. Hydraul. Eng.* **2015**, *21*, 255–264. [[CrossRef](#)]
9. Tipton, J.; Hooten, M.; Goring, S. Reconstruction of spatio-temporal temperature from sparse historical records using robust probabilistic principal component regression. *Adv. Stat. Clim. Meteorol. Oceanogr.* **2017**, *3*, 1–16. [[CrossRef](#)]
10. Ruan, W.; Xu, P.; Sheng, Q.Z. *Recovering Missing Values from Corrupted Spatio-Temporal Sensory Data via Robust Low-Rank Tensor Completion*; International Conference on Database Systems for Advanced Applications; Springer: Cham, Switzerland, 2017.
11. Lu, G.Y.; Wong, D.W. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. [[CrossRef](#)]
12. Bartier, P.M.; Keller, C.P. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Comput. Geosci.* **1996**, *22*, 795–799. [[CrossRef](#)]
13. Pesquer, L.; Cortés, A.; Pons, X. Parallel ordinary kriging interpolation incorporating automatic variogram fitting. *Comput. Geosci.* **2011**, *37*, 464–473. [[CrossRef](#)]
14. Bhattacharjee, S.; Mitra, P.; Ghosh, S.K. Spatial Interpolation to Predict Missing Attributes in GIS Using Semantic Kriging. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4771–4780. [[CrossRef](#)]
15. Dutilleul, P. *Spatio-Temporal Heterogeneity: Concepts and Analyses*; Cambridge University Press: Cambridge, UK, 2011.
16. Xu, C.; Wang, J.; Hu, M. Interpolation of Missing Temperature Data at Meteorological Stations Using P-BSHADE. *J. Clim.* **2013**, *26*, 7452–7463. [[CrossRef](#)]
17. Yozgatligil, C.; Aslan, S.; Iyigun, C. Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. *Theor. Appl. Climatol.* **2013**, *112*, 143–167. [[CrossRef](#)]
18. Gardner, E.S., Jr. Exponential smoothing: The state of the art—Part II. *Int. J. Forecast.* **2006**, *22*, 637–666. [[CrossRef](#)]
19. Li, Y.; Li, Z. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. *Transp. Res. Part C Emerg. Technol.* **2013**, *34*, 108–120. [[CrossRef](#)]
20. Ran, B.; Tan, H.; Wu, Y. Tensor based missing traffic data completion with spatial-temporal correlation. *Phys. A Stat. Mech. Appl.* **2016**, *446*, 54–63. [[CrossRef](#)]
21. Qi, H.; Liu, M.; Wang, D. Spatial-Temporal Congestion Identification Based on Time Series Similarity Considering Missing Data. *PLoS ONE* **2016**, *11*, e162043. [[CrossRef](#)] [[PubMed](#)]
22. Holland, R.C.; Jones, G.; Benschop, J. Spatio-temporal modelling of disease incidence with missing covariate values. *Epidemiol. Infect.* **2015**, *143*, 1777–1788. [[CrossRef](#)] [[PubMed](#)]
23. Reynolds, K.M.; Madden, L.V. Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology* **1988**, *78*, 240–246. [[CrossRef](#)]
24. Li, D.; Deogun, J.; Spaulding, W. Towards Missing Data Imputation: A Study of Fuzzy K-means Clustering Method. *Lect. Notes Comput. Sci.* **2004**, *3066*, 573–579.
25. Qu, L.; Li, L.; Zhang, Y. PPCA-based missing data imputation for traffic flow volume: A systematical approach. *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 512–522.
26. Kong, L.; Xia, M.; Liu, X.Y. Data Loss and Reconstruction in Sensor Networks. *IEEE Infocom* **2013**, *25*, 1654–1662.
27. Yi, X.; Zheng, Y.; Zhang, J. ST-MVL: Filling Missing Values in Geo-Sensory Time Series Data. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016; pp. 2704–2710.
28. Karydas, C.G.; Gitas, I.Z.; Koutsogiannaki, E. Evaluation of spatial interpolation techniques for mapping agricultural topsoil properties in Crete. *EARSeL eProceedings* **2009**, *8*, 26–39.
29. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representation by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]

30. Zheng, Y.; Yi, X.; Li, M. Forecasting fine-grained air quality based on big data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015.
31. Cesare, L.D.; Myers, D.E.; Posa, D. Estimating and modeling space-time correlation structures. *Stat. Probab. Lett.* **2001**, *51*, 9–14. [[CrossRef](#)]
32. Wang, J.F.; Li, X.H.; Christakos, G. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [[CrossRef](#)]
33. Fan, Z.; Gong, G.; Liu, B. A Space-time Interpolation Method of Missing Data Based on Spatio-temporal Heterogeneity. *Acta Geod. Cartogr. Sin.* **2016**, *45*, 458–465.
34. Kondrashov, D.; Ghil, M. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Proc. Geophys.* **2006**, *13*, 151–159. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).