Research Article

Temperature Field Data Reconstruction Using the Sparse Low-Rank Matrix Completion Method

Shan Wang [ID],[1] Jianhui Hu,[1] Huiling Shan,[1] Chun-Xiang Shi,[2] and Weimin Huang [ID][3]

[1]*School of Information Engineering, East China Jiaotong University, Nanchang, China*
[2]*National Meteorological Information Center, Beijing, China*
[3]*Department of Electrical and Computer Engineering, Memorial University, St. John's, Canada*

Correspondence should be addressed to Shan Wang; patrick_shan@163.com

Due to limited number of weather stations and interruption of data collection, the temperature field data may be incomplete. In the past, spatial interpolation is usually used for filling the data gap. However, the interpolation method does not work well for the case of the large-scale data loss. Matrix completion has emerged very recently and provides a global optimization for temperature field data reconstruction. A recovery method is proposed for improving the accuracy of temperature field data by using sparse low-rank matrix completion (SLR-MC). The method is tested using continuous gridded data provided by ERA Interim and the station temperature data provided by Jiangxi Meteorological Bureau. Experimental results show that the average signal-to-noise ratio can be increased by 12.5%, and the average reconstruction error is reduced by 29.3% compared with the matrix completion (MC) method.

1. Introduction

Temperature field data are measured at a height of at least 1.5 m above the ground, which is an important parameter to describe the environmental conditions of the land [1] and widely utilized in weather forecast. The initial research on city thermal environment phenomenon mainly employed the temperature data from meteorological stations [2]. The number of meteorological stations is often limited. In addition, the data are continuous, but not in space. Based on the above data characteristics, it is challenging to investigate the temperature-related problem over large area. Sparse temperature field data are highly correlated with low spatial variability; interpolation is usually used to obtain the temperature data missed in a region. And so far, no research on the sparse property of continuous temperature data has been conducted.

With the rapid development of sparse representation, matrix completion (MC) method [3–5], which extends the idea of compressed sensing to matrices, has been proposed recently. Matrix completion aims to recover a corrupted matrix from a small part of its entries. It is impossible to recover a corrupted matrix without any assumptions about the matrix. Candès and Recht [3] found if the given matrix is low rank or approximately low rank, the missing entries of the corrupted matrix can be recovered through minimizing the matrix rank. Mathematically, for a corrupted matrix $\mathbf{M} \in \mathrm{R}^{n1 \times n2}$, the low-rank matrix completion problem usually can be formulated as

$$\begin{aligned} \text{minimize} \quad & \text{rank}(\mathbf{X}), \\ \text{subject to} \quad & P_{\boldsymbol{\Omega}}(\mathbf{X}) = P_{\boldsymbol{\Omega}}(\mathbf{M}), \end{aligned} \tag{1}$$

where $\text{rank}(\mathbf{X})$ denotes the rank of matrix $\mathbf{X}$, $\boldsymbol{\Omega}$ represents the locations of sampling in matrix $\mathbf{X}$ which is the number of known entries, and $P_{\boldsymbol{\Omega}}$ is the sampling operator which obtains only the entries indexed by $\boldsymbol{\Omega}$.

Unfortunately, matrix completion problem is NP-hard because the rank is nonconvex and discontinuous in reality. To solve the problem, Candès and Tao proposed convex nuclear norm to solve the rank minimization problem [6].

While the known entries are sampled randomly and uniformly from the unknown matrix, the missing entries can be recovered accurately if the matrix satisfies low-rank structure and incoherence condition [7]. Nuclear norm represents the sum of the singular values and can be seen as a special case of l_1 norm. It is easy to know that nuclear norm is widely adopted as a low-rank convex surrogate [8], which can be solved via the convex optimization. In order to solve the convex problem, semidefinite programming (SDP) has been proposed. Since SDP has high computational cost, several faster algorithms which are more computationally efficient than the SDP-based methods have been proposed, such as singular value thresholding (SVT) [9], singular value projection (SVP) [10], and inexact augmented Lagrangian method (IALM) [11].

Unlike the interpolation method presented in [12–14], matrix completion requires the corrupted matrix to be low rank, and it works well for the case when a large portion of data is lost. Taking advantage of the low rank and spatiotemporal correlation of a matrix, MC can achieve good interpolation performance. Compared with the traditional spatial interpolation method, MC takes good use of the correlation between the data, and it could only use a few temperature field data to reconstruct the global temperature field. The reconstructed data quality is comparable to the spatial interpolation. In order to obtain good reconstruction resolution, the temperature matrix needs to be low rank based on matrix completion theory. However, the temperature field data matrix does not have a stable rank and the rank of matrix varies with time. So, we regard the gridded temperature field data as a new matrix, whose rank is more stable. Although matrix completion can recover the incomplete temperature field data perfectly, some information will still be lost in the process. In [15], the data matrix was supposed to be decomposed into a low-rank part and a sparse part, and it can be recovered individually by solving a very convenient convex program under some suitable assumptions. To recover the sparse and low-rank components of a matrix efficiently, the alternating direction method (ADM) has been proposed in [16], but the sparse part of gridded temperature field data is well suited for the application of compressed sensing (CS) due to extensive spatiotemporal correlations that result in sparser representations. The combination of compressed sensing and low-rank matrix completion represents an attractive proposition for further improving reconstruction.

In this paper, a method based on matrix completion and compressed sensing [17, 18] is presented and referred to as sparse low-rank matrix completion (SLR-MC). Different from the method proposed in [16], the low-rank part and sparse part of corrupted matrix were recovered by matrix completion and compressed sensing individually. Firstly, the temperature field data matrix is decomposed into a low-rank or an approximately low-rank matrix and a sparse matrix. Then, the low-rank matrix is reconstructed using the matrix completion method, and the sparse part is recovered using compressed sensing. The method is tested using the gridded incomplete temperature field data provided by ERA Interim and the station temperature data provided by Jiangxi Meteorological Bureau.

The rest of this paper is organized as follows. The temperature field matrix model and the proposed SLR-MC method are described in Section 2. The experiment data and results are presented in Section 3, in which the performance comparison between the MC and SLR-MC reconstruction is also given. In Section 4, a summary of the work is provided.

2. Method

In this section, we will describe the temperature field matrix model which is decomposed into a low-rank part and a sparse part, then the fundamentals of matrix completion are introduced, and finally the SLR-MC method is presented.

2.1. The Temperature Field Matrix Model. In order to overcome the influence of rank, the gridded temperature field data at each time can be regarded as a new low-rank matrix. According to [19], the gridded temperature field data T_1, T_2, ..., T_L collected over a period can be arranged in rows to a large matrix T, as shown in Figure 1.

Assume that the size of each matrix is m × n, the rank of T_1 is r_1, the rank of T_2 is r_2, ..., and the rank of T_L is r_L. For each single temperature field, the observation matrix may not satisfy the low-rank property; therefore, the MC method cannot be directly used to reconstruct missing data or lost data. However, due to structure similarity and strong correlation among the matrices T_1, T_2, ..., T_L, the rank R of matrix T is smaller than max $(r_1, r_2, ..., r_L)$, and a matrix can be decomposed into two parts: a low-rank matrix T_M (few nonzero singular values) and a sparse matrix T_S (few nonzero entries):

$$T = T_M + T_S, \tag{2}$$

where rank $(T_M) \ll \min (m, n)$ and sparsity $(T_S) \ll mn$.

Figure 2 illustrates an example of the decomposition result.

2.2. Fundamentals of Matrix Completion. Matrix completion is the technique of completing missing values of a matrix with a subset of entries selected randomly and uniformly from a low-rank matrix or an approximately low-rank matrix [3, 15]. The incomplete matrix M can be recovered by solving the following rank minimization problem [3]:

$$\begin{aligned} &\text{minimize} \quad \text{rank}(X), \\ &\text{subject to} \quad P_\Omega(X) = P_\Omega(M), \end{aligned} \tag{3}$$

where rank (X) denotes the rank of a matrix X, and the sampling operator $P_\Omega: R^{m \times n} \longrightarrow R^{m \times n}$ is defined as follows:

$$P_\Omega(X) = \begin{cases} X_{ij}, & (i, j) \in \Omega, \\ 0, & (i, j) \in \Omega'. \end{cases} \tag{4}$$

We use $|\Omega|$ to represent the cardinality of Ω which is the number of known entries. For example, suppose the matrix X is
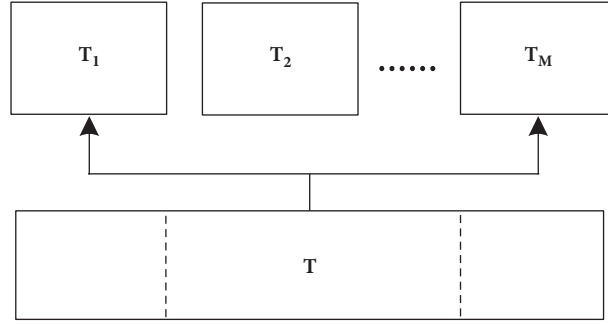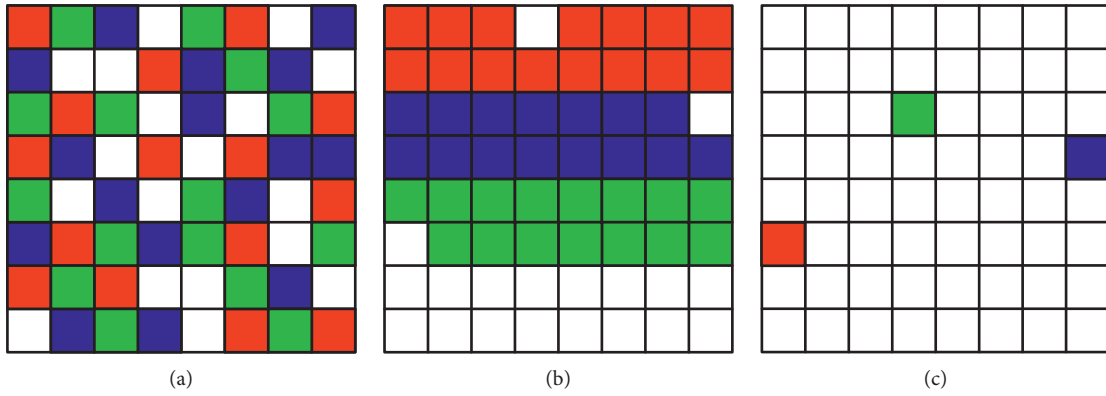
FIGURE 1: Temperature field matrix model.

FIGURE 2: (a) Temperature field matrix **T**, (b) low-rank matrix, and (c) sparse matrix.

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}. \tag{5}$$

If we have three elements known as $\boldsymbol{\Omega} = \{(1, 2), (2, 2), (2, 3)\}$, we can have

$$\mathbf{X}_{\boldsymbol{\Omega}} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 5 & 6 \end{bmatrix}. \tag{6}$$

However, the problem in equation (3) is NP-hard and impossible in practice. Candès and Recht proposed a nuclear norm minimization model to solve the following rank minimization model:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{X}_*\|, \\ \text{subject to} \quad & P_{\boldsymbol{\Omega}}(\mathbf{X}) = P_{\boldsymbol{\Omega}}(\mathbf{M}), \end{aligned} \tag{7}$$

where the nuclear norm $\|\mathbf{X}_*\|$ is the summation of the singular values of $\mathbf{X}$.

Unfortunately, we cannot recover any low-rank matrix (even its rank is 1) if the sampling entries in any row or column are completely missing. Suppose a matrix is of rank 1 and we do not have samples from the second column, the matrix cannot be recovered because no one can obtain all the exact entries of the second column using any method. In order to recover an unknown matrix, at least one observation in each row and column should be available. Candès and Recht [3] proved that if $\boldsymbol{\Omega}$ is sampled uniformly and randomly among all subset of cardinality m, we can solve the

problem (7) with high confidence where the number of samples should obey $m \geq Cn^{6/5}r \log n$.

In order to recover the incomplete matrix exactly, there is a restriction on the range of rank r. The selection of rank has a great influence on recovering low-rank matrix, and we use a small range of rank values and choose the value that results in the best performance (in Section 3, the rank is selected as 7).

2.3. Proposed Method. Considering model (2), the low-rank part $\mathbf{T_M}$ and sparse part $\mathbf{T_S}$ from the corrupted matrix $\mathbf{T}$ were supposed to be recovered. According to [20], a low-rank matrix or an approximate low-rank matrix can be reconstructed using the MC method. As shown in [21], a sparse matrix can be recovered with compressed sensing. Therefore, $\mathbf{T_M}$ and $\mathbf{T_S}$ in (2) can be obtained through

$$\begin{aligned} \min \quad & \text{rank}(\mathbf{T_M}) + \lambda\|\mathbf{T_S}\|_0, \\ \text{s.t.} \quad & \mathbf{T} = \mathbf{T_M} + \mathbf{T_S}. \end{aligned} \tag{8}$$

Problem (8) is a nonconvex optimization problem, where $\|\cdot\|_0$ denotes the number of nonzero value, and λ is a tuning weight that balances the contribution of the l_0-norm term relative to the rank minimization term and should be greater than 0. Problem (8) is extremely difficult to calculate and NP-hard, so it can be converted to the following convex optimization problem:

$$\min \quad \|\mathbf{T_M}\|_* + \lambda \|\mathbf{T_S}\|_0,$$
$$\text{s.t.} \quad \mathbf{T} = \mathbf{T_M} + \mathbf{T_S}, \tag{9}$$

where $\|\mathbf{T_M}\|_* = \sum_i \sigma_i(\mathbf{T_M})$ is the nuclear norm of $\mathbf{T_M}$ and $\sigma_i(\mathbf{T_M})$ represents the ith singular value of $\mathbf{T_M}$ (sorted in decreasing order). Problem (9) is also known as principal component pursuit (PCP), which can be solved by the augmented Lagrange multiplier (ALM) algorithm given in the following equation:

$$L(A, E, Y) = \|A\|_* + \lambda 1 \|E\|_0 + \langle Y, D - A - E \rangle$$
$$+ \frac{\mu}{2}\|D - A - E\|_F^2, \tag{10}$$

where μ is a positive scalar, $\lambda 1$ is a positive weighting parameter, the Lagrange multiplier Y is introduced to remove the equality constraint and $A + E = D$, $\|\cdot\|_F$ denotes the Frobenius norm $\|\cdot\|_F = \sqrt{\sum_i^m \sum_j^n \mathbf{X}_{ij}}$, and $\langle ., .\rangle$ represents the inner product operator. For a given Y, A and E are determined as the values that make $L(A, E, Y)$ reach the minimum. So, it is supposed that $\mathbf{T_M}$ can be recovered by problem (10). Different from the method proposed in [15, 16], to solve the sparse and low-rank matrix decomposition, the sparse part $\mathbf{T_S}$ was obtained by the compressed sensing method. This method represents a combination of augmented Lagrange multiplier used for matrix completion and compressed sensing used for sparse reconstruction.

3. Experimental Results

3.1. Gridded Temperature Field Data. We implemented our algorithms in MATLAB 2016. The experimental temperature field data used for testing the method are provided by ERA Interim of ECMWF (European Centre for Medium-Range Weather Forecasts) which can be obtained from the following website: https://apps.ecmwf.int/datasets. The data were collected from Asia at 00 am, 06 am, 12 pm, and 18 pm on January 1, 2014, at a height of 2 m above the ground, and the grid resolution was 0.75 degrees. The region of the study is at 20°E~160°W and 60°S~60°N, and the size of the region is 200×200.

Figure 3 shows the gridded temperature field data selected at 06 am on January 1, 2014, which is represented by a matrix of size 200×200. The value of the temperature is from 210 K to 320 K and the grid resolution is 0.75 degree. Figure 4 shows the sampled data of the global temperature at 06 am on January 1, 2014 (the sampling number is 15680), to which the reconstruction methods are applied.

4. Results

Both the MC method and the proposed algorithm are tested in this section. It can be seen from Figure 5(b) that the recovered global temperature field at 06 am on January 1, 2014, using the MC method agrees well with the original one
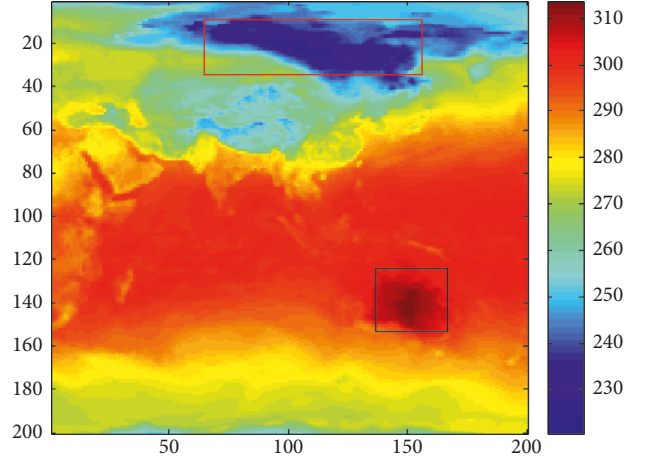

FIGURE 3: The data of gridded temperature field (blue color indicates the low temperature and red color indicates the high temperature).
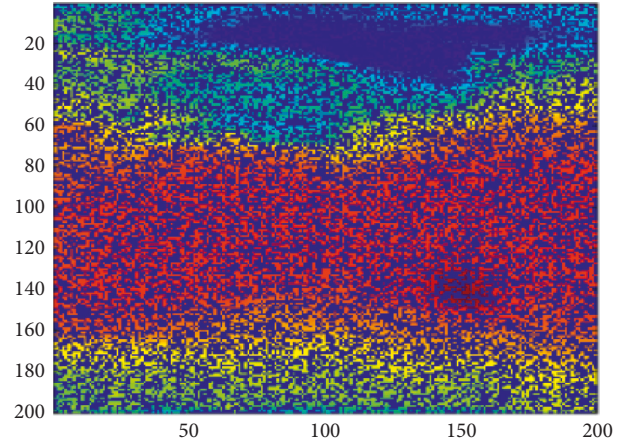

FIGURE 4: Sample data of gridded temperature field.

(the rank of Figure 5(a) can be estimated by LMaFit [22] and is selected as 7 in this section).

The results at high latitudes near the North Pole and low latitude areas near the equator are both satisfactory. Although the recovered global temperature field from the low-rank matrix using the MC method is good, the recovery results in some areas are not very satisfactory since the local temperature data property is not considered. For example, red rectangle in Figure 3 presents circumpolar latitude area, and its temperature varies from 220 K to 250 K. The corresponding recovery results are from 230 K to 250 K. In other words, the temperature data lower than 230 K have not been recovered successfully. Thus, the data in the red rectangle are selected for further analysis.

Table 1 ((a) and (b)) shows the original and recovery temperature field in the red rectangle on January 1, 2014, at 06 am, respectively. Comparison of the value at same position in Table 1 ((a) and (b)) shows that the difference is about 5 K to 9 K. The analysis results for low latitude area in the black rectangle shows similar performance in Figure 3.
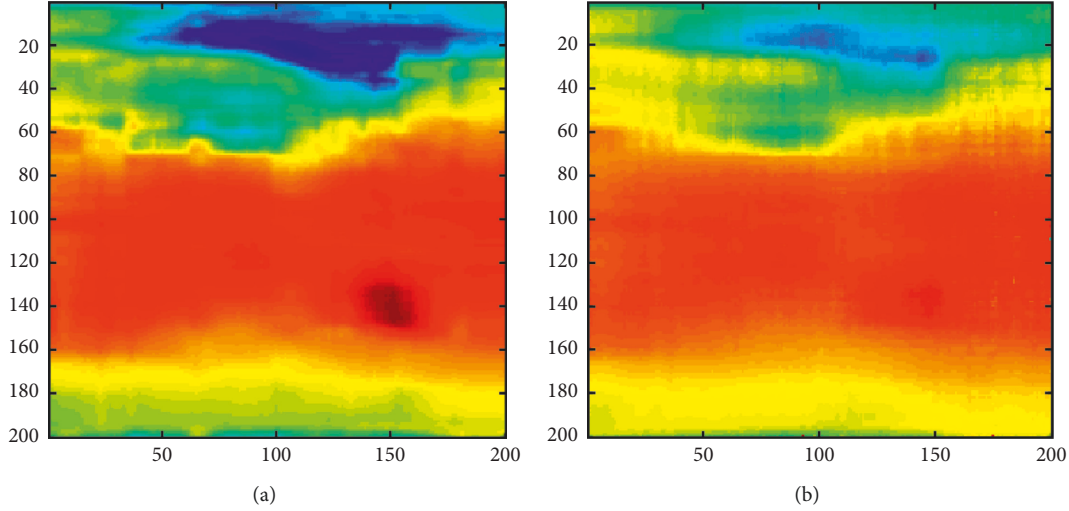
(a)   (b)

FIGURE 5: Reconstruction using the MC method. (a) Low-rank temperature field data ($r = 7$); (b) recovered field.

The true temperature data range from 300 K and 310 K. However, the recovery results using the MC method are less than 307 K, i.e., the temperature data higher than 307 K have not been recovered. From Table 2 ((a) and (b)), it can been seen that the recovered temperature field data are all lower than corresponding original data with an average temperature difference of 4 K. The above test shows that the performance of the MC method using low-rank matrix alone is not ideal. As mentioned earlier, the SLR-MC method can improve the reconstruction performance of global temperature field. The temperature field data collected at different times were used to test the proposed method. The original gridded temperature data (see Figure 6) at four moments (00 am, 06 am, 12 pm, and 18 pm) on January 1, 2014, were studied. The sampling number is 15680, and matrix rank is 7. The same data shown in Figure 3 (i.e., 06 am in Figure 6) were studied first. For the high latitude region in the red rectangle, the recovered temperature using the proposed method varies from 225 K to 250 K.

The point-to-point comparison is shown in Table 1 ((a) and (c)). It can be seen that the temperature difference is reduced from 7 K to 3 K, which is smaller than that in Table 1 (b). Using the SLR-MC method, the reconstruction error can be reduced significantly, which means the recovered temperature field is closer to the original one. Similarly, it is also found that the SLR-MC method can recover temperature field data higher than 307 K (in the black rectangle).

As illustrated in Table 2 ((a) and (c)), the recovered and original temperature field data at 06 am on January 1, 2014, were very close to each other. The average error was 1 K and less than that of MC. It can be concluded that the reconstruction results using SLR-MC are more accurate. For the regions with large temperature variation, the recovery performance is more satisfactory.

In this work, both reconstruction error (RE) and signal-to-noise ratio (SNR) are used to evaluate the recovery performance of the two methods. The RE is defined as follows:

TABLE 1: Comparison of original data and reconstruction data in red rectangle in Figure 3.

Latitude (N)	Longitude (E)					
	95.25°	96.00°	96.75°	97.50°	98.25°	99.00°
(a) Original data						
65.25°	223.1	222.0	222.0	222.0	223.4	224.8
64.50°	226.8	225.3	224.8	224.2	225.3	226.3
63.75°	232.1	231.0	230.2	229.4	229.1	228.8
63.00°	236.7	235.5	234.3	233.0	231.6	230.2
62.25°	240.5	238.9	237.4	236.1	234.9	233.3
61.50°	245.0	243.1	241.6	240.2	239.1	237.8
60.75°	250.6	248.3	246.3	244.4	242.9	241.4
60.00°	255.0	252.2	250.0	248.0	246.3	244.6
(b) Reconstruction results using MC						
65.25°	230.1	230.1	232.5	227.8	230.9	219.4
64.50°	231.8	232.4	230.7	229.4	230.3	192.6
63.75°	235.8	235.3	245.0	229.0	240.1	238.4
63.00°	239.3	239.9	237.1	235.9	235.4	191.7
62.25°	244.3	244.1	246.1	238.6	241.5	224.0
61.50°	247.2	248.9	231.8	242.8	233.9	142.4
60.75°	252.9	253.6	249.0	247.0	247.0	190.3
60.00°	257.5	258.1	251.7	252.6	249.2	195.2
(c) Reconstruction results using SLR-MC						
65.25°	224.9	224.3	224.3	224.8	225.1	229.3
64.50°	227.2	227.4	225.9	226.1	226.2	228.4
63.75°	230.6	231.9	231.7	232.0	227.3	232.0
63.00°	237.6	237.4	235.8	235.6	233.4	233.2
62.25°	242.2	242.1	238.4	235.9	234.6	234.4
61.50°	248.0	246.1	245.5	242.6	238.5	239.0
60.75°	253.2	250.7	248.5	246.8	242.6	238.9
60.00°	257.5	255.1	251.9	249.9	245.4	245.6

$$\text{RE} = \frac{\text{norm}\left(T - T_r\right)}{\text{norm}\left(T\right)}, \tag{11}$$

where T is the original temperature field data, T_r is the reconstructed data, and norm represents the 2-norm. The SNR is defined as

TABLE 2: Comparison of original data and reconstruction data in black rectangle in Figure 3.

| Latitude (S)	Longitude (E)			
	134.25°	135.00°	135.75°	136.50°
(a) Original data				
17.25°	306.9	306.5	306.5	306.2
18.00°	307.8	307.6	307.6	307.4
18.75°	309.1	309.1	308.9	308.8
19.50°	309.9	310.2	310.1	310.1
20.25°	310.3	310.7	311.1	311.5
21.00°	310.4	311.2	312.0	312.5
(b) Reconstruction results using MC				
17.25°	305.2	305.8	304.7	305.6
18.00°	294.3	305.6	293.5	308.8
18.75°	310.2	306.6	325.6	305.8
19.50°	315.4	309.7	312.0	308.4
20.25°	311.9	309.9	260.5	307.7
21.00°	307.2	309.1	309.2	308.6
(c) Reconstruction results using SLR-MC				
17.25°	305.8	304.3	304.7	303.9
18.00°	305.6	305.3	305.4	305.1
18.75°	306.2	309.4	308.2	308.3
19.50°	309.4	310.1	308.7	308.3
20.25°	309.7	309.7	309.7	309.6
21.00°	308.6	308.1	309.0	309.2
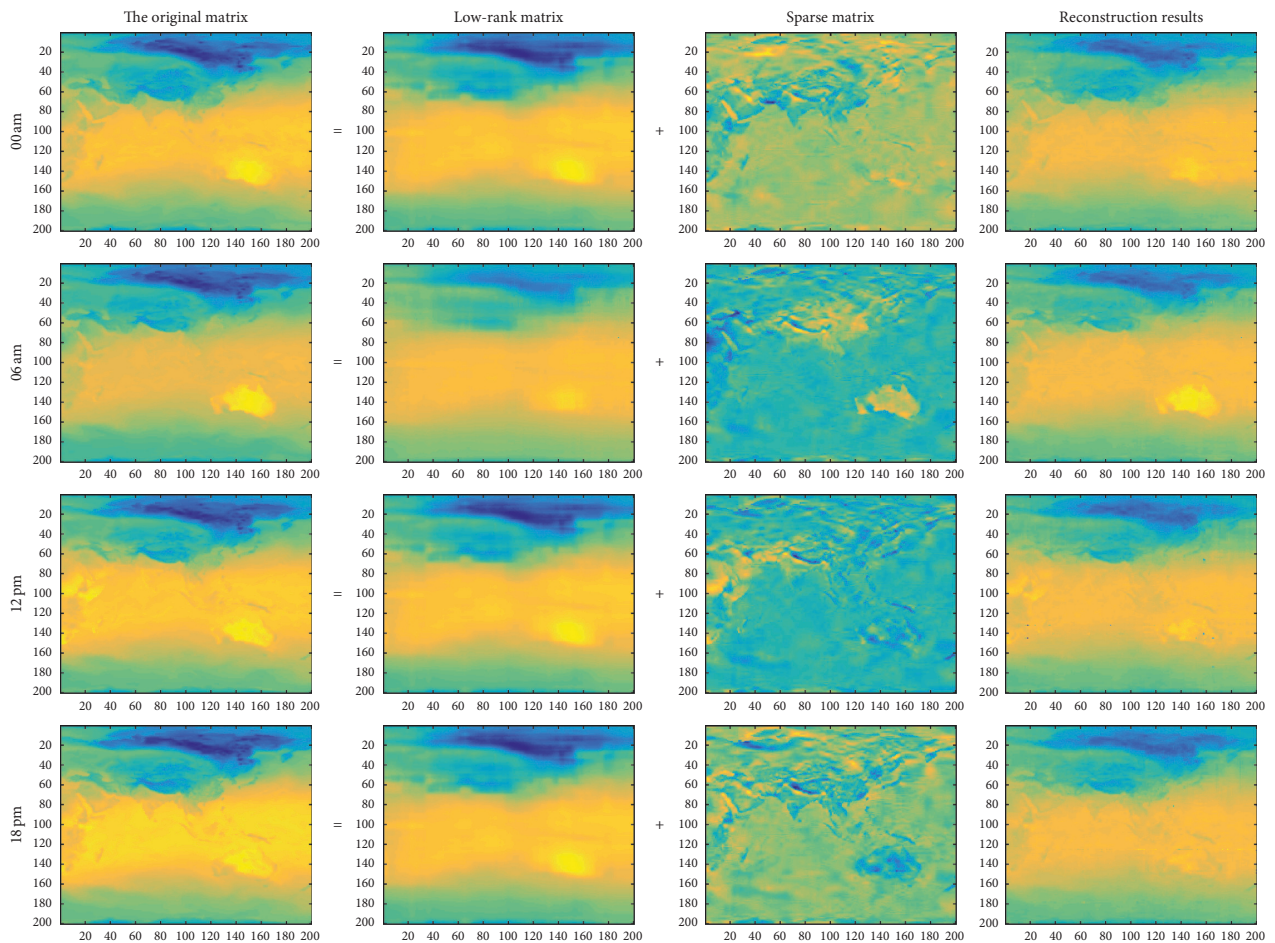
FIGURE 6: The reconstruction results of SLR-MC method.

$$SNR = 10 \log\left(\frac{p1}{p2}\right), \tag{12}$$

where $p1 = 1/[\text{length}(T) * \text{norm}(T)^2]$, $p2 = 1/[\text{length}(T) * \text{norm}(T - T_r)^2]$.

From Figures 7 and 8, it can be found that the RE and SNR of the SLR-MC method are lower and slightly higher, respectively, than those of the MC method for which the details are provided in Table 3. The average SNR for the four moments is increased by 12.5% using the proposed method, while the average error is 29.3% lower.

4.1. Station Temperature Data. In this section, we evaluate the performance of the SLR-MC method on the reconstruction of station temperature data. In this experiment, we have collected the temperature data at 92 national weather stations in Jiangxi, China. Figure 9 shows the longitude and latitude of stations, where the blue points represent the location of the national weather stations in Jiangxi. Each station reports its temperature data once a day to the monitoring center, and we have downloaded the data from January 2017 to March 2017. We put each station data into a vector and arrange the vectors into a large matrix. The data matrix has been set as $M = 87$ (only the data from 87 stations are used) and $T = 90$ (which represents the length from January 2017 to March 2017).

As shown in Figure 10, the row number of the temperature matrix represents the time from January 2017 to March 2017 and the column number represents the locations of 87 stations. Figure 11 shows the sampled temperature data matrix which is selected randomly and uniformly (the sampling number is 4698), and the blue dots represent the corrupted temperature data and the red dots represent the sampled temperature data. The size of matrix is 87×90, and the value of the temperature is from 0 K to 300 K. The reconstructed temperature data are shown in Figure 12. As shown in Figure 12, both MC and SLR-MC methods can capture the main feature of the original temperature data matrix.

The recovery results of the SLR-MC method can capture the local feature of original matrix and more key variation details, while the MC method often loses the information. The SLR-MC method may not have a significant improvement compared with the MC method because the changed temperature values only occupy a small portion of all temperature values in the matrix. Thus, the data in the white rectangle in Figure 12 are selected for further analysis.

The white rectangle in Figure 12 represents an area with significant temperature variation from 270 K to 290 K. Table 4 shows the original and reconstructed temperature data in the white rectangle. The white rectangle size is 8×6, which indicates the data matrix obtained by 8 stations (see Table 4 (a)) from time slots 30 to 35. Comparison of the value at same position in Table 4 ((a), (b), and (c)) shows that the difference is about 1 K to 7 K, which means both MC and SLR-MC methods can capture most information of the original
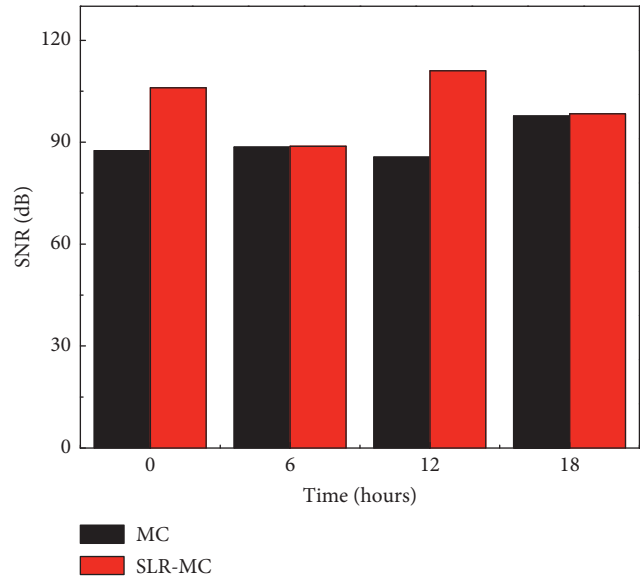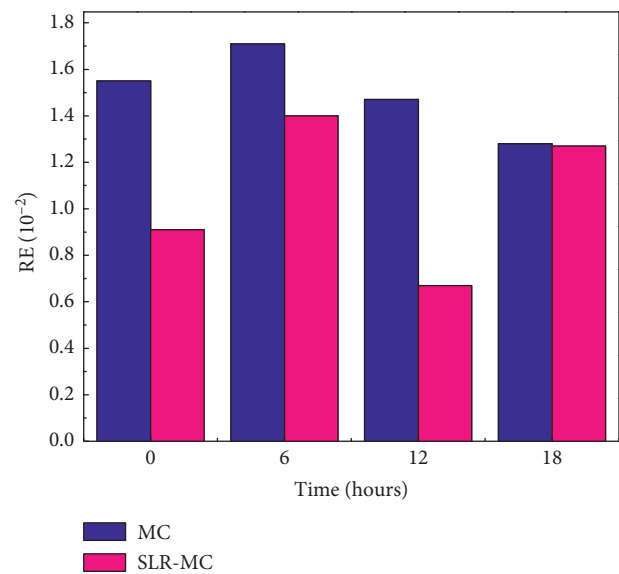

Figure 7: Comparison of SNR.

Figure 8: Comparison of RE.

Table 3: Comparison of the reconfiguration effect of MC and SLR-MC.

Prerequisite	Method	Time	RE (10^{-2})	SNR (dB)
		0	1.55	87.5
	MC	6	1.71	88.6
		12	1.47	85.6
Sampling = 15680		18	1.28	97.7
Rank = 7		0	0.91	106
	SLR-MC	6	1.40	88.8
		12	0.67	111
		18	1.27	98.4

matrix. Compared to MC, the data matrix in Table 4 (c) recovered by the SLR-MC method is closer to the original data matrix in Table 4 (a). For example, the reconstruction
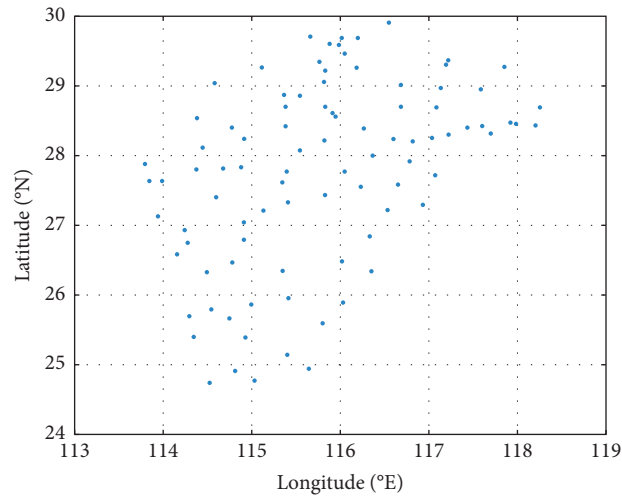
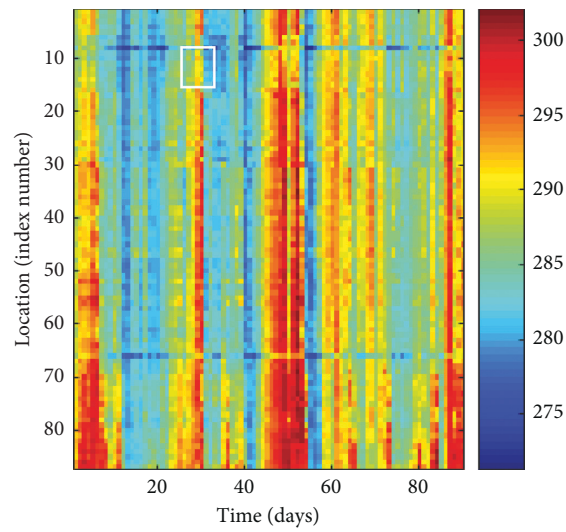FIGURE 9: The longitude and latitude of 92 weather stations in Jiangxi province.

FIGURE 10: The temperature data collected by 87 weather stations in Jiangxi from January 2017 to March 2017.
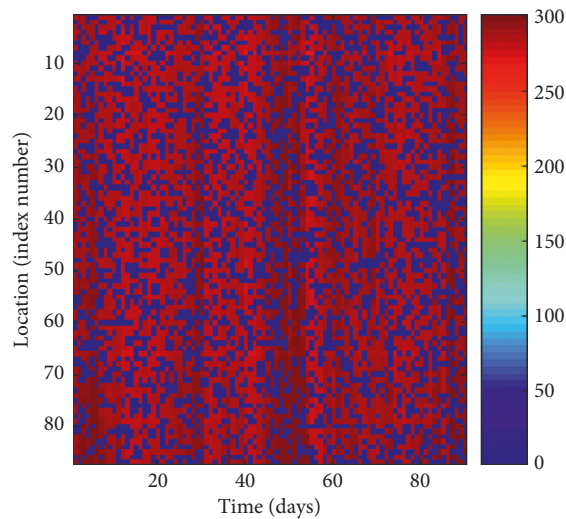

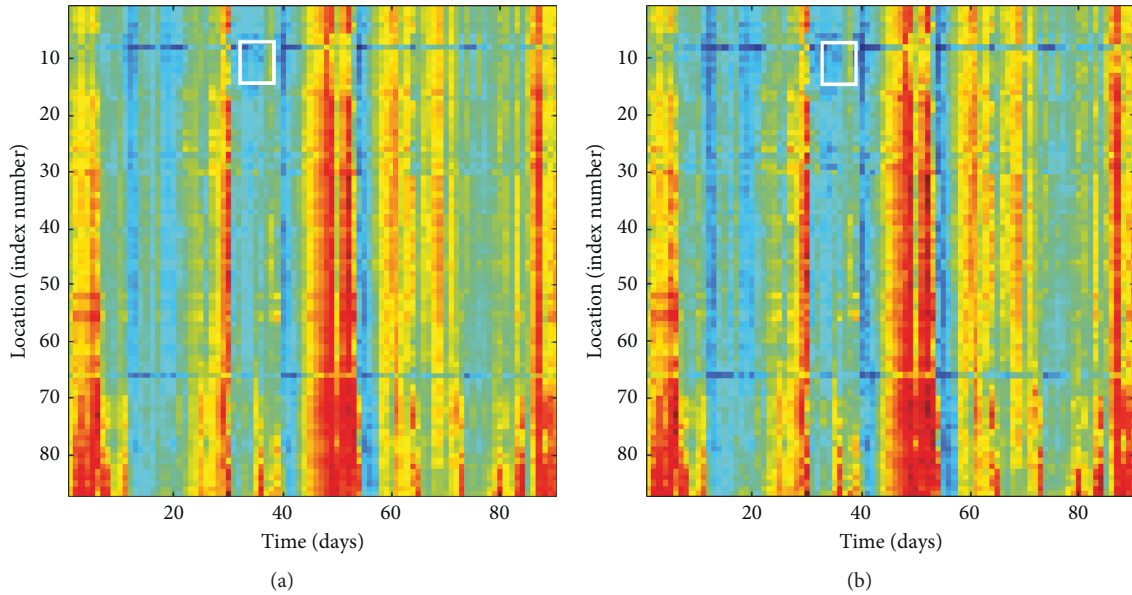FIGURE 11: Sampled temperature data matrix.

Figure 12: Comparison of corrupted temperature data matrix reconstruction. (a) Recovered by MC method; (b) recovered by SLR-MC method.

Table 4: Comparison of original data and reconstruction data in white rectangle in Figure 12.

Location	Time slot					
	30	31	32	33	34	35
(a) Original temperature data						
58503	288.05	274.65	279.95	273.55	278.95	280.15
58506	285.45	279.95	279.55	279.05	279.95	279.15
58510	288.75	279.65	279.35	279.05	280.05	278.55
58512	289.55	281.15	280.05	279.55	280.05	279.75
58514	290.55	280.05	280.35	279.25	280.45	280.05
58517	288.95	281.95	281.05	280.35	280.65	279.75
58508	290.85	281.15	280.65	280.85	280.95	280.35
58509	287.45	282.35	282.05	279.85	280.25	279.75
(b) Reconstruction results using MC						
58503	282.18	271.40	273.40	272.61	276.22	278.05
58506	282.01	282.91	276.68	279.33	280.90	276.27
58510	282.84	283.01	277.00	279.36	280.32	276.36
58512	282.70	282.84	277.23	279.63	281.69	277.05
58514	284.74	281.44	277.94	278.83	281.29	276.85
58517	284.16	283.17	278.15	279.92	281.99	277.01
58508	287.34	283.05	279.62	280.15	281.09	278.19
58509	283.07	283.93	277.42	279.73	280.84	275.36
(c) Reconstruction results using SLR-MC						
58503	283.68	272.23	276.00	272.41	278.98	279.46
58506	281.83	282.89	276.76	279.51	281.12	276.77
58510	286.69	282.98	276.61	279.63	280.59	276.57
58512	283.26	282.97	277.91	279.78	281.27	277.77
58514	284.90	281.13	277.45	278.57	281.73	277.22
58517	284.04	283.12	278.88	280.17	281.58	277.04
58508	287.59	282.61	279.02	280.31	281.22	278.38
58509	281.24	283.85	278.87	279.51	280.48	275.51

error (RE) between Table 4 (a) and Table 4 (b) is $1.08E(-2)$ while that between Table 4 (a) and Table 4 (c) is $9.43E(-3)$. The above test shows that the performance of the SLR-MC method is better than the MC method.

5. Conclusion

In this paper, the MC and SLR-MC methods were examined to determine which technique is appropriate for retrieving missing temperature data. Instead of using the alternating direction method (ADM) proposed in [16] to recover original corrupted matrix data, the SLR-MC method separates the clean low-rank matrix from the corrupted data effectively and applies matrix completion to fully exploit the low-rank features of temperature field data. The sparse matrix is reconstructed using compressed sensing to fully capture the sparse features of temperature field data. We have demonstrated the better performance of the SLR-MC method on gridded temperature field data and point temperature data from corrupted observations. Experimental results from gridded temperature field data confirm that the average SNR is increased by 12.5% and the average error is reduced by 29.3% using the SLR-MC method. The SLR-MC method can also be applied to many other meteorological data with appropriate modification.

Data Availability

The supplementary materials were provided by ERA Interim of ECMWF (European Centre for Medium-Range Weather Forecasts) and Jiangxi Meteorological Bureau. The data provided by ERA Interim were collected from Asia at 00 am, 06 am, 12 pm, and 18 pm on January 1, 2014, with a spatial resolution of 0.75 degrees, and the data provided by Jiangxi Meteorological Bureau were collected from 92 national weather stations in Jiangxi from January 2017 to March 2017.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.
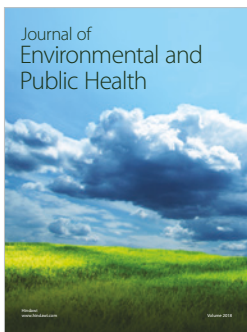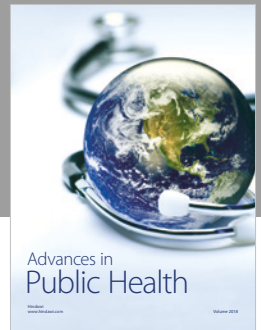
Acknowledgments

Supplementary Materials

Array size: 200×200; variable: 2 metre temperature unit = "K." The original temperature data have been recorded in Tabel original.xlsx. The original temperature data tested by the MC method have been recorded in Tabel MC.xlsx. The original temperature data tested by the SLR-MC method have been recorded in Tabel LRC-MC.xlsx. (*Supplementary Materials*)

References

[1] S. Zhang, "Remote sensing estimation of surface temperature and improvement of spatial interpolation data," Master's thesis, Capital Normal University, Beijing, China, 2014.

[2] K. Xie, X. Ning, X. Wang et al., "Recover corrupted data in sensor networks: a matrix completion solution," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1434–1448, 2017.

[3] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2008.

[4] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 2057–2078, 2012.

[5] S. Becker, V. Cevher, and A. Kyrillidis, *Randomized Singular Value Projection*, Eprint Arxiv, 2013.

[6] E. J. Candes and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[7] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[8] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, Stanford, CA, USA, 2002.

[9] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *Siam Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.

[10] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proceedings of the Neural Information Processing Systems*, pp. 937–945, Lake Tahoe, NV, USA, December 2012.

[11] Z. Lin, M. Chen, and M. Yi, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," vol. 9, 2010, https://arxiv.org/abs/1009.5055.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1953.

[13] L. Kong, D. Jiang, and M.-Y. Wu, "Optimizing the spatio-temporal distribution of cyber-physical systems for environment abstraction," in *Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, June 2010.

[14] H. Zhu, Y. Zhu, M. Li, and L. M. Ni, "SEER: metropolitan-scale traffic perception based on lossy sensory data," in *Proceedings of the 28th Conference on Computer Communications INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[15] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[16] X. M. Yuan and J. F. Yang, "Sparse and low-rank matrix decomposition via alternating direction method," *Pacific Journal of Optimization*, vol. 9, no. 1, pp. 167–180, 2013.

[17] C. Y. Li, L. Zhu, W. Z. Bao, Y. L. Jiang, C. A. Yuan, and D. S. Huang, "Convex local sensitive low rank matrix approximation," in *Proceedings of the International Joint Conference on Neural Networks*, Anchorage, AK, USA, May 2017.

[18] R. Tripathi, B. Mohan, and K. Rajawat, "Adaptive low rank matrix completion," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3603–3616, 2017.

[19] F. Ong and M. Lustig, "Beyond low rank + sparse: multi-scale low rank matrix decomposition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, March 2016.

[20] J. Hou, L.-P. Chau, N. Magnenat-Thalmann, and Y. He, "Sparse low-rank matrix approximation for data compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 5, pp. 1043–1054, 2017.

[21] J.-F. Cai and S. Osher, "Fast singular value thresholding without singular value decomposition," *Methods and Applications of Analysis*, vol. 20, no. 4, pp. 335–352, 2013.

[22] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.