Low-Rank Bi-directional Neighbor Tucker Decomposition for Traffic Data Imputation

Jiaxin Lu
Department of Statistics and Data Science
Southern University of Science and Technology, Shenzhen, China, 518055
Email: 12332881@mail.sustech.edu.cn

Wenwu Gong
Department of Statistics and Data Science
Southern University of Science and Technology, Shenzhen, China, 518055
Email: 12031299@mail.sustech.edu.cn

Lili Yang (Corresponding Author)
Department of Statistics and Data Science
Southern University of Science and Technology, Shenzhen, China, 518055
Email: yangll@sustech.edu.cn

Word Count: 5,834 words + 4 table (250 words per table) = 6,834 words

Submitted [November 28, 2024]

ABSTRACT

Modern sensors collect a large amount of spatiotemporal traffic data, which is crucial for understanding urban traffic patterns and improving daily commute efficiency. However, environmental noise often leads to data loss, undermining the effectiveness of intelligent transportation systems. Therefore, imputing missing traffic data to maintain preprocessing reliability becomes essential. The challenge lies in effectively modeling complex spatiotemporal relationships. Existing tensor-based methods have yet to fully exploit the unique properties of spatiotemporal traffic data, resulting in limited interpretability.

We propose a low-rank bi-directional neighbor Tucker decomposition (LBNTD) method to address this issue. This method leverages the data's self-correlation and incorporates bidirectional neighbor combinations to reveal hidden spatiotemporal features. By iteratively adjusting a learnable weight matrix, LBNTD can accommodate data heterogeneity, providing a solid foundation for its application. We develop a proximal alternating linear minimization algorithm to solve the LBNTD model. Numerical experiments on two real-world traffic datasets demonstrate that the LBNTD method significantly outperforms other baseline models.

Keywords: Traffic data imputation, self-correlation, heterogeneity, spatiotemporal feature, Tucker decomposition, bi-directional neighbor combination

INTRODUCTION

Spatiotemporal (ST) data are widely utilized in the real world, providing valuable insights into the behavior of complex systems. In transportation scenarios, ST data collected from modern sensors are crucial for monitoring traffic conditions and devising traffic operation strategies (*1*). However, equipment failures and communication issues often hinder data acquisition, resulting in unavoidable loss (*2*). These gaps can significantly impair the performance of intelligent transportation systems. Consequently, there is a great need to impute missing elements to improve data availability.

ST data exhibit two features: self-correlation and heterogeneity (*3*). Firstly, measurements in neighboring coordinates are not independent but are contextually connected. For example, traffic flows observed at nearby locations often display similar trends, and data fluctuations within a specific range tend to be smoother. Secondly, ST data exhibit different patterns across regions and periods, such as the traffic conditions on weekdays and weekends and in city centers and suburbs, which tend to differ significantly. **Figure 1** visualizes a sample of 20 days and 20 road sections from the Guangzhou urban speed dataset, highlighting the presence of these two qualities. Specifically, traffic status both upstream and downstream will mutually influence each other, alongside the travel behaviors that emerge on specific dates. Moreover, the traffic situations on different road sections within the city, as well as during different days of the week, clearly demonstrate local diversity.

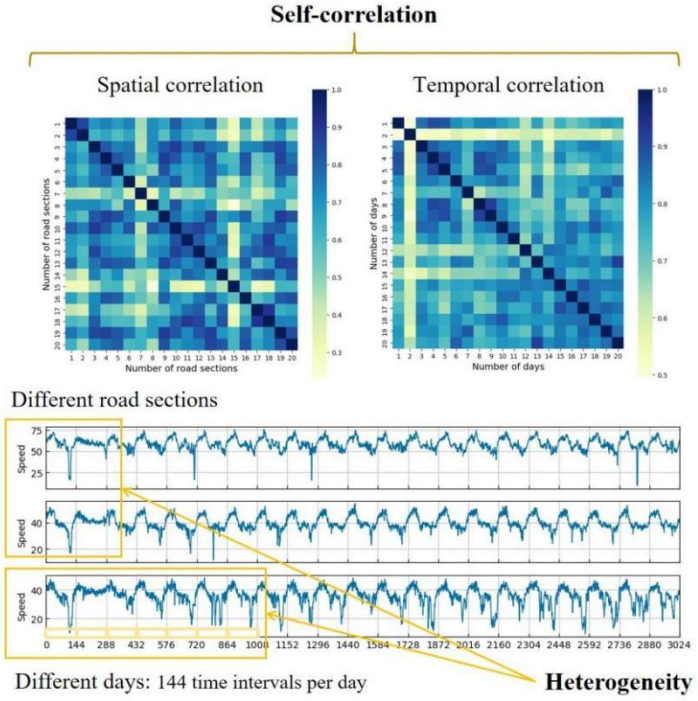

Figure 1 Visualisation of spatiotemporal data features.

Traditional imputation methods, such as KNN (*4*) and Cokriging (*5*), neglect the intrinsic spatiotemporal features of traffic data and result in poor capability (*6*). Since the multilayer structure of tensors can make the data more comprehensive, tensors tend to be used to present ST data (*7*). It is important to note that most existing models provide only a superficial framework for exploring the interactions between constituents in ST data. They fail to deeply investigate the connections between nodes across time and space, making them unsuitable for implementation in urban traffic situations characterized by diverse layouts and fluctuating states (*8*). In other words, there are short and long range impacts between sites, which lead to differences in the corresponding transport routes due to various regional functions. We propose a novel Low-rank Bi-directional Neighbor Tucker Decomposition

(LBNTD) model to fully leverage the spatiotemporal features in traffic data. It assumes that each missing element is represented by a linear combination of the surrounding elements (*9*). Unlike most methods that create a static weight matrix for spatiotemporal information metric (*10*), the proposed LBNTD model integrates the weight determination into the overall optimization algorithm and updates the weights iteratively. This dynamic approach allows spatiotemporal modeling in an inductive learning way, dealing with the heterogeneity of the ST data. The contributions of this paper are as follows:

- We depict the self-correlation of traffic data using Tucker decomposition, and exploit the hidden spatiotemporal features with bi-directional neighbors. The proposed LBNTD model yields a learnable weight matrix that fits different traffic datasets.
- We propose a proximal alternating linear minimization algorithm to solve non-convex optimization problems and discuss its numerical convergence.
- Extensive numerical experiments on two real-world traffic datasets demonstrate the superior performance of the LBNTD model compared to other state-of-the-art models.

RELATED WORK

In this section, we review contemporary imputation methods for spatiotemporal traffic data, organizing the discussion around two commonly used data representation tools. We primarily focus on the tensor-based approaches, outlining relevant techniques for mining spatiotemporal features and discerning their effectiveness.

Firstly, it is widely recognized that traffic data is directly described in matrix form. Su et al. (*11*) integrated the spatial information of the road network with the original traffic data matrix and employed a latent factor model-based algorithm to impute missing data in the combined matrix. Chen et al. (*12*) introduced an ensemble low-rank matrix completion method, which constructs a distance matrix to select the most similar samples for the filling via KNN search. However, they only capture spatiotemporal features in two dimensions, leading to insufficient precision and lack of scalability (*13*).

In contrast, tensor allows the inclusion of three or more dimensions, thereby preserving richer data characteristics (*14*). Wang et al. (*15*) transformed the raw data into a fourth-order Hankel tensor. They approximated the tensor rank by balancing the truncated nuclear norms of the spatiotemporal expansions. Ran et al. (*16*) proposed a traffic flow data fusion method that reconstructs the data in a four-way tensor model, fully utilizing spatiotemporal information from upstream and downstream detectors to improve estimation. With the help of tensors, many attempts to model spatiotemporal features in traffic data have gradually been derived (*17*).

Some research extracts spatiotemporal features by exploring the inherent correlation within traffic data. Wu et al. (*18*) discovered the nonlocal self-similarity prior in traffic data and proposed a corresponding spatiotemporal imputation method. They used tensor ring decomposition to depict the global low-rankness. Chen et al. (*19*) introduced a collaborative non-convex low-rank tensor completion model that delves into the intrinsic structure of traffic data. Multidimensional correlations and detailed spatiotemporal dependency are inscribed using an elastic net self-representation method. These methods show enhanced performance but do not clearly explain the rationale behind modeling spatiotemporal features in this particular way.

Meanwhile, other studies maintain better interpretability when encoding spatiotemporal features. Zhao et al. (*20*) introduced a spatiotemporal constrained low-rank tensor completion method using manifold embedding based on traffic flow's continuity, periodicity, and transmission mechanisms. This approach effectively utilizes the traffic tensor's global and local linkages. However, the decision degree between elements, as defined by these methods, is fixed from the start, discouraging the flexibility needed to uncover the internal rules of different datasets. Zhang et al. (*21*) developed a tensor-weighted Schatten-p norm model with neighbor regularization that preserves rank components of traffic data by assigning variable weights to the tensorial and decomposed mode dimension. It also mines the spatial consistency of the traffic network by utilizing the contributions of passenger flows. Nevertheless, solving for the weight matrices is cumbersome and involves executing specific deletion operations without the possibility of deriving a uniform closed-form solution.

Our proposed model overcomes the above deficiencies. It uses traffic data's self-correlation to construct a spatiotemporal regularization and elaborates on the underlying concepts. Moreover, a specialized algorithm has been developed to compute a learnable weight matrix, effectively managing traffic data heterogeneity.

NOTATIONS

In this paper, we adopt the following symbols for clarity. The mode-n unfolding of a tensor $\mathcal{X}$ is denoted by $\mathbf{X}_{(n)}$. We represent the (m,n)-th entry in a matrix $\mathbf{X}$ as $x_{m,n}$. The observed index set and its complement are denoted by Ω and $\overline{\Omega}$. The mode-n product is represented by $\times_n$, while the Kronecker product is denoted by $\otimes$. Norms are expressed as follows: the Frobenius norm by $\|\cdot\|_F$, the ℓ_1 norm by $\|\cdot\|_1$, and the ℓ_2 norm by $\|\cdot\|_2$, the nuclear norm is denoted by $\|\cdot\|_*$. Moreover, the inner product is expressed as $\langle\cdot,\cdot\rangle$. The soft-thresholding operator (*22*) with η is denoted by $S_\eta(\cdot)$, while the singular value shrinkage operator (*23*) with η is represented by $D_\eta(\cdot)$.

For a third-order tensor $\mathcal{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$, we define the projection operator $\mathcal{P}_\Omega(\mathcal{X})$, where $\mathcal{P}_\Omega(\mathcal{X})$ equals $\mathrm{x_{i,j,k}}$ if $(\mathrm{i},\mathrm{j},\mathrm{k}) \in \Omega$ and 0 otherwise. Then, we introduce the tensorization operator $\mathcal{Q}(\cdot)$. Through it, $\mathcal{X}$ is generated from a matrix Y as $\mathcal{X} = \mathcal{Q}(\mathrm{Y})$. The inverse operation, $\mathcal{Q}^{-1}(\cdot)$, converts $\mathcal{X}$ back into the original matrix Y, resulting in $\mathrm{Y} = \mathcal{Q}^{-1}(\mathcal{X})$ in $\mathbb{R}^{\mathrm{M}\times(\mathrm{IJ})}$. Let T represent the product of all dimensions except the first. The Tucker decomposition can be expressed as $\mathcal{X} = \mathcal{G}\times_1\mathbf{U}_1\cdots\times_{\mathrm{N}}\mathbf{U}_{\mathrm{N}} = \mathcal{G}\times_{\mathrm{n=1}}^{\mathrm{N}}\mathbf{U}_n$. This can be equivalently written as $\mathbf{X}_{(n)} = \mathbf{U}_n\mathbf{G}_{(n)}\mathbf{V}_n^{\mathrm{T}}$, where $\mathbf{V}_n = \mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \cdots \otimes \mathbf{U}_1$. It can be readily verified that $\mathrm{vec}(\mathcal{X}) = (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_n \otimes \cdots \otimes \mathbf{U}_1)\mathrm{vec}(\mathcal{G}) = (\otimes_{\mathrm{n=N}}^{1}\mathbf{U}_n)\mathrm{vec}(\mathcal{G})$.

$Cohen's\ d$ is a widely used metric for assessing between-group differences, measuring the degree of change in a variable. Its computational formula is $\mathrm{d} = (\overline{\mathrm{X}}_1 - \overline{\mathrm{X}}_2)/\mathrm{s}$, where $\overline{X}_1$ and $\overline{X}_2$ are the respective sample means for the two groups, s denotes the overall standard deviation of the two groups.

METHODOLOGY

In this section, we first present the construction principles of the low-rank bi-directional neighbor Tucker decomposition (LBNTD) model based on its intended purpose. Next, we propose a specific optimization algorithm to address the target problem and provide the relevant theoretical derivation.

Proposed Model

Since the tensor is an ideal representation of high-dimensional spatiotemporal data (*24*), we format the original traffic data as a third-order tensor of "sensor × time intervals × day." We assume that the tensor is restricted to be of low rank (*25*), which naturally leads to clustering of the data. We use Tucker decomposition to strengthen the commonality between dimensions by mining the cross-influence of the core tensor with the individual factor matrices (*26*).

Due to the self-correlation within the ST data, we consider using neighboring node information to guide imputation. Therefore, the missing elements are expressed through bi-directional neighbor combinations. Empirically, we give a set of neighbor metrics $\mathcal{R} = \{r_1, \cdots, r_d\}$, where the value of r_i is a positive integer indicating corresponding unit distances. Relying on the idea of bi-directional neighbors, we construct a spatiotemporal regularization:

$$\|\mathbf{Z}\|_{\mathbf{B},\mathcal{R}} = \sum_{s,m}\left(z_{s,m} - \sum_j b_{m,j}z_{s+r_j,m} - \sum_j b_{m,j}z_{s-r_j,m}\right)^2,$$

where $\mathbf{B}$ is the weight matrix that can be learned and responds to the heterogeneity of ST data.

The low-rank bi-directional neighbor Tucker decomposition (LBNTD) model is designed as

$$\min_{\mathcal{G},\{\mathbf{U}_n\},\mathcal{X},\mathbf{Z},\mathbf{B}} (1-\alpha)\sum_{n=1}^{N}\omega_n\|\mathbf{U}_n\|_* + \alpha\|\mathcal{G}\|_1 + \frac{\gamma}{2}\|\mathbf{Z}\|_{\mathbf{B},\mathcal{R}} \ \mathrm{s.t.}, \mathcal{X} = \mathcal{G}\underset{n=1}{\overset{N}{\times}}\mathbf{U}_n,\ \mathcal{X} = \mathcal{Q}(\mathbf{Z}),\ \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Y}) \quad (1)$$

where $0 < \alpha < 1$, $\omega_n = \prod_{i=1, i\neq n}^{3} 1/R_i$, $R_i = \sum \sigma(\mathbf{U}_i)$. The tensor $\mathcal{Y}$ represents the observed data, and γ is used to balance the low-rank term and spatiotemporal regularization.

It is necessary to state that the dataset studied in this paper is not yet anisotropic. Therefore, elements distributed at equal distances in both directions have the same contributions for center position.

Proposed Algorithm

To solve this problem, we begin by imposing penalties to the two constraints specified in **Equation 1**. Here, β and ρ serve as the corresponding parameters.

$$\begin{aligned} \min_{\mathcal{G},\{\mathbf{U}_n\},\mathcal{X},\mathbf{Z},\mathbf{B}} \quad & (1-\alpha)\sum_{n=1}^{N}\omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \frac{\gamma}{2}\|\mathbf{Z}\|_{\mathbf{B},\mathcal{R}} + \frac{\beta}{2}\|\mathcal{X} - \mathcal{G}\times_{n=1}^{N}\mathbf{U}_n\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathcal{X} - \mathcal{Q}(\mathbf{Z})\|_{\mathrm{F}}^2 \\ \text{s.t.}, \quad & \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{Y}) \end{aligned} \tag{2}$$

We utilize the proximal alternating linear minimization method to find a solution for **Equation 2**, which involves breaking down the solution process into several subproblems and taking an approximation strategy for some of them.

Optimization of $\mathcal{G}$, $\{U_n\}, n = 1, \ldots, N$.

To get the core tensor $\mathcal{G}$, we use the proximal gradient method in its vectorized form:

$$\begin{aligned} \widehat{\mathcal{G}} &= \operatorname*{argmin}_{\mathcal{G}} \alpha\|\mathcal{G}\|_1 + \frac{\beta}{2}\|\mathrm{vec}(\mathcal{X}) - (\bigotimes_{n=N}^{1}\mathbf{U}_n)\mathrm{vec}(\mathcal{G})\|_{\mathrm{F}}^2 \\ &\approx \operatorname*{argmin}_{\mathcal{G}} \alpha\|\mathcal{G}\|_1 + \langle \mathcal{G} - \widetilde{\mathcal{G}}, \nabla_{\mathcal{G}} f(\widetilde{\mathcal{G}})\rangle + \frac{L_{\mathcal{G}}}{2}\|\mathcal{G} - \widetilde{\mathcal{G}}\|_{\mathrm{F}}^2 \\ &= \mathcal{S}_{\frac{\alpha}{L_{\mathcal{G}}}}\left(\widetilde{\mathcal{G}} - \frac{1}{L_{\mathcal{G}}}\nabla_{\mathcal{G}} f(\widetilde{\mathcal{G}})\right). \end{aligned} \tag{3}$$

The gradient and Lipschitz constant are given by

$$\nabla_{\mathcal{G}} f(\mathcal{G}) = \beta\left(\mathcal{G}\times_{n=1}^{N}\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n - \mathcal{X}\times_{n=1}^{N}\mathbf{U}_n^{\mathrm{T}}\right),\ L_{\mathcal{G}} = \beta\prod_{n=1}^{N}\|\mathbf{U}_n^{\mathrm{T}}\mathbf{U}_n\|_2 .$$

To obtain the hidden factor matrix $\mathbf{U}_n$, we perform mode-n unfolding operation:

$$\begin{aligned} \widehat{\mathbf{U}}_n &= \operatorname*{argmin}_{\mathbf{U}_n} (1-\alpha)\omega_n\|\mathbf{U}_n\|_* + \frac{\beta}{2}\|\mathbf{X}_{(n)} - \mathbf{U}_n\mathbf{G}_{(n)}\mathbf{V}_n^{\mathrm{T}}\|_{\mathrm{F}}^2 \\ &\approx \operatorname*{argmin}_{\mathbf{U}_n} (1-\alpha)\omega_n\|\mathbf{U}_n\|_* + \langle \mathbf{U}_n - \widetilde{\mathbf{U}}_n, \nabla_{\mathbf{U}_n} f(\widetilde{\mathbf{U}}_n)\rangle + \frac{L_{\mathbf{U}_n}}{2}\|\mathbf{U}_n - \widetilde{\mathbf{U}}_n\|_{\mathrm{F}}^2 \\ &= \mathcal{D}_{\frac{(1-\alpha)\omega_n}{L_{\mathbf{U}_n}}}\left(\widetilde{\mathbf{U}}_n - \frac{1}{L_{\mathbf{U}_n}}\nabla_{\mathbf{U}_n} f(\widetilde{\mathbf{U}}_n)\right), \end{aligned} \tag{4}$$

The gradient and Lipschitz constant are given by

$$\nabla_{\mathbf{U}_n} f(\mathbf{U}_n) = \beta\left(\mathbf{U}_n\mathbf{G}_{(n)}\mathbf{V}_n^{\mathrm{T}}\mathbf{V}_n\mathbf{G}_{(n)}^{\mathrm{T}} - \mathbf{X}_{(n)}\mathbf{V}_n\mathbf{G}_{(n)}^{\mathrm{T}}\right),\ L_{\mathbf{U}_n} = \beta\|\mathbf{G}_{(n)}\mathbf{V}_n^{\mathrm{T}}\mathbf{V}_n\mathbf{G}_{(n)}^{\mathrm{T}}\|_2.$$

We update $\widetilde{\mathcal{G}}$ and $\widetilde{\mathbf{U}}_n$ by $\widetilde{\mathcal{G}}^k = \mathcal{G}^k + \eta_k\left(\mathcal{G}^k - \mathcal{G}^{k-1}\right),\ \widetilde{\mathbf{U}}_n^k = \mathbf{U}_n^k + \eta_k\left(\mathbf{U}_n^k - \mathbf{U}_n^{k-1}\right)$, for $k \geq 1$, where $\eta_k = \frac{t^{k-1}-1}{t^k},\ t^k = \frac{1+\sqrt{4(t^{k-1})^2+1}}{2},\ t^0 = 1.$

Optimization of Z.

The original problem can be reformulated in terms of $\mathbf{Z}$ as shown in **Equation 5**:

$$\begin{aligned}\widehat{\mathbf{Z}} &= \underset{\mathbf{Z}}{\operatorname{argmin}} \frac{\gamma}{2}\|\mathbf{Z}\|_{\mathbf{B},\mathcal{R}} + \frac{\rho}{2}\|\mathbf{Z} - \mathcal{Q}^{-1}(\mathcal{X})\|_F^2 \\ &= \underset{\mathbf{Z}}{\operatorname{argmin}} \sum_m \left[\frac{\gamma}{2}\|\Psi_0\mathbf{z}_m - \sum_j b_{m,j}\Psi_j\mathbf{z}_m\|_2^2 + \frac{\rho}{2}\|\mathbf{z}_m - \mathcal{Q}_m^{-1}(\mathcal{X})\|_2^2\right].\end{aligned} \tag{5}$$

By introducing some simple notation, we can express a closed-form solution to this problem:

$$\hat{\mathbf{z}}_m = \frac{\rho}{\gamma}\left(\mathbf{A}_m^{\mathrm{T}}\mathbf{A}_m + \frac{\rho}{\gamma}\mathbf{I}\right)^{-1} \cdot \mathcal{Q}_m^{-1}(\mathcal{X}), m = 1, ..., T, \tag{6}$$

where $\mathbf{A}_m = \Psi_0 - \sum_j b_{m,j}(\Psi_j^- + \Psi_j^+), m = 1, ..., T$

with $\Psi_0 = \left[\mathbf{0}_{(M-2r_d)\times r_d}\mathbf{I}_{M-2r_d}\mathbf{0}_{(M-2r_d)\times r_d}\right]$,

$\Psi_j^- = \left[\mathbf{0}_{(M-2r_d)\times(r_d-r_j)}\,\mathbf{I}_{M-2r_d}\,\mathbf{0}_{(M-2r_d)\times(r_d+r_j)}\right]$,

$\Psi_j^+ = \left[\mathbf{0}_{(M-2r_d)\times(r_d+r_j)}\,\mathbf{I}_{M-2r_d}\,\mathbf{0}_{(M-2r_d)\times(r_d-r_j)}\right]$.

They are defined based on neighbor set $\mathcal{R} = \{r_1, \cdots, r_d\}$.

Optimization of $\mathcal{X}$.

Integrating the known constraints allows us to present the comprehensive form of the result:

$$\widehat{\mathcal{X}}_{\bar{\Omega}} = \underset{\mathcal{X}}{\operatorname{argmin}} \frac{\beta}{2}\|\mathcal{X} - \mathcal{G}\overset{\mathrm{N}}{\underset{n=1}{\times}}\mathbf{U}_n\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\,\mathcal{X} - \mathcal{Q}(\mathbf{Z})\|_{\mathrm{F}}^2 = \left[\frac{1}{\beta+\rho}\left(\beta\mathcal{G}\overset{\mathrm{N}}{\underset{n=1}{\times}}\mathbf{U}_n + \rho\mathcal{Q}(\mathbf{Z})\right)\right]_{\bar{\Omega}},\ \widehat{\mathcal{X}}_{\Omega} = \mathcal{Y}_{\Omega}. \tag{7}$$

Optimization of B.

Finally, the weight matrix is obtained using Equation 8 below. Converting the original format into vectorized form will simplify the solution:

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}}{\operatorname{argmin}} \sum_{l,m}(\,z_{l,m} - \sum_j b_{m,j}z_{l+r_j,m} - \sum_j b_{m,j}z_{l-r_j,m})^2 = \underset{\mathbf{B}}{\operatorname{argmin}} \sum_m \|\mathbf{z}_{0,m} - \mathbf{W}_m^-\mathbf{b}_m - \mathbf{W}_m^+\mathbf{b}_m\|_2^2, \tag{8}$$

where $\mathbf{W}_m^- = \left(\mathbf{w}_{1+r_d}^-, \cdots, \mathbf{w}_{M-r_d}^-\right)^{\mathrm{T}}$,

$\mathbf{W}_m^+ = \left(\mathbf{w}_{1+r_d}^+, \cdots, \mathbf{w}_{M-r_d}^+\right)^{\mathrm{T}}$,

$\mathbf{w}_l^- = (z_{l-r_1,m}, \cdots, z_{l-r_d,m})^{\mathrm{T}}$,

$\mathbf{w}_l^+ = (z_{l+r_1,m}, \cdots, z_{l+r_d,m})^{\mathrm{T}}, l = 1 + r_d, ..., M - r_d$.

The solution can be explicitly determined using the following expression:

$$\widehat{\mathbf{b}}_m = (\mathbf{W}_m^- + \mathbf{W}_m^+)^{\dagger}\mathbf{z}_{0,m}, m = 1, ..., T, \tag{9}$$

where $\cdot^{\dagger}$ denotes the Moore-Penrose pseudo-inverse, $\mathbf{z}_{0,m} = (z_{1+r_d,m}, \cdots, z_{M-r_d,m})^{\mathrm{T}}, m = 1, ..., T$.

It is important to note that our algorithm features both an inner and an outer loop. Initially, the weight matrix is fixed to solve for the other variables, followed by a return to update itself.

The whole algorithm for LBNTD is detailed in **Algorithm 1** and yields $\widehat{\mathcal{X}}$ as the estimation result.

Algorithm 1: PALM-based LBNTD

Input : Missing tensor $\mathcal{Y}$, observed index set Ω.
Output: Recovered tensor $\hat{\mathcal{X}}$.
Initialize: $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ $(1 \le n \le N)$, $0 < \alpha < 1$, $\beta = \rho = 1$, $\gamma = \min\{\beta, \rho\}$, $s = 1$;
repeat
  for $k = 0$ **to** $K - 1$ **do**
    Update $\mathcal{G}^{s+1,k+1}$ using (3).
    for $n = 1$ **to** N **do**
      Update $\mathbf{U}_n^{s+1,k+1}$ using (4).
    end
    for $m = 1$ **to** T **do**
      Update $\boldsymbol{z}_m^{s+1,k+1}$ using (6).
    end
    Update $\mathcal{X}^{s+1,k+1}$ using (7).
  end
  for $m = 1$ **to** T **do**
    Update $\boldsymbol{b}_m^{s+1}$ using (9).
  end
  s=s+1;
until $\left\|\mathcal{X}^{s+1,K} - \mathcal{X}^{s,K}\right\|_{\mathrm{F}} \left\|\mathcal{X}^{s,K}\right\|_{\mathrm{F}}^{-1} < \epsilon$;

EXPERIMENTS

In this section, we assess the performance of the proposed model through experiments conducted on several real-world traffic datasets.

Traffic Datasets

- *Guangzhou Urban Traffic Speed Dataset*, denoted by the symbol *G*.
  It consists of traffic speed data collected at 10-minute intervals from 214 road segments in Guangzhou, China. The dataset has dimensions of $214 \times 144 \times 7$.
  (https://doi.org/10.5281/zenodo.1205229)
- *Hangzhou Metro Passenger Flow Dataset*, denoted by the symbol *H*.
  It records inbound passenger flow data at 10-minute intervals for 80 subway stations in Hangzhou, China. The 6 hours per day of unserved phases have been removed. The resulting dataset dimensions are $80 \times 108 \times 7$.
  (https://tianchi.aliyun.com/competition/entrance/231708/information)

Experimental Settings

To test the efficiency of the LBNTD model, we consider a scenario where missing data is randomly distributed. Besides, mean absolute percentage error (MAPE) and normalized mean absolute error (NMAE) were used as metrics:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100,\ \text{NMAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i|},$$

where y_i and $\hat{y}_i$ are real and imputed values, respectively.

To demonstrate the superiority of our model, we compared it with several imputation methods: ECLRMC, STRTD, LRSETD, stTT, and STH-LRTC. Both of them have a low-rank term and a spatiotemporal regularization. Their main characteristics are detailed in **Table 1**.

TABLE 1 Some Existing Traffic Data Imputation Methods

Methods	Types of the prior			Structures
	Low rankness	Spatial	Temporal	
ECLRMC (*12*)	√	√		matrix
LBNTD	√	√		3rd tensor
STRTD (*17*)	√	√	√	3rd tensor
LRSETD (*14*)	√	√	√	3rd tensor
stTT (*13*)	√	√	√	3rd tensor
STH-LRTC (*15*)	√	√	√	4th tensor

The LBNTD model involves hyperparameters tuning. Extensive numerical results show the parameters for the first dataset to be $\{\alpha = 0.98, \beta = 0.08, \rho = 10^{-14}, \gamma = 10^{-6}\}$, $\mathcal{R} = \{1,2,\cdots,6\}$. Whereas for the second dataset, let $\{\alpha = 0.99,\ \beta = 0.0004,\ \rho = 20^{-12},\ \gamma = 20^{-7}\}$, $\mathcal{R} = \{1,2,\cdots,7\}$.

We set the termination condition uniformly to tol = 10^{-5} and limit the maximum number of iterations to 300 for all experiments. We also configure the parameters for the comparison models according to the descriptions in the reference paper.

Results and Discussion

To begin, we focused solely on the terms characterizing low-rankness in each model compared, illustrating how two data structures — the matrix and the third-order tensor — impact the models' capability. It is important to note that the fourth-order tensor was excluded from this experiment because it inherently integrates spatiotemporal information during its generation. From **Figure 2**, we can find that the model based on the third-order tensor consistently achieves much lower errors. This suggests that the third-order tensor captures more detailed traffic data and confirms that self-correlation can effectively be used for completion.

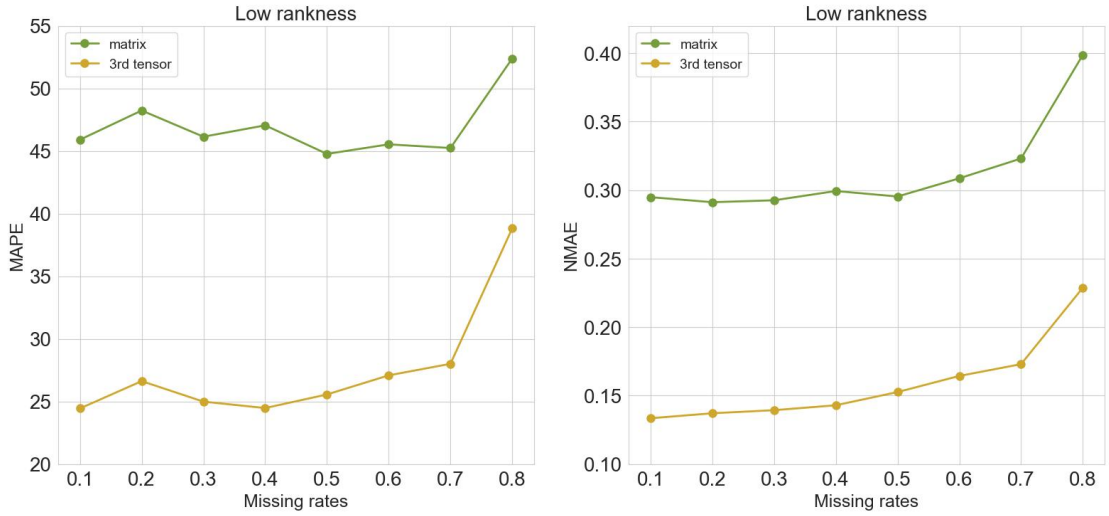

Figure 2 Dimensional effects of dataset *H*

Next, we display the performance changes in the proposed model before and after introducing the spatiotemporal prior by plotting the graphs. **Figure 3** indicates that the spatiotemporal regularization indeed plays a role in imputing missing elements. The calculation of $Cohen's\ d$ yields an effect size of roughly 0.3, suggesting that the introduction of the prior can contribute to data completion to some extent.

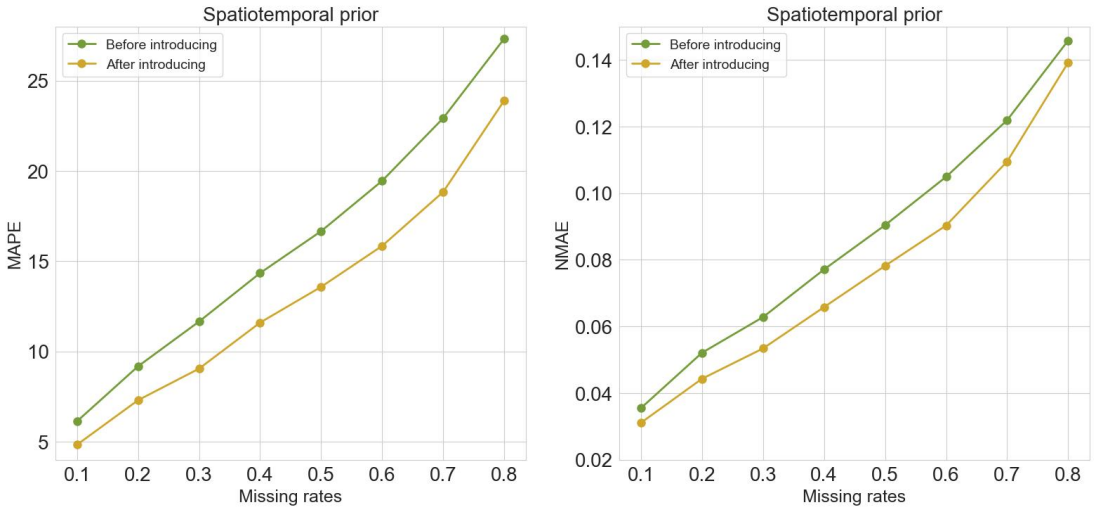

Figure 3 Impact testing of spatiotemporal prior in dataset *H*

We also present a heatmap of the weight matrix for a specific case in **Figure 4**, revealing that columns at odd positions predominantly take positive values, whereas columns at even positions tend towards negative values, suggesting interactions between segments. It is also noted that the nearest segments typically exhibit a facilitative effects, while the furthest segments are more inclined to inhibit each other, reflecting the heterogeneity of the spatiotemporal data.

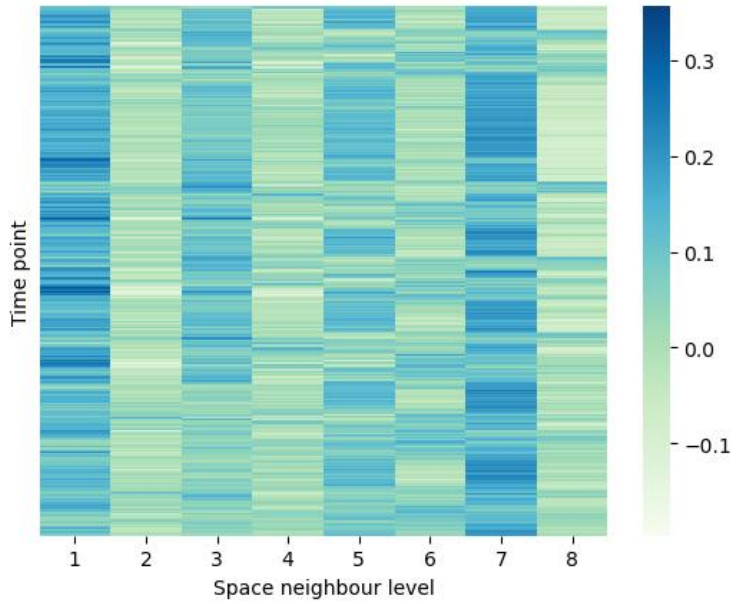

Figure 4 Visualisation of weight matrix under 20% missing dataset *G*

The imputation performance of LBNTD compared to several baseline models on two traffic datasets is summarized in **Table 2**, with the best results highlighted in bold. **Figure 5** illustrates the MAPE and NMAE values across different missing rates. The findings indicate that the LBNTD is superior to other models, delivering the highest imputation precision under the given missing conditions.

TABLE 2 Performance Comparison of LBNTD and Baseline Models

Dataset	Missing	LBNTD	ECLRMC	STRTD	LRSETD	stTT	STH-LRTC
(*G*)-MAPE	20%	**1.4146**	10.2601	7.2531	8.5893	10.8666	5.9454
	40%	**3.0684**	10.9905	7.8542	9.3302	10.9796	6.4935
	60%	**4.9221**	12.8423	8.3822	10.0502	10.9232	7.2259
	80%	**7.6427**	15.0101	9.3750	11.0326	11.1064	8.9860
(*G*)-NMAE	20%	**0.0112**	0.0743	0.0553	0.0640	0.0774	0.0460
	40%	**0.0237**	0.0830	0.0587	0.0686	0.0780	0.0495
	60%	**0.0375**	0.1036	0.0621	0.0732	0.0782	0.0549
	80%	**0.0569**	0.1272	0.0684	0.0801	0.0788	0.0674
(*H*)-MAPE	20%	**7.3002**	37.2798	19.9549	27.5051	27.9337	31.3135
	40%	**11.5988**	37.2128	21.0797	28.8972	26.0291	30.1205
	60%	**15.8487**	40.1555	22.4406	33.2615	30.5960	38.2953
	80%	**23.9022**	45.2961	24.5118	38.9334	31.7952	49.2579
(*H*)-NMAE	20%	**0.0443**	0.2698	0.1182	0.2001	0.1303	0.1320
	40%	**0.0658**	0.2819	0.1225	0.2214	0.1331	0.1409
	60%	**0.0903**	0.2950	0.1299	0.2492	0.1434	0.1588
	80%	**0.1392**	0.3911	0.1495	0.2852	0.1641	0.1997

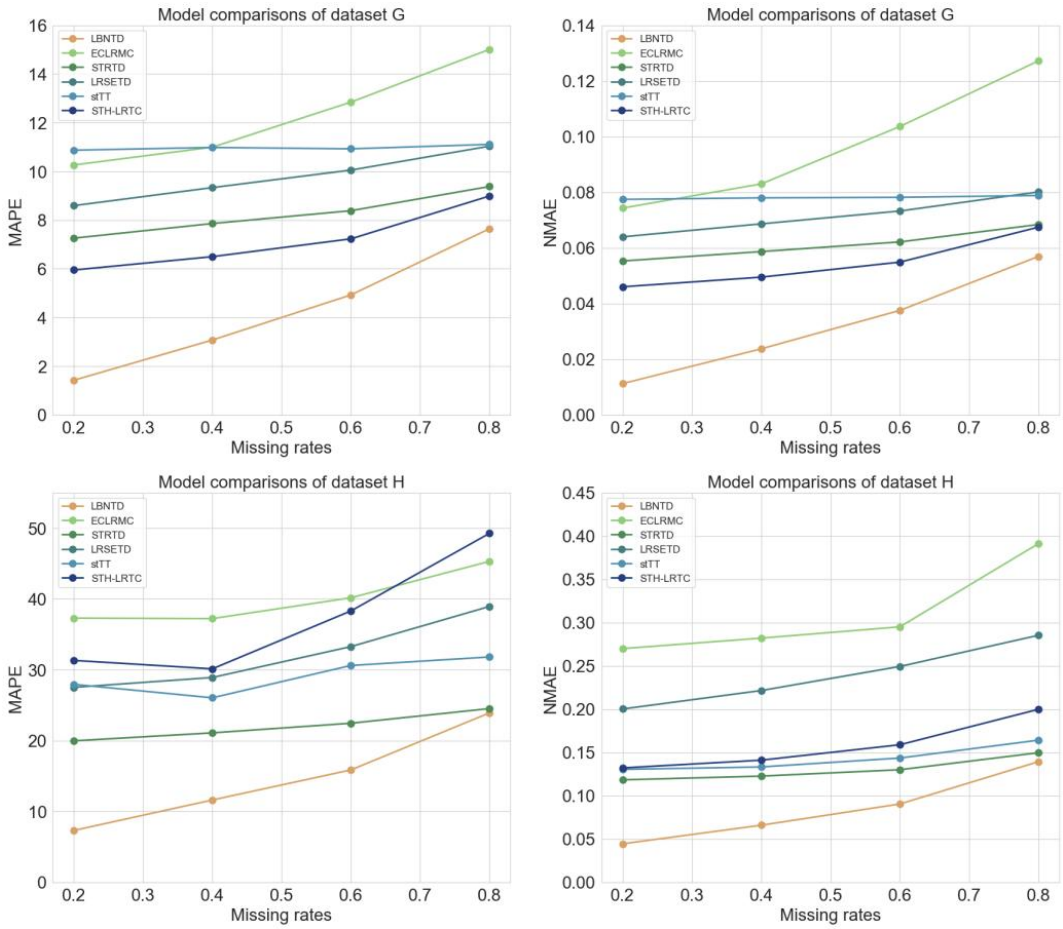

Figure 5 Performance comparison based on MAPE values (left) and NMAE values (right) across different missing rates for the *G* dataset (upper) and the *H* dataset (lower).

Additionally, we also present the computation time for all models at each missing rate in **Table 3**. The results indicate that our proposed model performs acceptably in terms of efficiency.

TABLE 3 Computation Time (in seconds) of Imputation Methods

Dataset	Missing	LBNTD	ECLRMC	STRTD	LRSETD	stTT	STH-LRTC
(*G*)-time	20%	576.0667	580.7604	28.8633	14.9848	0.3285	1236.6661
	40%	855.8848	710.1477	31.3499	19.4502	0.5282	1464.4715
	60%	577.2102	748.1811	35.8618	19.5849	0.8027	1451.5579
	80%	575.5270	860.7149	34.3666	23.9353	1.4635	1841.1340
(*H*)-time	20%	64.7596	142.4093	4.8572	5.2647	0.2302	144.6486
	40%	75.5777	197.9183	6.2006	5.8623	0.3294	168.3460
	60%	75.3918	277.2015	5.0135	5.7550	0.6477	188.3594
	80%	72.4656	336.7025	4.6138	6.5225	1.0951	203.2036

CONCLUSION

The proposed LBNTD model effectively imputes the traffic data with an explicable framework. It incorporates a bi-directional neighbor combination to capture spatiotemporal features and make surrounding elements aid in estimating missing values. The resulting learnable weight matrix is adaptable to various datasets. Experimental results on real-world traffic datasets demonstrate that the LBNTD model outperforms other state-of-the-art imputation methods.

For simplicity, we assume that each constituent is represented independently by a bidirectional neighbor combination and ignore their cross-interference. The consideration of multivariate schemes will further enhance the practicality of our model by making it more relevant to real-world scenarios. Meanwhile, We intend to engage with local transport authorities to seek specialized expert insights. This partnership will empower us to design more sophisticated analytical methods for multimodal data, thereby improving the efficiency of data-driven strategies. In the future, we remain committed to extending the generalisation of the model, ensuring its robustness in cross-regional contexts. We will unify GPS grid data and detector graph data into a single representation during the preprocessing stage. This will enable the implementation of data fusion and broaden the applicability of completion techniques.

ACKNOWLEDGMENT

This work is supported by the Shenzhen Science and Technology Program (Grant No. ZDSYS20210623092007023), the SUSTech Presidential Postdoctoral Fellowship, and the Educational Commission of Guangdong Province (Grant No. 2021ZDZX1069).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Jiaxin Lu, Wenwu Gong; data collection: Lili Yang; analysis and interpretation of results: Jiaxin Lu, Wenwu Gong; draft manuscript preparation: Jiaxin Lu, Wenwu Gong, Lili Yang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Chen, X., J. Yang, and L. Sun, A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies*, Vol. 117, 2020, p. 102673.

2. Said, A. B. and A. Erradi, Spatiotemporal tensor completion for improved urban traffic imputation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 7, 2022, pp. 6836–6849.

3. Atluri, G., A. Karpatne, and V. Kumar, Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, Vol. 51, No. 4, 2018, pp. 1–41.

4. Wang, Y., Y. Xiao, J. Lai, and Y. Chen, An adaptive k nearest neighbour method for imputation of missing traffic data based on two similarity. *Archives of Transport*, Vol. 54, No. 2, 2020, pp. 59–73.

5. Baffoe-Twum, E., E. Asa, and B. Awuku, Estimating annual average daily traffic (AADT) data on low-volume roads with the cokriging technique and census/population data. *Emerald Open Research*, Vol. 1, No. 5, 2023.

6. Pinto, F. C., J. G. Manchuk, and C. V. Deutsch, Decomposition of multivariate spatial data into latent factors. *Computers & Geosciences*, Vol. 153, 2021, p. 104773.

7. Zhao, Y., M. Tuo, H. Zhang, H. Zhang, J. Wu, and F. Gao, Nonnegative low-rank tensor completion method for spatiotemporal traffic data. *Multimedia Tools and Applications*, 2023, pp. 1–16.

8. Wang, Y., Y. Zhang, X. Piao, H. Liu, and K. Zhang, Traffic data reconstruction via adaptive spatial-temporal correlations. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 4, 2019, pp. 1531–1543.

9. Farzaneh, G., N. Khorasani, J. Ghodousi, and M. Panahi, Application of geostatistical models to identify spatial distribution of groundwater quality parameters. *Environmental Science and Pollution Research*, Vol. 29, No. 24, 2022, pp. 36512–36532.

10. Madani, N. and S. Abulkhair, A hierarchical cosimulation algorithm integrated with an acceptance–rejection method for the geostatistical modeling of variables with inequality constraints. *Stochastic Environmental Research and Risk Assessment*, Vol. 34, No. 10, 2020, pp. 1559–1589.

11. Su, X., W. Sun, C. Song, Z. Cai, and L. Guo, A latent-factor-model-based approach for traffic data imputation with road network information. *ISPRS International Journal of Geo-Information*, Vol. 12, No. 9, 2023, p. 378.

12. Chen, X., Z. Wei, Z. Li, J. Liang, Y. Cai, and B. Zhang, Ensemble correlation-based low-rank matrix completion with applications to traffic data imputation. *Knowledge-Based Systems*, Vol. 132, 2017, pp. 249–262.

13. Zhang, Z., C. Ling, H. He, and L. Qi, A tensor train approach for internet traffic data completion. *Annals of Operations Research*, 2021, pp. 1–19.

14. Pan, C., C. Ling, H. He, L. Qi, and Y. Xu, A low-rank and sparse enhanced Tucker decomposition approach for tensor completion. *Applied Mathematics and Computation*, Vol. 465, 2024, p. 128432.

15. Wang, X., Y. Wu, D. Zhuang, and L. Sun, Low-rank Hankel tensor completion for traffic speed estimation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24, No. 5, 2023, pp. 4862–4871.

16. Ran, B., H. Tan, Y. Wu, and P. J. Jin, Tensor based missing traffic data completion with spatial–temporal correlation. *Physica A: Statistical Mechanics and its Applications*, Vol. 446, 2016, pp. 54–63.

17. Gong, W., Z. Huang, and L. Yang, Spatiotemporal regularized Tucker decomposition for traffic data imputation. *arXiv*:2305.06563, 2023, to be published.

18. Wu, P.-L., M. Ding, and Y.-B. Zheng, Spatiotemporal traffic data imputation by synergizing low tensor ring rank and nonlocal subspace regularization. *IET Intelligent Transport Systems*, Vol. 17, No. 9, 2023, pp. 1908–1923.

19. Chen, X., K. Wang, Z. Li, Y. Zhang, and Q. Ye, A novel nonconvex low-rank tensor completion approach for traffic sensor data recovery from incomplete measurements. *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, 2023, pp. 1–15.

20. Zhao, Z., L. Tang, M. Fang, X. Yang, C. Li, and Q. Li, Toward urban traffic scenarios and more: a spatio-temporal analysis empowered low-rank tensor completion method for data imputation. *International Journal of Geographical Information Science*, Vol. 37, No. 9, 2023, pp. 1936–1969.

21. Zhang, X., Y. Zhao, S. Wang, Y. Sun, and B. Yin, A tensorial weighted Schatten-p norm model with neighbor regularization for traffic data completion and traffic system correlation exploration. *Neurocomputing*, Vol. 559, 2023, p. 126765.

22. Wright, S. J., R. D. Nowak, and M. A. Figueiredo, Sparse reconstruction by separable approximation. *IEEE Transactions on signal processing*, Vol. 57, No. 7, 2009, pp. 2479 – 2493.

23. Cai, J.-F., E. J. Candès, and Z. Shen, A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, Vol. 20, No. 4, 2010, pp. 1956 – 1982.

24. Li, D., Y. Teng, X. Zhou, J. Zhang, W. Luo, B. Zhao, Z. Yu, and L. Yuan, A tensor-based approach to unify organization and operation of data for irregular spatio-temporal fields. *International Journal of Geographical Information Science*, Vol. 36, No. 9, 2022, pp. 1885–1904.

25. Feng, X., H. Zhang, C. Wang, and H. Zheng, Traffic data recovery from corrupted and incomplete observations via spatial-temporal TRPCA. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 10, 2022, pp. 17835–17848.

26. Gong, W., Z. Huang, and L. Yang, LSPTD: Low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain, 2023, pp. 460–465.