# ROBUST LOW-RANK TENSOR DECOMPOSITION FRAMEWORK FOR TRANSPORT DATA IMPUTATION

**Cheng Lyu, Corresponding Author**
Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany
Email: cheng.lyu@tum.de

**Qing-Long Lu**
Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany
Email: qinglong.lu@tum.de

**Xinhua Wu**
Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA.
Email: wu.xinh@northeastern.edu

**Constantinos Antoniou, Ph.D.**
Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany
Email: c.antoniou@tum.de

Word Count: 5824 words + 5 table(s) $\times$ 250 = 7074 words

Submission Date: September 5, 2024

**ABSTRACT**

Missing values are prevalent in spatio-temporal transport data, undermining the quality of data-driven analysis. While prior works have demonstrated the promise of tensor completion methods for imputation, their performance remains limited for complicated composite missing patterns. This paper proposes a novel imputation framework combining tensor decomposition and rank minimization, which is effective in capturing key traffic dynamics and eliminates the need for exhaustive rank tuning. The framework is further supplemented with time series decomposition to account for trends, spatio-temporal correlations, and outliers, with the intention of improving the robustness of imputation results. A proximal ADMM algorithm is designed to solve the resulting multi-block nonconvex optimization efficiently. Experiments on four real-world transport datasets suggest that the proposed framework outperforms state-of-the-art imputation methods, especially in the context of complex missing patterns with high missing rates. These results underscore the potential benefits of incorporating temporal characteristics for more reliable imputation. Additionally, a sensitivity analysis provides initial evidence of the model's robustness in relation to hyperparameters.

*Keywords*: Missing data imputation, tensor completion, tensor decomposition, rank minimization, ADMM

## INTRODUCTION

Transport data, enriched by contemporary data collection techniques and open data initiatives, propels a multitude of applications including traffic forecasting, traffic state estimation, and traffic flow analysis (*1, 2*). These data-driven studies have enabled improved decision-making for the efficiency, safety, and sustainability of urban transport systems. However, the integrity of transport data is often undermined by missing values, particularly in time series data like traffic state, public transport ridership, and shared mobility usage. Even with the progression of transportation management systems, the issue of missing data is still not rare. For instance, a considerable number of traffic sensors with over half of their readings missing were reported by Laña et al. (*3*), where only 11% of sensors in the study area demonstrated more than 98% data completeness. Another evidence from bicycle volume data indicated that only 54% of the data is viable for use (*4*). These issues pose significant challenges for time series modeling, leading to a compromise in terms of analysis and prediction accuracy (*5*). Hence, devising accurate data imputation methods becomes essential to address the challenges of missing values in transport data.

Transport data can be corrupted for various reasons, including communication failure, sensor failure, infrastructure upgrading, and database failure, resulting in different missing patterns. Notably, missing data does not necessarily indicate the absence of entries. As summarized in *AASHTO Guidelines for Traffic Data Programs*, erroneous records, such as repeated, extreme, or error-coded values, also represent missing data as they are unusable for further analysis (*6*). One common missing data scenario in transport pertains to occasional random missing entries, which can generally be addressed by simple methods like temporal interpolation and spatial weighting. However, structured missing entries, characterized by missing data in the form of spatial and temporal clusters, pose more significant challenges for imputation methods. It was early emphasized by Smith et al. (*7*) that missing entries in transport data can happen in both temporal and spatial domains, with varying time spans of missing values.

A wide collection of transport data is spatio-temporal data, which can be naturally organized as matrices spanning space and time. The periodicity of transport time series further allows wrapping the temporal dimension into higher number of dimensions like time-of-the-day and day-of-the-week. Given such inherent tensorial structure, tensor/matrix decomposition and completion have emerged as one of the most promising solutions for transport data imputation, offering better imputation accuracy and higher computational efficiency over other methods like simple interpolation and statistical learning. Initially developed within the domain of computer vision and recommendation system (*8, 9*), these techniques are capable of modeling intricate multilinear relationship and filling in missing data by exploiting of the low-rank property of transport data. They have proven effective for understanding urban mobility dynamics (*10*) and improving imputation accuracy across various missing patterns (*11–13*).

Nevertheless, the performance of many existing tensor-based imputation methods are still limited when applied on spatio-temporal transport data with complex missing patterns. Similar to the cases in image recovery (*14*), some imputation methods have been found to falter faced with composite missing patterns, especially when spatial and temporal missing happens simultaneously. Additionally, data outliers—especially extreme values—can incur instability in imputation results, while variations in traffic dynamics due to supply-side changes, such as fleet expansion or infrastructure damage, may also adversely affect imputation accuracy. To develop an effective imputation method for practical transport data, a significant challenge lies in enhancing the robustness to handle various missing patterns, outliers, and temporal pattern shifts. As such, bridging the

gap between existing tensor methods and the multifaceted real-world complexity remains an open research problem.

## Literature Review

The effective imputation of missing values in transport data primarily revolves around exploiting temporal patterns and spatial similarities. These ideas are often realized through interpolation of observed values, fitting a statistical learning model, and optimizing a priori objective. A broad range of methodologies adhering to these ideas have been explored over the years, falling into three general paradigms: simple interpolation, machine learning, and tensor-based methods.

Typical simple interpolation methods include historical averaging, and weighted averaging across neighboring time periods and sensors (*7*). While straightforward to implement, these methods are only effective for isolated missing entries due to their strong dependency on historical or surrounding data. The imputation quality can be improved via parametric models like autoregressive models (*15*), which can relieve the difficulty in determining weights for averaging and reduce the impacts of outliers.

To address complicated missing patterns, recent research resort to machine learning models because of their strong predictive power. One line of research leverages clustering to extract temporal patterns, thereby facilitating missing entry imputation (*16*). Clustering can also be integrated with classification models to counteract its limitation with whole-day missing scenarios (*3*). The importance of explicit temporal modeling, e.g., the Prophet model (*17*) and Gaussian process with a periodical kernel (*18*), is later acknowledged for handling complex missing scenarios. However, these methods can become limited for either complicated workflows or high computational complexity. Hence, deep learning, which has seen significant success in traffic forecasting, has been recently introduced for imputation. Noteworthy methods include denoising stacked autoencoder (*19*), graph convolutional neural network (*20*), and attention mechanism (*21*). The common idea behind them is to encode the spatio-temporal patterns in latent spaces or a memory module. However, generating appropriate training samples is non-trivial. Most deep learning-based models use a sliding window strategy to prepare samples, which are short time series segments ranging from 15 minutes to one day. Such small window sizes restrict the contextual information input to the model, limiting their applicability in scenarios with longer missing periods.

Tensor-based methods provide a streamlined imputation formulation by exploiting the inherent low-rank property and multi-dimensional relationships present in transport data. One approach is tensor decomposition, which breaks transport data into several latent factors; it degenerates into matrix decomposition when the transport data is two dimensional (*12*). Each decomposed factor encodes the information of a specific dimension, indicating key time series patterns and correlations among sensors. Missing value imputation can be realized by minimizing the reconstruction error from factors (*11*) or maximizing the posterior likelihood in the Bayesian context (*13*). Regularization can also be applied to enforce spatial and temporal smoothness (*22*). However, a significant limitation of tensor decomposition is the need for predefined ranks as hyperparameters, which are hard to determine in practice. Tensor completion, on the other hand, seeks a low-rank approximation of transport data by directly minimizing the rank of the reconstructed tensor. Given that tensor rank is non-differentiable, many studies focused on finding a viable rank approximation, e.g., truncated nuclear norm (*23*), Schatten-$p$ norm (*24*), and tubal rank (*25*). However, tensor rank and its approximations are all permutation invariant; thus, the imputation may fail in the absence of time series modeling. Integrating tensor completion with autoregressive models has proven to

be an effective solution to this problem (*26*).

### Objectives and Contribution

This study, following the paradigm of tensor-based method, aims to design a new transport data imputation framework based on tensor decomposition, which is robust for various composite missing patterns. Our contribution can be summarized into the following points:

- The framework leverages tensor decomposition techniques combined with rank minimization, effectively eliminating the need for predefined ranks and enhancing the accuracy of imputation.
- To account for outliers and temporal pattern shifts, the framework is embedded with a time series decomposition model, enabling effective modeling of temporal evolution in transport data.
- An improved alternating direction method of multiplier (ADMM) algorithm is designed to solve the multi-block nonconvex separable problem in the proposed framework.

This paper is structured as follows. In Section 2, we present the fundamentals of tensors and establish the problem definition. Section 3 details our novel transport data imputation framework. In Section 4, we present the experimental results, demonstrating the effectiveness of our approach. Finally, in Section 5, we provide conclusion and outline potential future research directions.

## PRELIMINARIES

### Notations

In this paper, scalars are represented by italic letters, vectors by boldface lowercase letters, matrices by boldface uppercase letters, and tensors by calligraphic letters, for example, $x$, $\boldsymbol{x}$, $\boldsymbol{X}$, and $\mathcal{X}$, respectively.

A tensor with $k$ modes, also referred to as a $k$-way tensor or a $k$-dimensional tensor, is represented as $\mathcal{X} \in \mathbb{R}^{N_1,N_2,\ldots,N_k}$, where $N_n$ indicates the size of the $n$-th mode. Subscripts added to a tensor, such as $x_{i_1 i_2 \ldots i_k}$, indicate the $(i_1, i_2, \ldots, i_k)$-th entry of the tensor.

In reference to the transport data discussed in this paper, we begin with a generic matrix representation, $\boldsymbol{X} \in \mathbb{R}^{N_s \times N_t}$, that includes a time series of measurements over a specific period gathered from a range of sensors. Here, $N_s$ represents the number of sensors and $N_t$ refers to the number of time slots. This matrix structure can be reorganized into a tensor structure by reshaping along the temporal mode. For example, a four-way tensor, $\mathcal{X} \in \mathbb{R}^{N_s \times N_w \times N_d \times N_m}$, can be obtained by reshaping the temporal mode into weeks and days-of-the-week, where $N_w$, $N_d$, and $N_m$ indicates the number of weeks, days in a week, and time slots in a day, respectively. It is worth noting that, when $N_t$ is not divisible by the product of $N_w$, $N_d$ and $N_m$, the matrix can be padded with null values in the temporal mode to allow proper reshaping. Also, more temporal modes, such as month-of-the-year, can be incorporated to transform the matrix into a higher-order tensor when dealing with data spanning a long time period.

### Tensor Basics

A tensor can be converted from and to a matrix through the folding and unfolding operations along its $n$-th mode,

$$\mathcal{X} = \text{fold}_n(\boldsymbol{X}_{(n)}; N_1, \ldots, N_k) \in \mathbb{R}^{N_1,N_2,\ldots,N_k}, \tag{1}$$

$$\boldsymbol{X}_{(n)} = \text{unfold}_n(\mathcal{X}) \in \mathbb{R}^{N_n \times (N_1 \times \cdots \times N_{n-1} \times N_{n+1} \times \cdots \times N_k)}. \tag{2}$$

Similar to a matrix, a tensor can be characterized by a range of different norms. The $L^1$ norm of a tensor $\|\cdot\|_1$ can be defined as the absolute sum of all its entries; and the Frobenius norm of a tensor $\|\cdot\|_F$ can be defined as the square root of the squared sum of all its entries,

$$\|\mathcal{X}\|_1 = \sum_{i_1,i_2,\ldots,i_k} |x_{i_1,i_2,\ldots,i_k}|, \quad \|\mathcal{X}\|_F = \sqrt{\sum_{i_1,i_2,\ldots,i_k} x^2_{i_1,i_2,\ldots,i_k}}. \tag{3}$$

Tensor decomposition refers to decomposing a tensor into several factors. In this paper, we focus on Tucker decomposition, which can factorize a tensor into the $k+1$ components, including a core tensor and $k$ factor matrices (*27*),

$$\mathcal{X} = \mathcal{G} \times_1 \boldsymbol{U}_1 \times_2 \cdots \times_k \boldsymbol{U}_k. \tag{4}$$

where $\mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_k}$ is the core tensor and $\boldsymbol{U}_n \in \mathbb{R}^{N_n \times r_n} (n \in [k])$ are the factor matrices, respectively. $r_1, \ldots, r_k$ are the rank of factor matrices. The operator $\times_n$ is the $n$-mode product.

Given an arbitrary tensor $\mathcal{A} \in \mathbb{R}^{r_1 \times \cdots \times r_k}$ and a matrix $\boldsymbol{U} \in \mathbb{R}^{N_n \times r_n}$, their $n$-mode product, denoted by $\mathcal{B} \in \mathbb{R}^{r_1 \times \cdots \times r_{n-1} \times N_n \times r_{n+1} \times \cdots \times r_k}$, can be defined by:

$$\mathcal{B} = \mathcal{A} \times_n \boldsymbol{U} \Leftrightarrow \mathcal{B} = \text{fold}_n(\boldsymbol{U}\boldsymbol{A}_{(n)}). \tag{5}$$

The definition can also be written using an element-wise expression, $b_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_k} = \sum_{i_n} a_{i_1 \ldots i_k} u_{j i_n}$. Likewise, the Tucker decomposition can be written element-wisely:

$$x_{i_1 \ldots i_k} = \sum_{j_1=1}^{r_1} \cdots \sum_{j_k=1}^{r_k} g_{j_1 \ldots j_k} u_{1,i_1 j_1} \cdots u_{k,i_k j_k}. \tag{6}$$

One natural definition of tensor rank coming with Tucker decomposition is the tensor $n$-rank, defined as the rank of $n$-mode unfolding matrix of a tensor. The Tucker rank is then defined as the set of $n$-rank of all unfolding matrices $(\text{rank}_{(1)}(\mathcal{X}), \text{rank}_{(2)}(\mathcal{X}), \ldots, \text{rank}_{(k)}(\mathcal{X}))$, where

$$\text{rank}_{(n)}(\mathcal{X}) = \text{rank}(\boldsymbol{X}_{(n)}). \tag{7}$$

## Problem Statement

The main goal of transport data imputation is to estimate the missing values in a transport tensor using the observed values. Formally, given an incomplete tensor $\mathcal{Y}_\Omega = \mathcal{Y} \odot \Omega \in \mathbb{R}^{N_1 \times \cdots \times N_k}$, where $\mathcal{Y}$ is the unknown ground truth tensor and $\Omega \in \{0,1\}^{N_1 \times \cdots \times N_k}$ is the binary mask tensor, the task is to obtain a restored tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_k}$.

# IMPUTATION FRAMEWORK

In this section, we first present the idea of using a time series decomposition model as the underlying temporal model of transport data. Then, based on the time series decomposition, we propose a new optimization model combining tensor decomposition and rank minimization to enable robust imputation of missing values.

## Time Series Decomposition

In order to effectively model the temporal evolution of time series in transport data, an additive time series decomposition is formulated and later embedded in the proposed framework. We base the rationale behind the decomposition on the primary characteristics of transport data derived from the example traffic volume data collected by a loop detector, as shown in Figure 1. (i) An apparent drop in traffic volume can be noticed around Aug 15, 2019, which is not observed elsewhere. This may signify a long-term supply-side change, such as road construction. Another sharp volume drop is evident around Oct 7, 2019, which quickly recovers to the usual level, possibly due to a temporary lane closure. (ii) Daily periodicity can be easily observed from the plot. (iii) A few

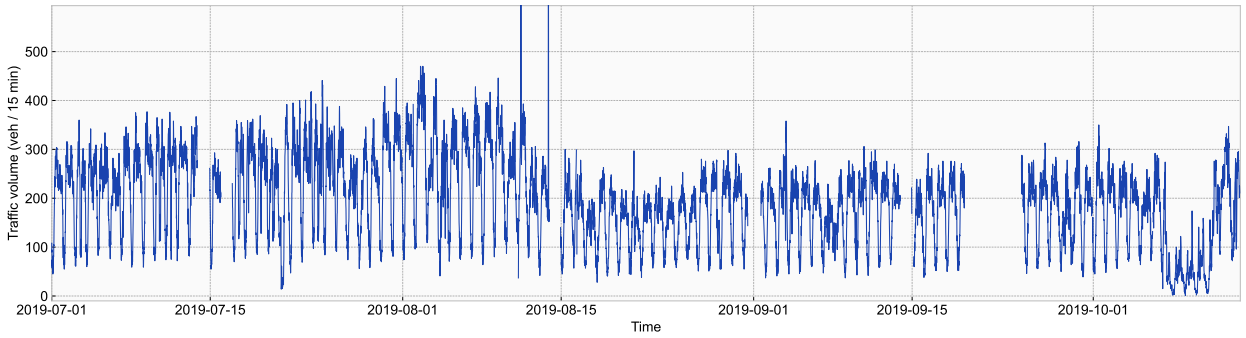



**FIGURE 1**: An example time series of traffic volume data collected by a loop detector in London. The blank spaces in the figure indicate missing values.

spurious extreme values can be seen around Aug 11, 2019 and Aug 14, 2019, which need to be carefully handled by the imputation model.

Based on the observations above, we decompose the time series into three components, namely trend, seasonality and error. The trend component assumes a locally constant trend to provide a stable prior for the imputation. The seasonal component captures the periodicity of the time series and the correlations among sensors, aligning with the low rank property of the transport data tensor. The error component accounts for the occasional outliers in the time series. To sum up, a transport data tensor can be expressed as the sum of the following three components,

$$\mathcal{X} = \mathcal{T} + \mathcal{S} + \mathcal{E}, \tag{8}$$

where $\mathcal{T}$, $\mathcal{S}$, and $\mathcal{E}$ denote the trend tensor, seasonality tensor, and error tensor respectively.

An illustration of the time series decomposition is demonstrated in Figure 2. The trend tensor is established through a preliminary step of changepoint detection, such that major changes in temporal evolution, e.g., the volume drop in Figure 1, can be identified. In changepoint detection, we hold the assumption that changepoints occur independently across sensors. Therefore, the tensor is first unfolded back to a matrix, and changepoint detection is performed on time series of each sensor individually. The locations of changepoints in time series are examined using the pruned exact linear time (PELT) algorithm (*28*), where the homogeneity of each segment is measured using the Gaussian kernel function (*29*). The identified changepoints are then used to partition the time series into a sequence of segments. Within each segment, we assume a constant trend component, in an effort to enhance the robustness of imputation results.

The seasonal component of the time series decomposition is managed using a Tucker decomposition model, which factorizes the low-rank seasonal tensor into a core tensor and factor matrices, allowing us to effectively capture the spatio-temporal correlations in the data, including the temporal periodicity and cross-sensor correlations. Details of addressing Tucker decomposition will be explained in the subsequent subsection. Finally, the additional error component is responsible to manage the occasional outliers in the time series, thereby mitigating their impacts on the imputation results.

### Optimization Problem

To obtain a proper imputation of the missing values, the objective of tensor-based methods is often either minimizing the reconstruction error in the case of tensor decomposition, or minimizing the tensor rank when working on the full-sized tensor. Based on the time series decomposition

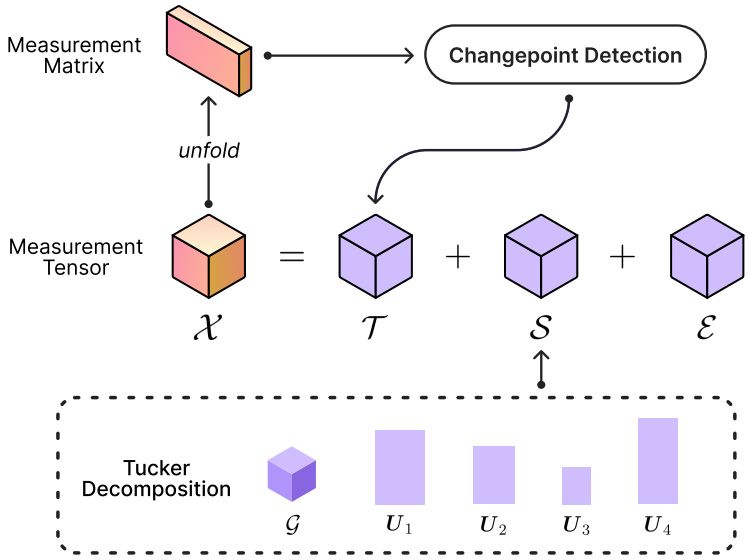



**FIGURE 2**: An overview of the time series decomposition. Note that tensors are illustrated as 3-dimensional cubes merely for demonstration purposes. They can be of higher modes to incorporate higher dimensional information.

model (Equation 8), one may want to directly minimize the error component because of the tensor decomposition. However, as reviewed in *Introduction*, the difficulty in determining appropriate ranks for Tucker decomposition is one of its major limitations. It is reasonable to apply additional penalty on large Tucker ranks in order to relieve the burden of rank determination. Additionally, it was suggested by Goulart et al. (*30*) that imposing parsimony on the core tensor can help alleviate the adverse effects of misspecified ranks. Thus, the following objective is designed to tackle the aforementioned limitation,

$$f(\{\boldsymbol{U}_i\},\mathcal{G},\mathcal{E}) = \|\mathcal{E}\|_1 + \mu\|\mathcal{G}\|_1 + \lambda\sum_{i=1}^{k}\text{rank}(\boldsymbol{U}_i) + \frac{\xi}{2}\sum_{i=2}^{k}\|\boldsymbol{D}_i\boldsymbol{U}_i\|_F^2. \tag{9}$$

The objective function consists of four components, including the $L^1$ norm of the error term, the $L^1$ norm of the core tensor, the ranks of factor matrices, and the total variation (TV) regularization. The $L^1$ regularization of both the error term and the core tensor aims to obtain a sparse solution. The last TV regularization term aims to stabilize the tensor completion result by applying a smoothness prior on factor matrices (*31*) with the help of the difference matrix:

$$\boldsymbol{D}_i = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{N_i \times N_i}, \tag{10}$$

In addition, the TV regularization is only imposed on the temporal factors, as the first mode, i.e., the mode of sensors, is not necessarily smooth in the local neighborhood.

Most tensor completion methods based on rank minimization aim to minimize the sum of the ranks of all unfolding matrices, i.e., $\sum_i \text{rank}(\boldsymbol{S}_{(i)})$ (*8, 23, 24*). However, with tensor decomposition in the optimization problem, optimizing over unfolding matrices introduces additional

difficulty in designing solution algorithm. To circumvent this issue, we directly optimize over the sum of the ranks of all factor matrices, denoted as $\sum_i \text{rank}(\boldsymbol{U}_i)$, which has been shown to be equivalent to minimizing the sum of ranks of unfolding matrices (*32*).

To ensure the objective function is differentiable, a smooth surrogate for matrix rank is necessary (*24, 33*). Given that the matrix rank is equal to the number of non-zero values of its singular values, approximations are usually designed in the direction of finding a better surrogate for the $L^0$ norm. Compared with nuclear norm, which is the convex envelop of matrix rank, nonconvex surrogates, such as truncated nuclear norm and Schatten-$p$ norm, have been shown to provide superior approximations. In our formulation, we employ the $\gamma$-norm for a better approximation,

$$\|\boldsymbol{U}\|_\gamma = \sum_i \frac{(1+\gamma)\sigma_i(\boldsymbol{U})}{\gamma+\sigma_i(\boldsymbol{U})}, \tag{11}$$

where $\sigma_i(\boldsymbol{U})$ denote the $i$-th singular value of factor matrix $\boldsymbol{U}$, and $\gamma$ is a shape parameter.

Substitute the matrix rank in Equation 9 by the $\gamma$-norm above, and the complete optimization problem can be formulated as,

$$\min_{\{\boldsymbol{U}_i\},\mathcal{G},\mathcal{E},\boldsymbol{c},\mathcal{X}} \|\mathcal{E}\|_1 + \mu\|\mathcal{G}\|_1 + \lambda\sum_{i=1}^{k}\|\boldsymbol{U}_i\|_\gamma + \frac{\xi}{2}\sum_{i=2}^{k}\|\boldsymbol{D}_i\boldsymbol{U}_i\|_F^2 \tag{12}$$

$$\text{s.t.} \quad \mathcal{X} = \sum_i c_i \,\text{fold}_1(\boldsymbol{\Psi}_i) + \mathcal{G}\times_1\boldsymbol{U}_1\times_2\cdots\times_k\boldsymbol{U}_k + \mathcal{E} \tag{13}$$

$$\mathcal{X}_\Omega = \mathcal{Y}_\Omega \tag{14}$$

where the first constraint is the time series decomposition of the transport data tensor, as outlined in Equation 8, where $\mathcal{T} = \sum_i c_i \,\text{fold}_1(\boldsymbol{\Psi}_i)$ and $\mathcal{S} = \mathcal{G}\times_1\boldsymbol{U}_1\times_2\cdots\times_k\boldsymbol{U}_k$. The trend term $\mathcal{T}$ connects the changepoints identified by the PELT algorithm with the decision variables $\boldsymbol{c}$. Each time series segment partitioned by changepoints is assigned a variable $c_i$, which represents the corresponding long-term trend. Additionally, a binary masking matrix, $\boldsymbol{\Psi}_i \in \{0,1\}^{N_s\times N_t}$, is aligned with each segment. The seasonal term $\mathcal{S}$ here is substituted by the Tucker decomposition. The last constraint ensures that the values of all observed entries are identical to the imputation results.

## SOLUTION ALGORITHM

While the optimization problem above with two equality constraints can be solved using gradient descend by absorbing the constraints into the objective function, it can be computationally inefficient. Fortunately, it is feasible to separate the objective as the sum of several decoupled functions, allowing acceleration with ADMM (*34*). The basic idea behind ADMM is to decompose the problem into smaller sub-problems, each of which can be solved independently and iteratively.

### Problem Separation

To enable successful separation, auxiliary variables $\{\boldsymbol{V}_i\}_{i=1}^k$ are introduced to decouple the third and last components in the objective, resulting in the modified problem,

$$\min_{\{\boldsymbol{U}_i,\boldsymbol{V}_i\},\mathcal{G},\mathcal{E},\boldsymbol{c},\mathcal{X}} \|\mathcal{E}\|_1 + \mu\|\mathcal{G}\|_1 + \lambda\sum_{i=1}^{k}\|\boldsymbol{V}_i\|_\gamma + \frac{\xi}{2}\sum_{i=2}^{k}\|\boldsymbol{D}_i\boldsymbol{U}_i\|_F^2 \tag{15}$$

$$\text{s.t.} \quad \mathcal{X} = \sum_i c_i \,\text{fold}_1(\boldsymbol{\Psi}_i) + \mathcal{G}\times_1\boldsymbol{U}_1\times_2\cdots\times_k\boldsymbol{U}_k + \mathcal{E} \tag{16}$$

$$\mathcal{X}_\Omega = \mathcal{Y}_\Omega \tag{17}$$

$$\boldsymbol{U}_i = \boldsymbol{V}_i \tag{18}$$

The problem is now separable with four decoupled blocks in the objective. Denote the

Lagrangian multipliers by $\mathcal{M}$ and $\mathbf{\Gamma}$, the augmented Lagrangian is given as follows:

$$L(\{\boldsymbol{U}_i,\boldsymbol{V}_i,\boldsymbol{\Gamma}_i\},\mathcal{G},\mathcal{E},\boldsymbol{c},\mathcal{X},\mathcal{M}) = \|\mathcal{E}\|_1+\mu\|\mathcal{G}\|_1+\lambda\sum_{i=1}^{k}\|\boldsymbol{V}_i\|_\gamma+\frac{\xi}{2}\sum_{i=2}^{k}\|\boldsymbol{D}_i\boldsymbol{U}_i\|_F^2$$
$$+\frac{\rho}{2}\|\mathcal{X}-\sum_i c_i\,\text{fold}_1(\boldsymbol{\Psi}_i)-\mathcal{G}\times_1\boldsymbol{U}_1\times_2\cdots\times_k\boldsymbol{U}_k-\mathcal{E}+\mathcal{M}\|_F^2 \tag{19}$$
$$+\sum_{i=1}^{k}\frac{\rho}{2}\|\boldsymbol{U}_i-\boldsymbol{V}_i+\boldsymbol{\Gamma}_i\|_F^2.$$

**Proximal ADMM Algorithm**

While the problem has become separable after modification, the convergence of standard ADMM, which is initially designed for two-block problems, is not guaranteed when extended problems with a higher number of blocks (*35*). To overcome this challenge, we incorporate an additional proximal term in the sub-problems, which has been demonstrated to effectively resolve this issue (*36, 37*).

*Solving Sub-problem of* $\boldsymbol{U}_i$

We begin with the first factor matrix $\boldsymbol{U}_1$, and the sub-problem can be described as:

$$\boldsymbol{U}_1^{(t+1)} = \arg\min_{\boldsymbol{U}_1} L(\{\boldsymbol{U}_{i\neq1}^{(t)},\boldsymbol{V}_i^{(t)},\boldsymbol{\Gamma}_i^{(t)}\},\boldsymbol{U}_1,\mathcal{G}^{(t)},\mathcal{X}^{(t)},\mathcal{E}^{(t)},\boldsymbol{c}^{(t)},\mathcal{M}^{(t)})+\frac{\eta}{2}\|\boldsymbol{U}_1-\boldsymbol{U}_1^{(t)}\|_F^2, \tag{20}$$

where $\eta$ is a penalty parameter of the proximal term. All variables excluding $\boldsymbol{U}_1$ are fixed, resulting in a convex problem. The solution can be easily derived by letting the gradient be zero:

$$\boldsymbol{U}_1^{(t+1)} = (\rho\boldsymbol{Z}_{(1)}\boldsymbol{P}_1(\boldsymbol{G}_{(1)}^{(t)})^\top-\rho\boldsymbol{Q}_1-\eta\boldsymbol{U}_1^{(t)})(\rho\boldsymbol{G}_{(1)}^{(t)}\boldsymbol{P}_1^\top\boldsymbol{P}_1(\boldsymbol{G}_{(1)}^{(t)})^\top-(\rho+\eta)\boldsymbol{I})^{-1}, \tag{21}$$

where $\boldsymbol{P}_1=\boldsymbol{U}_2^{(t)}\otimes\cdots\otimes\boldsymbol{U}_k^{(t)}$ and $\boldsymbol{Q}_1=\boldsymbol{V}_1^{(t)}-\boldsymbol{\Gamma}_1^{(t)}$. The operator $\otimes$ here denotes the Kronecker product.

Similarly, the sub-problems of other factor matrices, which involve an additional TV regularizer, can be simplified to solving a Sylvester equation:

$$\boldsymbol{A}\boldsymbol{U}_i+\boldsymbol{U}_i\boldsymbol{B}=\boldsymbol{R}, \tag{22}$$

where $\boldsymbol{A}=\rho\boldsymbol{I}+\xi\boldsymbol{D}_i^\top\boldsymbol{D}_i$, $\boldsymbol{B}=-\eta\boldsymbol{I}-\rho\boldsymbol{G}_{(i)}^{(t)}\boldsymbol{P}_2^\top\boldsymbol{P}_i(\boldsymbol{G}_{(i)}^{(t)})^\top$, and $\boldsymbol{R}=\rho\boldsymbol{Q}_i+\eta\boldsymbol{U}_i^{(t)}-\rho\boldsymbol{Z}_{(i)}\boldsymbol{P}_2(\boldsymbol{G}_{(i)}^{(t)})^\top$. Herein, $\boldsymbol{Z}_{(i)}=\boldsymbol{X}_{(i)}^{(t)}-\boldsymbol{T}_{(i)}^{(t)}-\boldsymbol{E}_{(i)}^{(t)}+\boldsymbol{M}_{(i)}^{(t)}$, $\boldsymbol{P}_i=\bigotimes_{j\neq i}\boldsymbol{U}_j^{(t)}$, and $\boldsymbol{Q}_i=\boldsymbol{V}_i^{(t)}-\boldsymbol{\Gamma}_i^{(t)}$. The Sylvester equation can be solved using the Bartels-Stewart algorithm (*38*).

*Solving Sub-problem of* $\boldsymbol{V}_i$

Let $\boldsymbol{W}_i=\frac{1}{\rho+\eta}(\rho(\boldsymbol{U}_i^{(t+1)}+\boldsymbol{\Gamma}_i^{(t)})+\eta\boldsymbol{V}_i^{(t)})$, the $\boldsymbol{V}_i$ sub-problem can be simplified as:

$$\boldsymbol{V}_i^{(t+1)} = \arg\min_{\boldsymbol{V}_i}\frac{\lambda}{\rho+\eta}\|\boldsymbol{V}_i\|_\gamma+\frac{1}{2}\|\boldsymbol{V}_i-\boldsymbol{W}_i\|_F^2, \tag{23}$$

where the objective is the sum of $\gamma$-norm and Frobenius norm.

Note that non-convex problems in similar forms can usually be solved using various thresholding algorithms, such as soft thresholding and singular value thresholding (*26, 30, 39*). Here, we manage to solve this problem by relaxing it to a weighted $\ell_1$-minimization problem to make it easily tractable.

**Lemma 1.** *Let* $\sigma_i$ *denote the i-th singular value of a matrix* $\boldsymbol{V}$. *For any concave antimonotone*

*function $g(\cdot)$, the function $h(\boldsymbol{V},\boldsymbol{W}) = \frac{\lambda}{\rho+\eta}\sum_{i=1}^{k} g(\sigma_i) + \frac{1}{2}\|\boldsymbol{V}-\boldsymbol{W}\|_F^2$ can be relaxed to:*

$$\tilde{h}(\boldsymbol{V},\boldsymbol{W}) = \frac{\lambda}{\rho+\eta}\sum_{i=1}^{k} g(\sigma_i^{(t)}) + \nabla g(\sigma_i^{(t)})(\sigma_i^{(t)} - \sigma_i) + \frac{1}{2}\|\boldsymbol{V}-\boldsymbol{W}\|_F^2 \tag{24}$$

$$= \frac{\lambda}{\rho+\eta}\sum_{i=1}^{k} \nabla g(\sigma_i^{(t)})\sigma_i + \frac{1}{2}\|\boldsymbol{V}-\boldsymbol{W}\|_F^2. \tag{25}$$

With Lemma 1, the $\gamma$-norm can be relaxed to the weighted $\ell_1$-norm of $\boldsymbol{V}_i$. Then, the $\boldsymbol{V}_i$ sub-problem can be solved by the weighted singular thresholding (WST) algorithm (*39*). Let the singular value decomposition (SVD) of $\boldsymbol{W}_i$ be $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^\top$, the update rule is given by,

$$\boldsymbol{V}_i^{(t+1)} = \boldsymbol{A}\mathfrak{T}_{\frac{\lambda}{\rho+\eta}\boldsymbol{\omega}}(\boldsymbol{\Sigma})\boldsymbol{B}^\top, \tag{26}$$

where $\omega_i = \nabla g(\sigma_i^{(t)}) = \nabla \|\boldsymbol{V}_i\|_\gamma = (1+\gamma)\gamma/(\gamma+\sigma_i^{(t)})^2$. The WST operator can be defined by $\mathfrak{T}_{\frac{\lambda}{\rho+\eta}\boldsymbol{\omega}}(\boldsymbol{\Sigma}) = \mathrm{diag}((\Sigma_{jj} - \frac{\lambda}{\rho+\eta}\omega_j)_+)$, and $(\cdot)_+ = \max(\cdot, 0)$.

*Solving Sub-problem of $\mathcal{G}$*

Denote $\mathcal{Z} = \mathcal{X}^{(t)} - \mathcal{T}^{(t)} - \mathcal{E}^{(t)} + \mathcal{M}^{(t)}$, the sub-problem of $\mathcal{G}$ can be written as:

$$\mathcal{G}^{(t+1)} = \arg\min_{\mathcal{G}} \mu\|\mathcal{G}\|_1 + \frac{\rho}{2}\|\mathcal{G}\times_1 \boldsymbol{U}_1^{(t+1)} \times_2 \cdots \times_k \boldsymbol{U}_k^{(t+1)} - \mathcal{Z}\|_F^2 + \frac{\eta}{2}\|\mathcal{G}-\mathcal{G}^{(t)}\|_F^2. \tag{27}$$

Direct analysis with $n$-mode product is inconvenient. Hence, the objective of this sub-problem is rewritten as matrix operations using Equation 5. Let us define:

$$\phi(\mathcal{G}) = (\rho\|\mathcal{G}\times_1 \boldsymbol{U}_1^{(t+1)} \times_2 \cdots \times_k \boldsymbol{U}_k^{(t+1)} - \mathcal{Z}\|_F^2 + \eta\|\mathcal{G}-\mathcal{G}^{(t)}\|_F^2)/(2\mu), \tag{28}$$

$$\varphi(\boldsymbol{G}_{(1)}) = (\rho\|\boldsymbol{U}_1^{(t+1)}\boldsymbol{G}_{(1)}\boldsymbol{P}_1^\top - \boldsymbol{Z}_{(1)}\|_F^2 + \eta\|\boldsymbol{G}_{(1)} - \boldsymbol{G}_{(1)}^{(t)}\|_F^2)/(2\mu), \tag{29}$$

where $\boldsymbol{P}_1 = \boldsymbol{U}_2^{(t+1)} \otimes \cdots \otimes \boldsymbol{U}_k^{(t+1)}$. The tensors are unfolded along the first mode without loss of generality. The sub-problem can then be rewritten as a general $\ell_1$-minimization problem:

$$\mathcal{G}^{(t+1)} = \arg\min_{\mathcal{G}} \|\mathcal{G}\|_1 + \phi(\mathcal{G}) \quad \Leftrightarrow \quad \boldsymbol{G}_{(1)}^{(t+1)} = \arg\min_{\boldsymbol{G}_{(1)}} \|\boldsymbol{G}_{(1)}\|_1 + \varphi(\boldsymbol{G}_{(1)}). \tag{30}$$

Considering that function $\varphi(\cdot)$ is differentiable and convex, the optimal solution can be given by the following soft thresholding operation (*40*):

$$\boldsymbol{G}_{(1)}^{(t+1)} = \mathrm{sgn}(\boldsymbol{H}) \odot (|\boldsymbol{H}| - \delta)_+, \tag{31}$$

where $\boldsymbol{H} = \boldsymbol{G}_{(1)}^{(t)} - \delta\nabla\varphi(\boldsymbol{G}_{(1)}^{(t)})$. We set the step size $\delta = 1/\|\boldsymbol{U}_1^{t+1}\|_2\|\boldsymbol{P}_1^\top\|_2$ as the reciprocal of the Lipschitz constant of $\nabla\varphi(\boldsymbol{G}_{(1)}^{(t)})$, as suggested by (*41*), where $\|\cdot\|_2$ denotes the spectral norm. The gradient $\nabla\varphi(\boldsymbol{G}_{(1)})$ can be computed based on Equation 29:

$$\nabla\phi(\boldsymbol{G}_{(1)}) = \frac{\rho}{\mu}(\boldsymbol{U}_1^{(t+1)})^\top(\boldsymbol{U}_1^{(t+1)}\boldsymbol{G}_{(1)}\boldsymbol{P}_1^\top - \boldsymbol{Z}_{(1)})\boldsymbol{P}_1 + \frac{\eta}{\mu}(\boldsymbol{G}_{(1)} - \boldsymbol{G}_{(1)}^{(t)}). \tag{32}$$

*Solving Other Sub-problems*

The $\mathcal{E}$ sub-problem is a standard $\ell_1$-minimization problem, which can be solved via soft thresholding (*34*). Denote $\mathcal{S} = \mathcal{G}\times_1 \boldsymbol{U}_1 \times_2 \cdots \times_k \boldsymbol{U}_k$, the update rule for $\mathcal{E}$ is,

$$\mathcal{E}^{(t+1)} = \mathrm{sgn}(\mathcal{B}) \odot \left(|\mathcal{B}| - \frac{1}{\rho+\eta}\right)_+, \tag{33}$$

where $\mathcal{B} = \frac{1}{\rho+\eta}\left(\rho(\mathcal{X}^{(t)} - \mathcal{T}^{(t)} - \mathcal{S} + \mathcal{M}^{(t)}) + \eta\mathcal{E}^{(t)}\right)$, and $\odot$ is the Hadamard product.

Closed-form solutions to the two convex sub-problems of $\boldsymbol{c}$ and $\mathcal{X}$ can both be easily obtained by letting gradient be zero:

$$\boldsymbol{c}^{(t+1)} = \frac{1}{\rho+\eta}\left(\rho\frac{1}{\|\boldsymbol{\Psi}_i\|_0}\mathrm{tr}(\boldsymbol{I}\,\mathrm{vec}(\mathrm{fold}_i(\boldsymbol{\Psi}_i)\odot(\mathcal{X}^{(t)}-\mathcal{S}-\mathcal{E}^{(t+1)}+\mathcal{M}^{(t)})))+\eta\boldsymbol{c}^{(t)}\right), \tag{34}$$

$$\mathcal{X}^{(t+1)} = \frac{1}{\rho+\eta}\left(\rho(\mathcal{T}^{(t+1)}+\mathcal{S}+\mathcal{E}^{(t+1)}-\mathcal{M}^{(t)})+\eta\mathcal{X}^{(t)}\right). \tag{35}$$

Finally, the overall algorithm is concluded by Algorithm 1. Before updating variables, we first initialize $\mathcal{X}$ using historical average. Then, $\mathcal{G}$ and all $\boldsymbol{U}_i$ are computed using higher-order SVD (*27*). The auxiliary variables $\boldsymbol{V}_i$ are set identical to $\boldsymbol{U}_i$. Other variables, including $\mathcal{E},\mathcal{M}$ and all $\boldsymbol{\Gamma}_i$, are initialized as zero.

**Algorithm 1** Proximal ADMM for Robust Transport Data Imputation

**Input**: Incomplete tensor $\mathcal{Y}_\Omega$; missing mask $\Omega$; hyperparameters $\mu,\lambda,\xi,\{r_i\},\rho,\eta$; convergence threshold $\varepsilon$.
**Output**: Imputed tensor $\mathcal{X}$.
Initialize primal variables and Lagrangian multipliers; set tolerance $\Delta\leftarrow\infty$; set $t\leftarrow 1$.
**repeat**
  **for** $i\leftarrow 1,k$ **do**
    Solve sub-problem of $\boldsymbol{U}_i$ using Equation 21 or Equation 22.
  **end for**
  **for** $i\leftarrow 1,k$ **do**
    Solve sub-problem of $\boldsymbol{V}_i$ using Equation 26.
  **end for**
  Solve sub-problem of $\mathcal{G}$ using Equation 31.
  Compute seasonal component $\mathcal{S}\leftarrow\mathcal{G}\times_1\boldsymbol{U}_1\times_2\cdots\times_k\boldsymbol{U}_k$.
  Solve sub-problem of $\mathcal{E}$ using Equation 33.
  Solve sub-problem of $\boldsymbol{c}$ using Equation 34.
  $\bar{\mathcal{X}}\leftarrow\mathcal{X}$. Solve sub-problem of $\mathcal{X}$ using Equation 35.
  Update Lagrangian multiplier $\mathcal{M}\leftarrow\mathcal{M}+\mathcal{X}-\mathcal{T}-\mathcal{S}-\mathcal{E}$.
  **for** $i\leftarrow 1,k$ **do**
    Update Lagrangian multiplier $\boldsymbol{\Gamma}_i\leftarrow\boldsymbol{\Gamma}_i+\boldsymbol{U}_i-\boldsymbol{V}_i$.
  **end for**
  $\Delta\leftarrow\|\mathcal{X}-\bar{\mathcal{X}}\|_F^2/\|\bar{\mathcal{X}}\|_F^2$.
**until** $\Delta<\varepsilon$

## EXPERIMENT SETTINGS

To evaluate the imputation performance of the proposed framework, four datasets are employed in this study. This section presents a detailed overview of the metadata and basic statistics of these datasets, as well as the preprocessing methods. Furthermore, we provide the details of the experimental settings, including the design of missing patterns, base models for comparison, and configuration of hyperparameters.

## Data Description

The datasets used in this study include one traffic speed data from Guangzhou (China), and three traffic volume data from London (UK), Madrid (Spain) and Melbourne (Australia), respectively[1]. To ensure the reliability of our analysis, we remove any traffic sensor data with less than 95% completeness, excluding time periods where all traffic detectors have no reading. For the traffic volume data, we select sensors located in the city centers, where typically experience higher traffic densities. Detailed metadata for all four datasets can be found in Table 1.

**TABLE 1**: Metadata of Datasets

| City | # Sensor | Duration | Interval | Measurement (Unit) |
|---|---|---|---|---|
| Guangzhou | 209 | Aug 1, 2016 — Sep 30, 2016 | 10 min | Speed (km/h) |
| London | 276 | Jul 1, 2019 — Oct 13, 2019 | 15 min | Volume (veh/15min) |
| Madrid | 324 | Jun 1, 2021 — Sep 13, 2021 | 15 min | Volume (veh/15min) |
| Melbourne | 313 | Jun 1, 2020 — Sep 13, 2020 | 15 min | Volume (veh/15min) |

The singular values of the data matrices of all four cities are demonstrated in Figure 3. The dominance of large singular values across all datasets is evident, substantiating the rationality of exploiting the low-rank property for missing value imputation. Furthermore, for the proposed framework, the data matrices are organized into four-way tensors $\mathcal{Y} \in \mathbb{R}^{N_s \times N_w \times N_d \times N_m}$ with the structure of sensor $\times$ week $\times$ day-of-the-week $\times$ time-of-the-day.

## Missing Patterns

Contrary to the majority of studies, which typically create individual missing scenarios for each pattern, our experiment comprises a composite missing mask by integrating four fundamental missing patterns. The mixing-up of missing patterns can be expressed formally as:

$$\Omega = \Omega_{\text{BM}} \odot \Omega_{\text{RM}} \odot \Omega_{\text{DM}} \odot \Omega_{\text{TM}}, \tag{36}$$

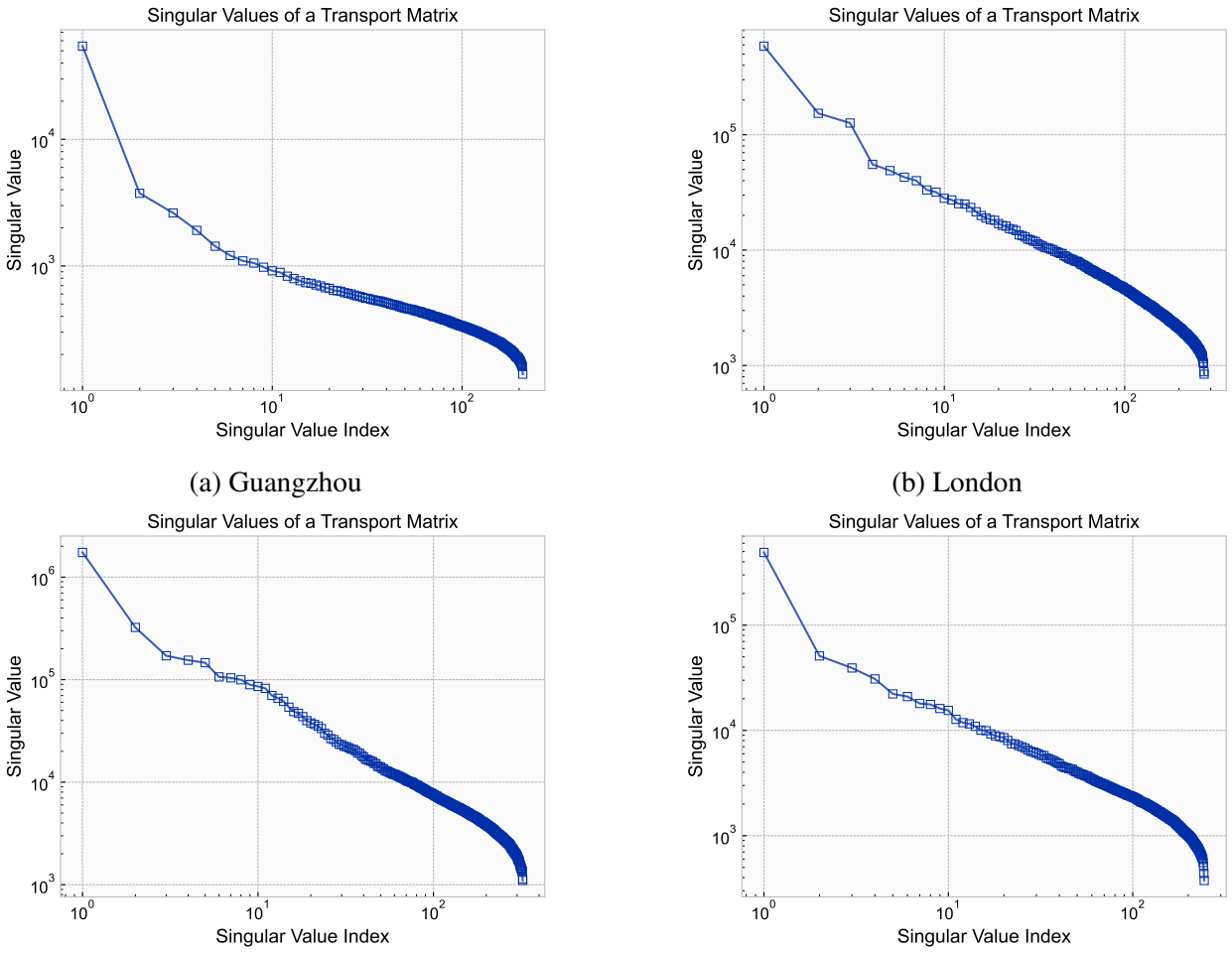
where BM represents the *blackout missing* scenario, indicating an entire day with no observations from any sensor. The term RM is indicative of a *random missing* scenario, wherein data points are randomly absent. The labels DM and TM denote *day missing* and *time-of-the-day missing*, respectively, implying random missing data along the dimensions of day and time. Each of these missing patterns is subjected to two basic missing rates, namely, 10% and 30%. Consequently, this yields a total of 16 unique missing scenarios. The resultant composite missing rates, a product of the combination of these patterns, vary between 34% and 77%. The detailed list of all missing scenarios can be found along with the imputation performance in the next section.

## Base Models and Hyperparameters

In our experiment, we compare the proposed imputation framework with several simple baselines and state-of-the-art tensor-based transport data imputation models.

- **Historical average (HA)**. HA is a naïve baseline that averages the observed values over each time-of-the-day.

[1]The Guangzhou data is available at `https://zenodo.org/record/1205229`. The other three datasets were gathered by the NeurIPS Traffic4cast 2022 Challenge at `https://www.iarai.ac.at/traffic4cast/challenge/`, where the actual source data can be found at `https://roads.data.tfl.gov.uk/`, `https://datos.madrid.es/egob/catalogo/202468-0-intensidad-trafico`, and `https://discover.data.vic.gov.au/dataset/traffic-signal-volume-data`




(a) Guangzhou (b) London

(c) Madrid (d) Melbourne

**FIGURE 3**: Singular values of transport data from four cities

- **Low-Rank Tensor Completion with Truncated Schattern-$p$ Norm (LRTC-TSpN)**. LRTC-TSpN is the state-of-the-art transport data imputation model based on rank minimization paradigm (*24*).
- **Bayesian Gaussian CANDECOMP/PARAFAC Decomposition (BGCP)**. BGCP is a Bayesian model based on tensor decomposition paradigm, which extends the Bayesian probabilistic matrix factorization model (*9*) from matrix to tensor (*13*).
- **Temporal Regularized Matrix Factorization (TRMF)**. TRMF is a matrix decomposition-based model with an integrated autoregressive model for temporal modeling (*42*).
- **Low-Rank Autoregressive Tensor Completion (LATC)**. LATC is a rank minimization-based model with an integrated autoregressive model for temporal modeling (*26*).

The hyperparameters of base models are set optimally through preliminary experiments. For LRTC-TSpN, we set norm parameter $p = 0.9$, truncation rate $\theta = 0.05$, decay rate $\beta = 2$, and updating step size $\rho = 10^{-5}$. For BGCP, we set tensor rank $r = 30$. For TRMF, we set tensor rank $r = 10$, coefficients $\lambda_x = \lambda_w = \lambda_\theta = 1$, and updating step size $\eta = 10^{-3}$. For LATC, we set truncation parameter $r = 10$, coefficient $\lambda = 10^{-5}$, and updating step size $\rho = 10^{-5}$. Finally, for

our model, we use tensor ranks $\boldsymbol{r} = \{30, 9, 5, 20\}$, coefficients $\mu = 0.1, \lambda = 0.1, \xi = 1.0$, norm parameter $\gamma = 0.01$, and updating step size $\rho = \eta = 10^{-4}$. While our method seems to have many hyperparameters, it will be shown later that the influence of hyperparameters on the imputation performance is minor.

Two performance metrics, including an absolute metric (mean absolute error, MAE) and a relative metric (symmetric mean absolute percentage error, SMAPE), are used to quantify the imputation quality of the models, as defined below:

$$\text{MAE} = \frac{1}{|I(\tilde{\Omega})|} \sum_{i \in I(\tilde{\Omega})} |x_i - y_i|, \quad \text{SMAPE} = \frac{1}{|I(\tilde{\Omega})|} \sum_{i \in I(\tilde{\Omega})} \frac{|x_i - y_i|}{|x_i| + |y_i|} \times 100\%, \tag{37}$$

where $I(\tilde{\Omega})$ is the index set of testing entries given the mask $\tilde{\Omega}$.

## EXPERIMENT RESULTS

In this section, the proposed imputation framework is first compared with base models on all evaluation scenarios. Additionally, the effect of hyperparameters on the imputation performance is examined through sensitivity analysis.

### Imputation Performance

The imputation performance of all the models on 16 missing scenarios in the four cities are listed in Tables 2–5. On a general note, the proposed imputation framework achieves the lowest error in most of the scenarios, particularly, with higher advantage over base models in London and Madrid as well as scenarios with higher missing rates. It is worth noting that some state-of-the-art imputation models, despite exemplary performance on individual missing scenarios substantiated by prior literature and empirical observations, can suffer from compromised reliability in the presence of intricate composite missing scenarios, resulting in imputation error higher than that of historical average.

The performance of the proposed imputation framework is consistently better than base models on London and Madrid datasets. Larger missing rates usually result in lower imputation accuracy due to less observed information, conforming with the error values in the tables. Nevertheless, a smaller degradation can be observed for the proposed model, even faced with high missing rates, indicating its robustness over various missing patterns. For instance, the error gap between the best and the worse scenario in London for the proposed model is approximately 7.5 veh/15min. However, for LATC, the best base model in our experiment, the gap is around 11.4 veh/15min. The gap can be even larger for other models except historical average, which is robust but with higher imputation error.

Compared with London and Madrid, the imputation error on Guangzhou and Melbourne is generally smaller; and the proposed model does not show discernible merit in scenarios with low missing rates, which could be attributed to less challenging traffic dynamics there. Clues can be found by recalling the distribution of singular values of all cities demonstrated in Figure 3, where the primary singular values of Guangzhou and Melbourne show stronger dominance compared with the other cities, implying higher volatility and more pattern changes in temporal dynamics (e.g., see Figure 1. Still, in these two cities, lower imputation error of the proposed model over base models can be noticed in scenarios with a high BM missing rate.

**TABLE 2**: Imputation Performance of All Models on Guangzhou Dataset

| BM | RM | DM | TM | Overall | HA | LRTC-TSpN | TRMF | BGCP | LATC | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | 10% | 10% | 10% | 34% | 4.18/6.26 | 3.68/5.50 | 3.17/4.92 | 2.82/4.44 | 2.61/4.10 | 2.64/4.18 |
| 10% | 10% | 10% | 30% | 48% | 3.94/5.95 | 3.44/5.21 | 3.16/4.91 | 2.86/4.49 | 2.68/4.21 | 2.73/4.31 |
| 10% | 10% | 30% | 10% | 49% | 5.22/7.52 | 3.34/5.07 | 3.13/4.86 | 2.77/4.35 | 2.58/4.06 | 2.62/4.14 |
| 10% | 10% | 30% | 30% | 60% | 4.87/7.09 | 3.28/5.01 | 3.17/4.94 | 2.80/4.40 | 2.66/4.19 | 2.67/4.23 |
| 10% | 30% | 10% | 10% | 48% | 3.94/5.95 | 3.31/5.02 | 3.12/4.86 | 2.73/4.29 | 2.55/4.03 | 2.59/4.10 |
| 10% | 30% | 10% | 30% | 60% | 3.84/5.82 | 3.26/4.98 | 3.17/4.93 | 2.79/4.40 | 2.64/4.16 | 2.69/4.24 |
| 10% | 30% | 30% | 10% | 60% | 4.87/7.09 | 3.22/4.93 | 3.14/4.90 | 2.74/4.30 | 2.60/4.10 | 2.61/4.13 |
| 10% | 30% | 30% | 30% | 69% | 4.69/6.86 | 3.21/4.92 | 3.22/5.02 | 2.77/4.35 | 2.68/4.22 | 2.66/4.20 |
| 30% | 10% | 10% | 10% | 52% | 3.83/5.77 | 4.51/6.56 | 4.68/6.78 | 3.30/5.04 | 3.15/4.82 | 2.91/4.56 |
| 30% | 10% | 10% | 30% | 62% | 3.78/5.72 | 4.27/6.27 | 4.45/6.52 | 3.29/5.05 | 3.08/4.74 | 2.91/4.56 |
| 30% | 10% | 30% | 10% | 62% | 4.76/6.93 | 4.19/6.16 | 4.38/6.43 | 3.23/4.94 | 3.03/4.66 | 2.84/4.45 |
| 30% | 10% | 30% | 30% | 71% | 4.62/6.77 | 4.02/5.95 | 4.20/6.22 | 3.25/4.98 | 3.01/4.66 | 2.89/4.53 |
| 30% | 30% | 10% | 10% | 62% | 3.78/5.73 | 4.18/6.14 | 4.39/6.46 | 3.19/4.90 | 3.01/4.63 | 2.87/4.48 |
| 30% | 30% | 10% | 30% | 71% | 3.77/5.72 | 4.01/5.93 | 4.21/6.26 | 3.20/4.93 | 3.00/4.64 | 2.88/4.52 |
| 30% | 30% | 30% | 10% | 71% | 4.62/6.77 | 3.97/5.89 | 4.17/6.19 | 3.12/4.81 | 2.97/4.60 | 2.82/4.42 |
| 30% | 30% | 30% | 30% | 77% | 4.54/6.68 | 3.87/5.76 | 4.02/6.03 | 3.16/4.87 | 2.98/4.63 | 2.87/4.50 |

**TABLE 3**: Imputation Performance of All Models on London Dataset

| BM | RM | DM | TM | Overall | HA | LRTC-TSpN | TRMF | BGCP | LATC | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | 10% | 10% | 10% | 34% | 65.41/12.89 | 61.13/11.65 | 52.55/10.57 | 56.18/9.42 | 36.77/7.79 | 36.32/7.81 |
| 10% | 10% | 10% | 30% | 48% | 64.33/12.59 | 57.44/10.68 | 56.02/10.89 | 78.23/9.71 | 38.52/8.01 | 38.35/8.12 |
| 10% | 10% | 30% | 10% | 49% | 74.98/14.61 | 56.42/10.35 | 50.06/10.40 | 49.96/9.04 | 42.57/8.02 | 36.61/7.71 |
| 10% | 10% | 30% | 30% | 60% | 72.58/14.11 | 55.91/10.12 | 54.58/10.62 | 112.93/9.56 | 43.34/8.26 | 38.26/8.05 |
| 10% | 30% | 10% | 10% | 48% | 64.25/12.51 | 52.68/10.28 | 49.39/10.14 | 49.82/8.92 | 36.27/7.66 | 35.83/7.65 |
| 10% | 30% | 10% | 30% | 60% | 63.89/12.40 | 53.21/10.07 | 51.94/10.49 | 62.91/9.28 | 38.25/7.95 | 38.11/8.04 |
| 10% | 30% | 30% | 10% | 60% | 72.52/14.06 | 52.77/9.88 | 47.65/9.99 | 46.92/8.77 | 42.32/8.08 | 36.79/7.73 |
| 10% | 30% | 30% | 30% | 69% | 71.26/13.80 | 54.14/9.94 | 50.61/10.35 | 78.02/9.42 | 43.55/8.36 | 38.49/8.04 |
| 30% | 10% | 10% | 10% | 52% | 64.74/12.71 | 94.84/17.49 | 63.20/13.01 | 68.60/11.02 | 43.60/9.10 | 40.01/8.37 |
| 30% | 10% | 10% | 30% | 62% | 64.22/12.50 | 86.16/15.80 | 68.67/13.11 | 79.93/10.94 | 43.73/9.02 | 41.32/8.61 |
| 30% | 10% | 30% | 10% | 62% | 72.73/14.15 | 84.95/15.52 | 57.55/11.98 | 54.68/10.32 | 47.35/9.10 | 40.36/8.33 |
| 30% | 10% | 30% | 30% | 71% | 71.44/13.84 | 81.24/14.74 | 62.56/12.41 | 80.49/10.47 | 47.70/9.18 | 42.17/8.65 |
| 30% | 30% | 10% | 10% | 62% | 64.11/12.47 | 83.42/15.57 | 58.97/12.01 | 53.20/10.21 | 41.95/8.74 | 40.38/8.54 |
| 30% | 30% | 10% | 30% | 71% | 63.96/12.38 | 79.93/14.77 | 63.27/12.22 | 93.50/10.71 | 42.92/8.84 | 42.94/8.83 |
| 30% | 30% | 30% | 10% | 71% | 71.36/13.82 | 79.24/14.59 | 55.09/11.40 | 52.45/10.02 | 46.78/9.02 | 40.27/8.32 |
| 30% | 30% | 30% | 30% | 77% | 70.63/13.65 | 77.56/14.17 | 58.99/11.76 | 69.94/10.18 | 47.70/9.19 | 43.34/8.82 |

**TABLE 4**: Imputation Performance of All Models on Madrid Dataset

| BM | RM | DM | TM | Overall | HA | LRTC-TSpN | TRMF | BGCP | LATC | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | 10% | 10% | 10% | 34% | 194.42/21.14 | 207.74/20.04 | 88.13/13.38 | 85.06/11.80 | 67.35/9.59 | 53.36/9.08 |
| 10% | 10% | 10% | 30% | 48% | 182.28/19.96 | 204.24/19.34 | 80.88/12.57 | 77.00/11.20 | 65.53/9.55 | 54.19/9.28 |
| 10% | 10% | 30% | 10% | 49% | 242.42/25.66 | 212.25/19.76 | 79.45/12.44 | 75.73/10.98 | 85.42/9.82 | 65.29/9.51 |
| 10% | 10% | 30% | 30% | 60% | 225.80/24.09 | 217.01/21.30 | 76.78/12.09 | 71.89/10.79 | 81.87/9.84 | 65.11/9.60 |
| 10% | 30% | 10% | 10% | 48% | 181.95/19.94 | 185.18/17.60 | 80.10/12.53 | 73.95/10.91 | 64.31/9.42 | 53.16/9.15 |
| 10% | 30% | 10% | 30% | 60% | 176.67/19.43 | 204.42/20.68 | 76.81/12.18 | 71.86/10.72 | 63.94/9.47 | 54.99/9.36 |
| 10% | 30% | 30% | 10% | 60% | 225.67/24.08 | 209.52/20.80 | 77.45/12.32 | 69.85/10.66 | 80.97/9.78 | 64.25/9.50 |
| 10% | 30% | 30% | 30% | 69% | 216.53/23.21 | 229.33/24.48 | 75.19/12.03 | 70.04/10.66 | 79.84/9.87 | 64.57/9.69 |
| 30% | 10% | 10% | 10% | 52% | 182.76/19.97 | 328.55/31.66 | 160.92/20.72 | 118.78/14.38 | 70.72/10.07 | 63.01/9.78 |
| 30% | 10% | 10% | 30% | 62% | 177.05/19.45 | 303.96/28.97 | 143.67/19.11 | 108.99/13.62 | 69.28/9.96 | 59.46/9.70 |
| 30% | 10% | 30% | 10% | 62% | 225.74/24.06 | 304.56/28.62 | 139.31/18.18 | 106.29/13.45 | 86.01/10.29 | 67.30/9.69 |
| 30% | 10% | 30% | 30% | 71% | 216.60/23.22 | 294.69/28.06 | 129.68/17.34 | 100.68/13.03 | 85.05/10.35 | 71.58/10.09 |
| 30% | 30% | 10% | 10% | 62% | 177.04/19.44 | 290.48/27.88 | 146.32/20.24 | 106.76/13.49 | 67.67/9.84 | 58.59/9.50 |
| 30% | 30% | 10% | 30% | 71% | 174.01/19.17 | 287.60/27.92 | 134.10/18.46 | 101.29/13.10 | 68.60/9.90 | 58.33/9.63 |
| 30% | 30% | 30% | 10% | 71% | 216.57/23.21 | 287.96/27.42 | 132.31/18.51 | 99.94/13.06 | 83.39/10.24 | 69.88/9.92 |
| 30% | 30% | 30% | 30% | 77% | 210.97/22.69 | 293.14/29.11 | 124.91/17.37 | 97.91/12.80 | 85.34/10.45 | 72.94/10.29 |

**TABLE 5**: Imputation Performance of All Models on Melbourne Dataset

| BM | RM | DM | TM | Overall | HA | LRTC-TSpN | TRMF | BGCP | LATC | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | 10% | 10% | 10% | 34% | 73.74/25.20 | 42.88/17.63 | 28.00/16.00 | 30.28/14.70 | 18.54/9.87 | 18.74/9.52 |
| 10% | 10% | 10% | 30% | 48% | 68.21/23.54 | 34.86/15.25 | 30.29/16.01 | 39.60/14.32 | 18.91/10.33 | 19.65/9.99 |
| 10% | 10% | 30% | 10% | 49% | 95.81/31.62 | 34.52/15.05 | 25.98/15.32 | 27.38/14.11 | 19.50/10.20 | 19.05/9.59 |
| 10% | 10% | 30% | 30% | 60% | 88.25/29.40 | 32.10/14.45 | 26.37/15.24 | 31.09/14.08 | 20.03/10.70 | 19.48/9.95 |
| 10% | 30% | 10% | 10% | 48% | 68.32/23.53 | 34.18/14.97 | 27.94/15.85 | 26.41/13.83 | 18.06/9.94 | 19.03/9.58 |
| 10% | 30% | 10% | 30% | 60% | 65.85/22.80 | 31.63/14.37 | 27.82/15.64 | 29.96/13.72 | 18.81/10.47 | 19.07/9.79 |
| 10% | 30% | 30% | 10% | 60% | 88.30/29.39 | 31.44/14.25 | 25.45/15.15 | 25.33/13.55 | 19.48/10.45 | 19.02/9.50 |
| 10% | 30% | 30% | 30% | 69% | 84.13/28.17 | 30.48/14.14 | 25.81/15.23 | 29.04/13.69 | 20.16/10.93 | 19.45/9.88 |
| 30% | 10% | 10% | 10% | 52% | 67.95/23.32 | 82.76/29.13 | 59.37/24.37 | 42.41/17.67 | 23.67/11.74 | 21.99/10.28 |
| 30% | 10% | 10% | 30% | 62% | 65.58/22.64 | 71.41/25.92 | 56.26/22.86 | 45.19/17.36 | 22.87/11.75 | 22.16/10.33 |
| 30% | 10% | 30% | 10% | 62% | 87.75/29.20 | 70.90/25.67 | 47.88/21.62 | 38.43/16.77 | 23.45/11.71 | 24.39/10.99 |
| 30% | 10% | 30% | 30% | 71% | 83.80/28.05 | 64.60/24.03 | 48.43/21.29 | 38.79/16.47 | 23.32/11.92 | 22.20/10.50 |
| 30% | 30% | 10% | 10% | 62% | 65.65/22.64 | 70.87/25.65 | 47.04/21.30 | 37.95/16.64 | 22.28/11.43 | 21.59/10.24 |
| 30% | 30% | 10% | 30% | 71% | 64.38/22.28 | 64.50/24.02 | 47.11/20.88 | 38.99/16.42 | 22.22/11.67 | 20.56/9.96 |
| 30% | 30% | 30% | 10% | 71% | 83.78/28.03 | 64.17/23.84 | 42.75/19.91 | 35.41/16.04 | 22.87/11.72 | 21.03/9.99 |
| 30% | 30% | 30% | 30% | 77% | 81.37/27.32 | 60.80/23.06 | 41.73/19.59 | 37.57/16.13 | 23.11/12.02 | 22.47/10.47 |

It can also be concluded from the results that explicit or implicit temporal modeling is crucial to robust imputation under complicated missing scenarios. For models without temporal modeling, such as LRTC-TSpN, the performance is not satisfactory in most evaluation scenarios due to the permutation invariance of tensor/matrix rank, which is an inherent limitation of the rank minimization paradigm. In other words, by minimizing the rank of a tensor, the solution remains unchanged regardless of the order of any two fibers (analogous to columns/rows of matrix in higher order). Consequently, when many fibers or complete slices are missing, imputation models are unable to correctly recover missing values unless equipped with a time series prior.

Among all missing patterns, RM, DM and TM are missing patterns that are easier to handle. An increase in the missing rate regarding these missing patterns, as per the tables, incurs limited perturbations in terms of the errors in most cases. In comparison, most models are more sensitive to changes in the missing rates of BM. Although models like TRMF, BGCP, and LATC are all embedded with time series models, they still suffer from discernible performance degradation or instability during imputation.

**Sensitivity Analysis**

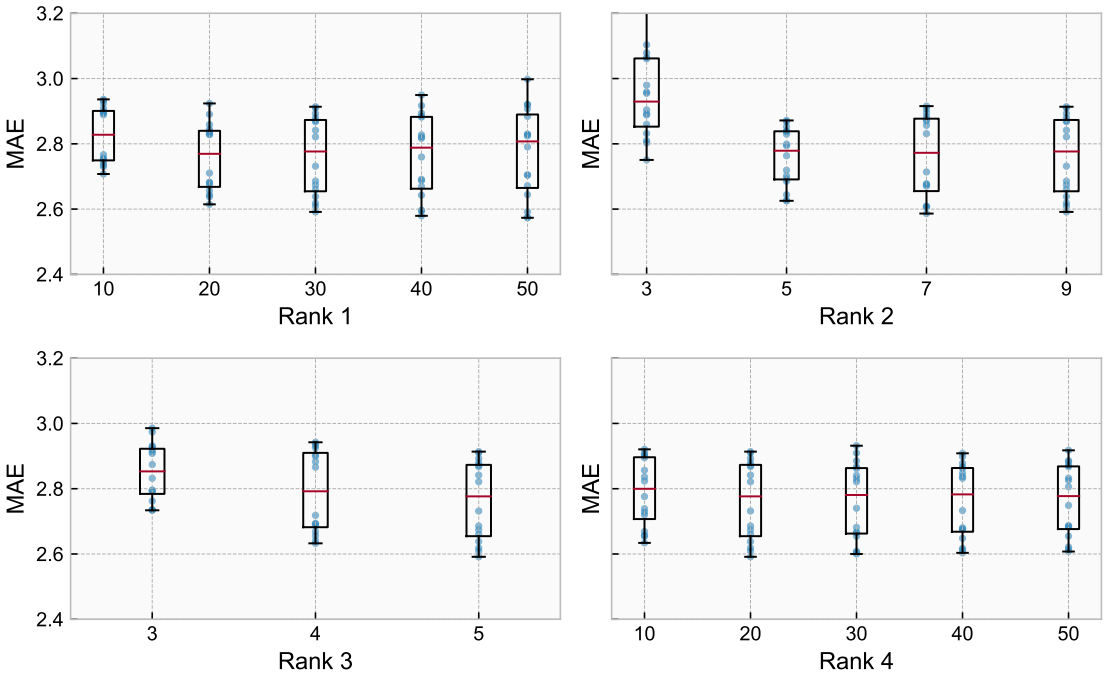
While the proposed imputation framework outperforms base models in terms of accuracy and robustness, one may question its practical feasibility due to the requirement of tuning more hyperparameters. A sensitivity analysis was first conducted specifically focusing on tensor ranks for Tucker decomposition, as demonstrated in Figure 4.

In general, most rank configurations do not exert a significant impact on the imputation error, showing an observable degree of consistency. Minor deviations do exist, however, the overall performance remains comparably stable against base models. There are also very few exceptions, where increases in imputation error can be noted, e.g., small rank values in Guangzhou data. This is possibly because of under-fitting with small rank values, which may limit the model from capturing all necessary details for accurately recovering the traffic dynamics. Therefore, in practice, a trade-off between quality and efficiency is needed for practical application. Nonetheless, thanks to the additional rank minimization, the risk of over-fitting for selecting a large rank value is minimized, which greatly relieves the burden of rank tuning compared with other tensor decomposition-based models.

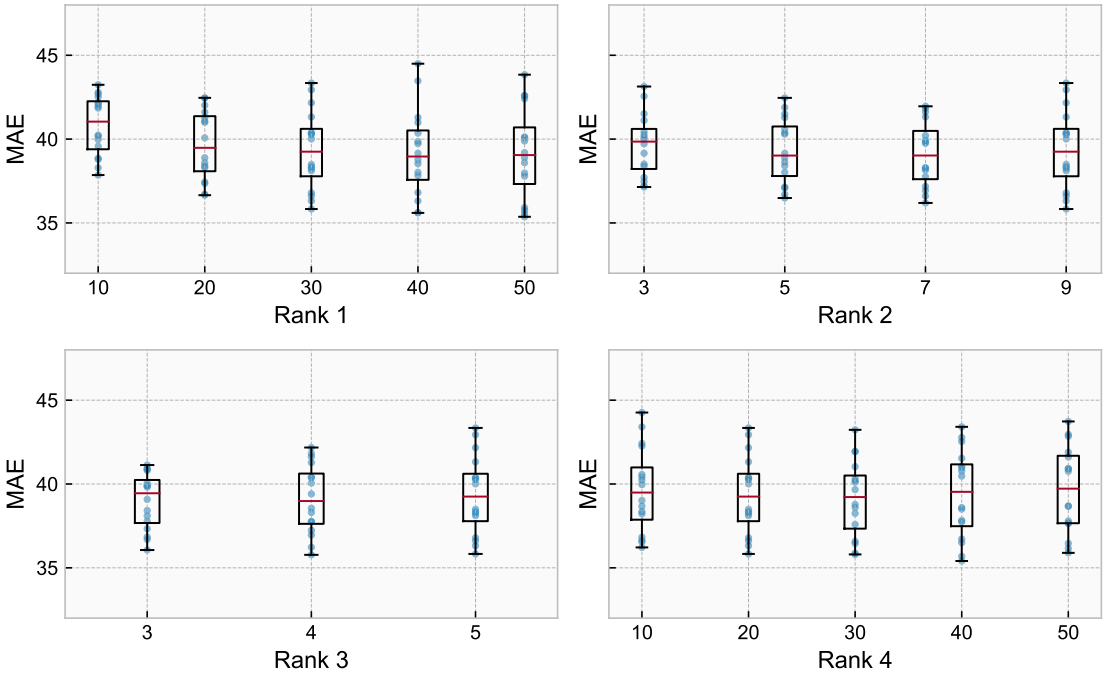
Further sensitivity analysis was performed on other model parameters, encompassing coefficients $\mu, \lambda, \xi$ and the norm parameter $\gamma$. In general, these hyperparameters do not exert a substantial effect on the imputation error, with the exception of the coefficient $\mu$. However, the impact of $\mu$ on imputation error is as minor as other hyperparameters when it is less than 1. This analysis underscores the robustness of the proposed imputation framework in relation to hyperparameter sensitivity, suggesting that avoiding overly small rank values and overly large $\mu$ will suffice to achieve optimal performance, while the influences of other parameters remain mostly invariant.

## CONCLUSION

In this paper, we addressed the pervasive problem of missing values in transport data, which undermines the integrity and efficacy of data-driven transportation analysis. As one of the most promising solutions to this problem, current tensor-based methods are still limited in terms of robustness facing complicated composite missing patterns. To amend this gap, we proposed a novel tensor-based imputation framework that integrates a time series decomposition model to simultaneously

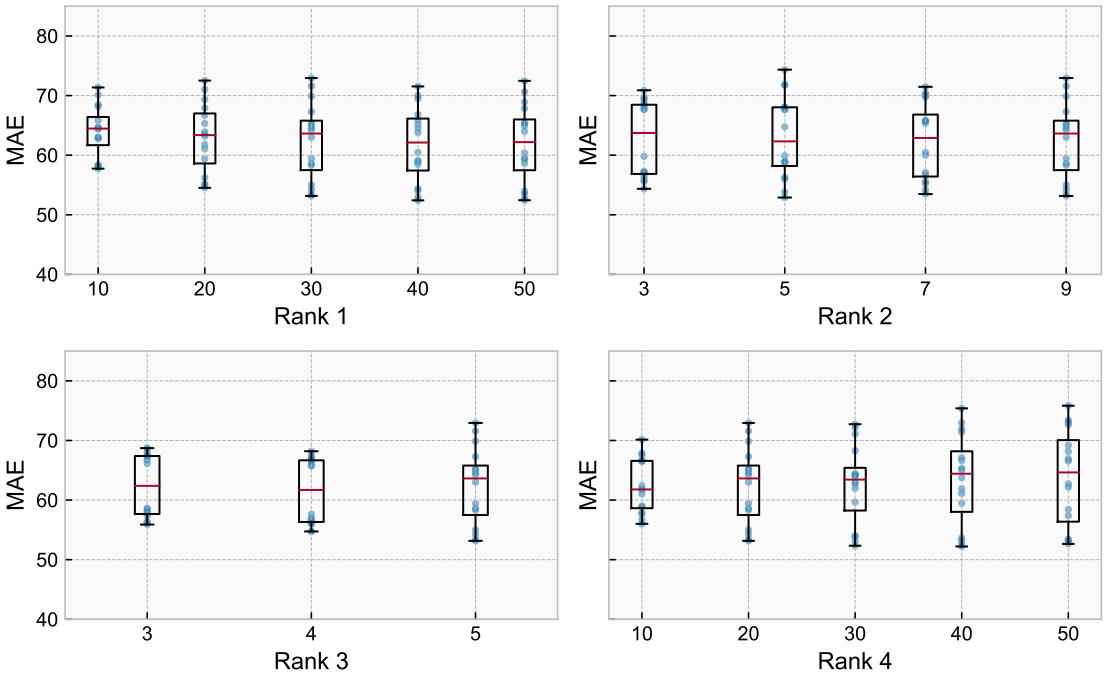



(a) Guangzhou

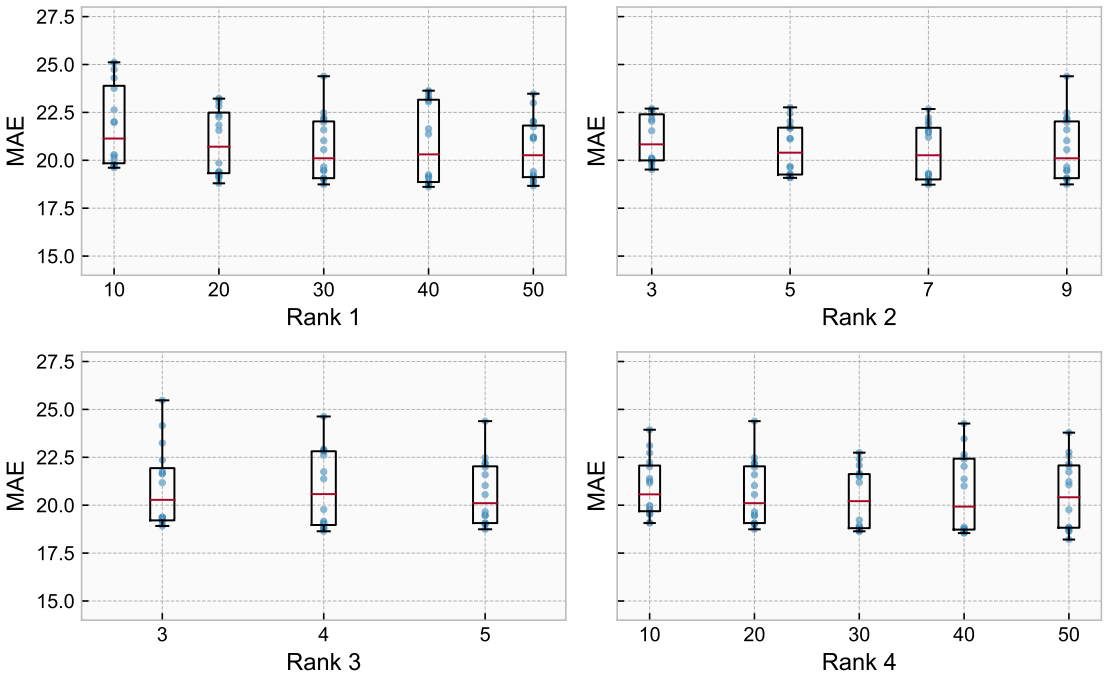



(b) London

**FIGURE 4**: Sensitivity analysis on tensor ranks

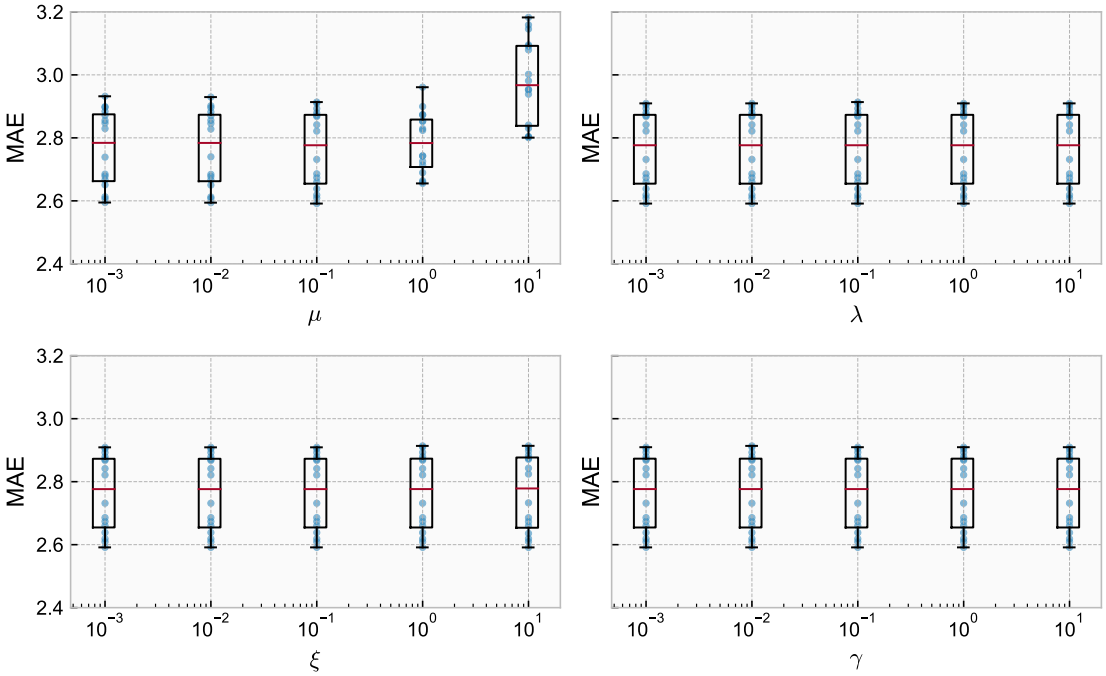



(c) Madrid

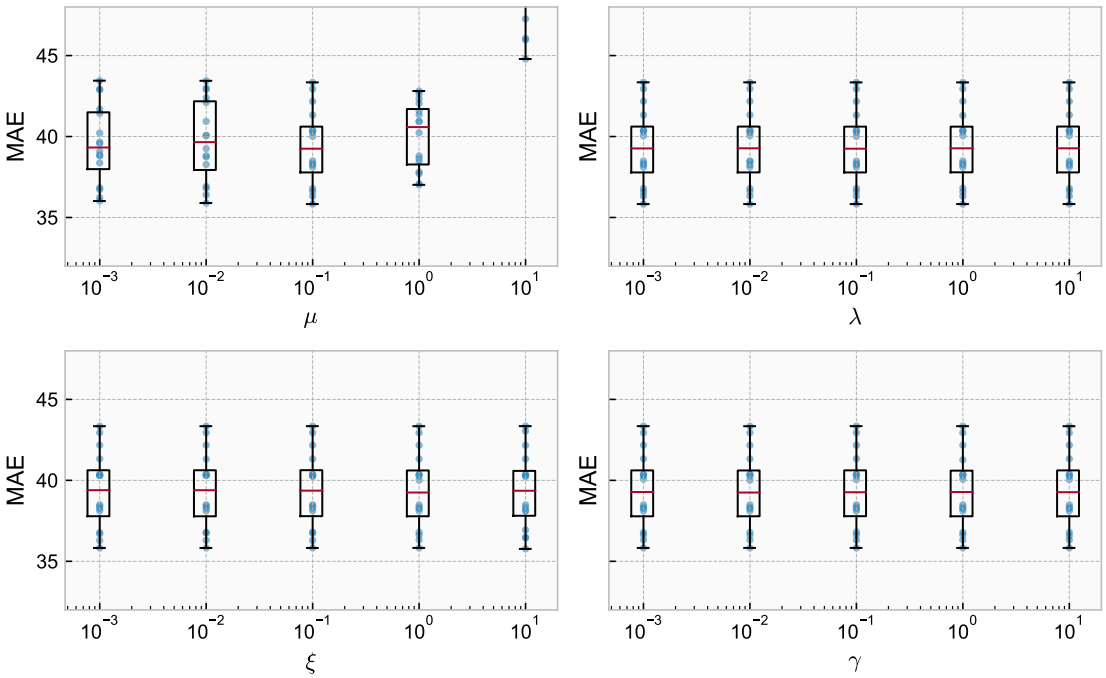



(d) Melbourne

**FIGURE 4**: Sensitivity analysis on tensor ranks (cont'd)

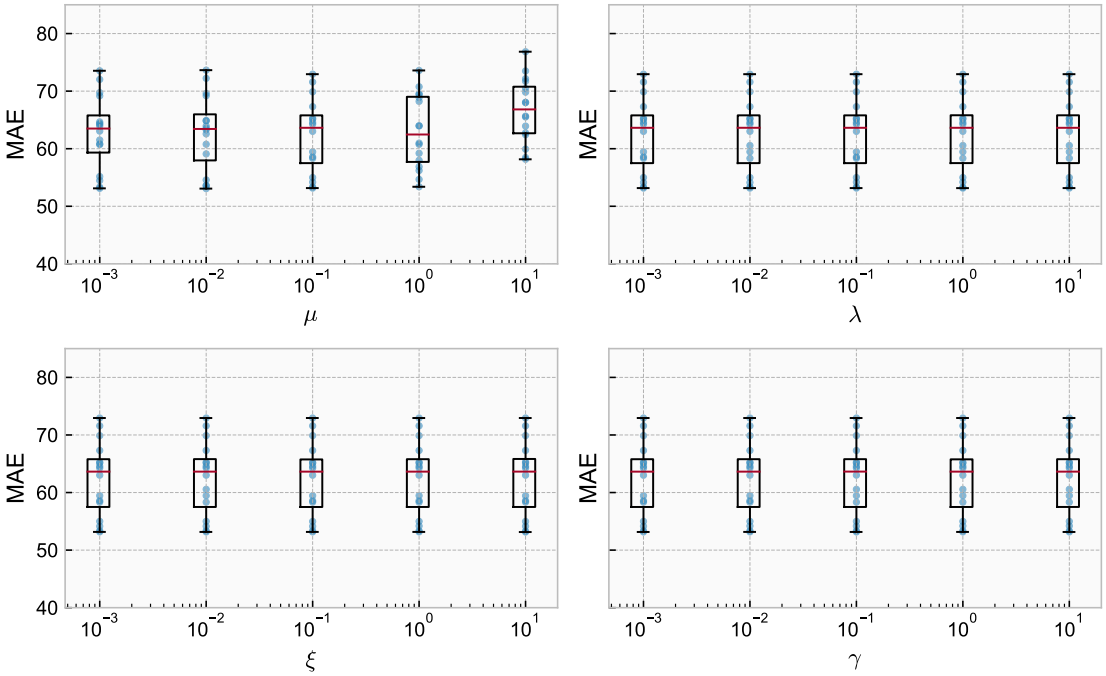



(a) Guangzhou

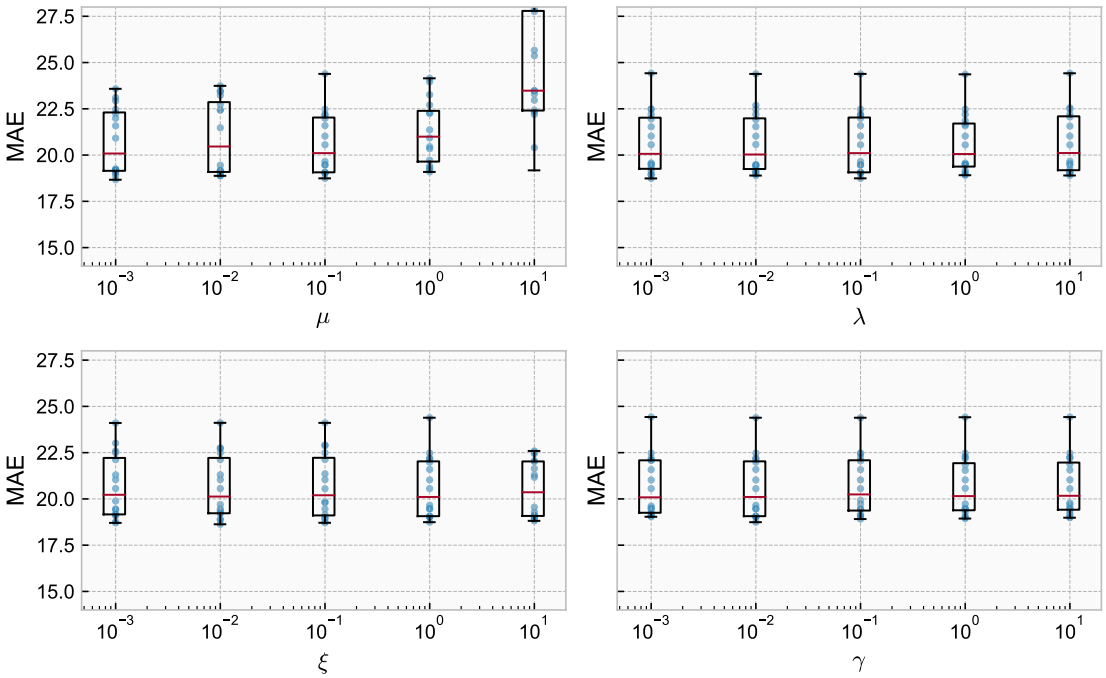



(b) London

**FIGURE 5**: Sensitivity analysis on other model parameters




(c) Madrid




(d) Melbourne

**FIGURE 5**: Sensitivity analysis on other model parameters (cont'd)

account for long-term trends, spatio-temporal correlations, and outliers in the data. The combination of tensor decomposition and rank minimization also eliminates the need for exhaustive rank tuning in conventional tensor decomposition-based methods. In addition, an improved ADMM algorithm is developed to solve the resulting multi-block separable nonconvex optimization problem efficiently.

Experiments on four real-world transport datasets demonstrate that the proposed framework can outperform state-of-the-art imputation methods, especially in the presence of complex missing patterns with high missing rates. The results highlight the importance of integrating temporal modeling in tensor completion framework. Sensitivity analysis also underscores the stability of our framework with respect to hyperparameter settings.

Future work in this direction may focus on incorporating supplementary information to further assist imputation, such as road topology, weather and events. From a methodological perspective, it is also worthwhile to investigate better solutions for temporal modeling as well as more efficient solution algorithms. Overall, this research contributes towards advancing data-driven transport applications through improving the integrity and reliability of input data.

## ACKNOWLEDGEMENTS

This work was supported by the PANAMERA project (Grant Number: 19I21016F) funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). This research was also partially supported by the European Interest Group CONCERT-Japan DARUMA project (Grant 01DR21010) funded by the German Federal Ministry of Education and Research (BMBF).

## REFERENCES


1. Mahajan, V., N. Kuehnel, A. Intzevidou, G. Cantelmo, R. Moeckel, and C. Antoniou, Data to the People: A Review of Public and Proprietary Data for Transport Models. *Transport Reviews*, Vol. 42, No. 4, 2022, pp. 415–440.
2. Liu, Y., C. Lyu, Y. Zhang, Z. Liu, W. Yu, and X. Qu, DeepTSP: Deep Traffic State Prediction Model Based on Large-Scale Empirical Data. *Communications in Transportation Research*, Vol. 1, 2021, p. 100012.
3. Laña, I., I. I. Olabarrieta, M. Vélez, and J. Del Ser, On the Imputation of Missing Data for Road Traffic Forecasting: New Insights and Novel Techniques. *Transportation Research Part C: Emerging Technologies*, Vol. 90, 2018, pp. 18–33.
4. El Esawey, M., A. I. Mosa, and K. Nasr, Estimation of Daily Bicycle Traffic Volumes Using Sparse Data. *Computers, Environment and Urban Systems*, Vol. 54, 2015, pp. 195–203.
5. Fang, S., C. Zhang, S. Xiang, and C. Pan, AutoMSNet: Multi-Source Spatio-Temporal Network via Automatic Neural Architecture Search for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24, No. 3, 2023, pp. 2827–2841.
6. AASHTO, *AASHTO Guidelines for Traffic Data Programs*. American Association of State Highway and Transportation Officials, Washington, D.C, 2009.
7. Smith, B. L., W. T. Scherer, and J. H. Conklin, Exploring Imputation Techniques for Missing Data in Transportation Management Systems. *Transportation Research Record*, Vol. 1836, No. 1, 2003, pp. 132–142.
8. Liu, J., P. Musialski, P. Wonka, and J. Ye, Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 1, 2013, pp. 208–220.
9. Salakhutdinov, R. and A. Mnih, Bayesian Probabilistic Matrix Factorization Using Markov Chain Monte Carlo. In *Proceedings of the 25th International Conference on Machine Learning*, ACM Press, Helsinki, Finland, 2008, pp. 880–887.
10. Sun, L. and K. W. Axhausen, Understanding Urban Mobility Patterns with a Probabilistic Tensor Factorization Framework. *Transportation Research Part B: Methodological*, Vol. 91, 2016, pp. 511–524.
11. Tan, H., G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, A Tensor-Based Method for Missing Traffic Data Completion. *Transportation Research Part C: Emerging Technologies*, Vol. 28, 2013, pp. 15–27.
12. Li, L., Y. Li, and Z. Li, Efficient Missing Data Imputing for Traffic Flow by Considering Temporal and Spatial Dependence. *Transportation Research Part C: Emerging Technologies*, Vol. 34, 2013, pp. 108–120.
13. Chen, X., Z. He, and L. Sun, A Bayesian Tensor Decomposition Approach for Spatiotemporal Traffic Data Imputation. *Transportation Research Part C: Emerging Technologies*, Vol. 98, 2019, pp. 73–84.
14. Yamamoto, R., H. Hontani, A. Imakura, and T. Yokota, Fast Algorithm for Low-rank Tensor Completion in Delay-embedded Space. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, New Orleans, LA, USA, 2022, pp. 2048–2056.
15. Tight, M., E. Redfern, S. Watson, and S. Clark, *Outlier Detection and Missing Value Estimation in Time Series Traffic Count Data*. The University of Leeds, Leeds, UK, 1993.

16. Tang, J., G. Zhang, Y. Wang, H. Wang, and F. Liu, A Hybrid Approach to Integrate Fuzzy C-means Based Imputation Method with Genetic Algorithm for Missing Traffic Volume Data Estimation. *Transportation Research Part C: Emerging Technologies*, Vol. 51, 2015, pp. 29–40.
17. Li, H., M. Li, X. Lin, F. He, and Y. Wang, A Spatiotemporal Approach for Traffic Data Imputation with Complicated Missing Patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 119, 2020, p. 102730.
18. Jiang, W., N. Zheng, and I. Kim, Missing Data Imputation for Transfer Passenger Flow Identified from In-Station WiFi Systems. *Transportmetrica B*, 2022, pp. 1–18.
19. Duan, Y., Y. Lv, Y.-L. Liu, and F.-Y. Wang, An Efficient Realization of Deep Learning for Traffic Data Imputation. *Transportation Research Part C: Emerging Technologies*, Vol. 72, 2016, pp. 168–181.
20. Chen, Y. and X. M. Chen, A Novel Reinforced Dynamic Graph Convolutional Network Model with Data Imputation for Network-Wide Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 143, 2022, p. 103820.
21. Liang, Y., Z. Zhao, and L. Sun, Memory-Augmented Dynamic Graph Convolution Networks for Traffic Data Imputation with Diverse Missing Patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 143, 2022, p. 103826.
22. Chen, X., Y. Cai, Q. Ye, L. Chen, and Z. Li, Graph Regularized Local Self-Representation for Missing Value Imputation with Applications to on-Road Traffic Sensor Data. *Neurocomputing*, Vol. 303, 2018, pp. 47–59.
23. 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.
24. Nie, T., G. Qin, and J. Sun, Truncated Tensor Schatten P-Norm Based Approach for Spatiotemporal Traffic Data Imputation with Complicated Missing Patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 141, 2022, p. 103737.
25. Chen, X., Y. Chen, N. Saunier, and L. Sun, Scalable Low-Rank Tensor Learning for Spatiotemporal Traffic Data Imputation. *Transportation Research Part C: Emerging Technologies*, Vol. 129, 2021, p. 103226.
26. Chen, X., M. Lei, N. Saunier, and L. Sun, Low-Rank Autoregressive Tensor Completion for Spatiotemporal Traffic Data Imputation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 8, 2022, pp. 12301–12310.
27. Kolda, T. G. and B. W. Bader, Tensor Decompositions and Applications. *SIAM Review*, Vol. 51, No. 3, 2009, pp. 455–500.
28. Killick, R., P. Fearnhead, and I. A. Eckley, Optimal Detection of Changepoints with a Linear Computational Cost. *Journal of the American Statistical Association*, Vol. 107, No. 500, 2012, pp. 1590–1598.
29. Arlot, S., A. Celisse, and Z. Harchaoui, A Kernel Multiple Change-Point Algorithm via Model Selection. *Journal of Machine Learning Research*, Vol. 20, No. 162, 2019, pp. 1–56.
30. Goulart, J. d. M., A. Kibangou, and G. Favier, Traffic Data Imputation via Tensor Completion Based on Soft Thresholding of Tucker Core. *Transportation Research Part C: Emerging Technologies*, Vol. 85, 2017, pp. 348–362.

31. Wang, Y., J. Yang, W. Yin, and Y. Zhang, A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, Vol. 1, No. 3, 2008, pp. 248–272.
32. Yu, Q., X. Zhang, Y. Chen, and L. Qi, Low Tucker Rank Tensor Completion Using a Symmetric Block Coordinate Descent Method. *Numerical Linear Algebra with Applications*, 2022, p. e2464.
33. Kang, Z., C. Peng, and Q. Cheng, Robust PCA via Nonconvex Rank Approximation. In *Proceedings of the 15th IEEE International Conference on Data Mining*, IEEE, Atlantic City, USA, 2015, pp. 211–220.
34. Boyd, S., Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, Vol. 3, No. 1, 2010, pp. 1–122.
35. Chen, C., B. He, Y. Ye, and X. Yuan, The Direct Extension of ADMM for Multi-Block Convex Minimization Problems Is Not Necessarily Convergent. *Mathematical Programming*, Vol. 155, No. 1-2, 2016, pp. 57–79.
36. Bolte, J., S. Sabach, and M. Teboulle, Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, Vol. 146, No. 1-2, 2014, pp. 459–494.
37. Wang, F., W. Cao, and Z. Xu, Convergence of Multi-Block Bregman ADMM for Nonconvex Composite Problems. *Science China Information Sciences*, Vol. 61, No. 12, 2018, p. 122101.
38. Bartels, R. H. and G. W. Stewart, Solution of the Matrix Equation $AX + XB = C$. *Communications of the ACM*, Vol. 15, No. 9, 1972, pp. 820–826.
39. Chen, K., H. Dong, and K.-S. Chan, Reduced Rank Regression via Adaptive Nuclear Norm Penalization. *Biometrika*, Vol. 100, No. 4, 2013, pp. 901–920.
40. Yin, W., S. Osher, D. Goldfarb, and J. Darbon, Bregman Iterative Algorithms for $\ell_1$-Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, Vol. 1, No. 1, 2008, pp. 143–168.
41. Hale, E. T., W. Yin, and Y. Zhang, Fixed-Point Continuation for $\ell$1-Minimization: Methodology and Convergence. *SIAM Journal on Optimization*, Vol. 19, No. 3, 2008, pp. 1107–1130.
42. Yu, H.-F., N. Rao, and I. S. Dhillon, Temporal Regularized Matrix Factorization for High-dimensional Time Series Prediction. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ACM Press, Barcelona, Spain, 2016, pp. 847–855.