Traffic & Transportation

A Bayesian Robust Tensor Ring Decomposition Model with Gaussian-Wishart Prior for Missing Traffic Data Completion

Longsheng HUANG[1], Yu ZHU[2], Hanzeng SHAO[3], Yun ZHU[4], Gaohang YU[5], Jun WANG[6]

Original Scientific Paper
Submitted: 27 Mar 2024
Accepted: 15 Oct 2024

[1] huanglongsheng@gnnu.edu.cn, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China
[2] zzhuy1028@163.com, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China
[3] 914554245@qq.com, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China
[4] Corresponding author, zhuyun@gnnu.edu.cn, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China
[5] maghyu@hdu.edu.cn, School of Sciences, Hangzhou Dianzi University, Hangzhou, China
[6] wangjungnnu@163.com, School of Physics and Electronic Information, Gannan Normal University, Ganzhou, China

This work is licensed under a Creative Commons Attribution 4.0 International License.

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
The rapid development of Intelligent Transportation Systems (ITS) is often hindered by missing data due to technical or equipment failures, impacting data analysis and application. To address this issue and leverage the advantages of tensor completion methods in multidimensional data imputation, we propose a Bayesian Robust Tensor Ring Decomposition Model with Gaussian-Wishart priors (BRTRC). The BRTRC model structures traffic data into tensor formats such as "road segment × day × time of day" and "road segment × week × day of the week × time of day." Using tensor ring decomposition, the model reduces tensor complexity and redundancy. To approximate real values more accurately, Gaussian-Wishart priors are applied to the horizontal and frontal slices of the core factors and conjugate priors are set on the hyperparameters, allowing automatic characterisation of data changes in Bayesian modelling to avoid overfitting. Additionally, the BRTRC model introduces a sparse tensor to identify outliers in the data and employs a variational Bayesian inference method to estimate model parameters and latent variables. Experiments on two real traffic datasets demonstrate that the BRTRC model effectively imputes traffic data under both random and non-random missing patterns, robustly identifies and removes outliers and improves data quality and reliability. Extending the model to the fourth order shows superior performance in recovering high-dimensional characteristics of intersecting data compared to other models.

KEYWORDS
traffic data; tensor ring decomposition; Gaussian-Wishart prior; robust tensor; variational Bayesian.

1. INTRODUCTION

With the progress of society and the improvement of the quality of life, the number of private cars has increased significantly. This surge in private car ownership has heightened the dependence on transportation resources. Unfortunately, the development of transportation infrastructure has not kept pace with this demand, leading to frequent traffic accidents and congestion [1]. In the 1990s, the United States introduced the concept of Intelligent Transportation Systems (ITS). The core of the ITS is the collection of spatiotemporal traffic data from various sensors (e.g. cameras, floating cars and loop detectors). This real-time data – including traffic flow, speed and congestion levels – provides reliable information for decision-making in traffic policy,

planning and design, traffic control and management and information dissemination, such as traffic signal control and travel time prediction. However, equipment and network failures often result in the loss of spatiotemporal traffic data. This data is crucial for planning traffic routes and predicting vehicle travel. Missing data can decrease prediction accuracy, severely impacting the development of smart transportation [2]. Therefore, interpolating missing traffic data has become a critical research issue in ITS [3]. A substantial body of literature addresses the imputation of missing data in transportation systems. Li et al. [4] and Ni et al. [5] conducted comprehensive reviews of imputation methods for traffic data, categorising them into three types: prediction-based methods, interpolation methods and statistical learning-based methods.

Traffic data exhibit distinct temporal and spatial correlations [6]. Temporally, traffic flow and speed data collected over time show that adjacent time intervals generally have similar traffic patterns. Additionally, there is a notable similarity in traffic data between adjacent days for each road segment and detector. Spatially, roads are intricately connected and influence each other. For example, when an accident or congestion occurs downstream, the upstream speed significantly decreases while traffic volume increases. Conversely, an increase in upstream speed and volume leads to noticeable increases in downstream speed and volume. Therefore, conducting a detailed spatiotemporal analysis of traffic data can uncover intrinsic features that are beneficial for imputing missing data. To achieve this, traffic data can be represented in tensor form. For instance, traffic data collected from multiple road segments over several weeks can be organised into different representations: a third-order tensor (road segment × day × time of day), a fourth-order tensor (road segment × week × day of week × time of day), etc [7].

For the modelling of traffic data, we should first consider the strong spatiotemporal correlation model. Tensor decomposition can effectively capture the potential relationship between different dimensions of high-dimensional data and model the original multidimensional data through multi-linear structure information [8]. In tensor decomposition, common frameworks such as CANDECOMP/PARAFAC (CP) [9, 10] and Tucker decomposition [11–14] have been widely used in data recovery, image denoising, infrared detection and so on. The primary problem with interpolation of traffic data is to make robust predictions about those unobservable entries, rather than estimating model parameters. This has inspired researchers to adopt a probabilistic approach to solving the missing data interpolation problem. First, Salakhuttinov and Mnih et al. developed a Bayesian matrix decomposition algorithm, which provided the basis for this probabilistic algorithm [15]. Chen et al. extended this Bayesian matrix algorithm to higher orders and developed a Bayesian Gaussian CP rank tensor decomposition model (BGCP) to learn underlying statistical patterns in spatio-temporal traffic data and developed a Markov Chain Monte Carlo algorithm (MCMC) to estimate this model [16]. Tucker rank decomposition decomposes the original tensor into a core tensor and a set of factor matrices, which is more expressible than the CP decomposition for high-dimensional data. Tan et al. developed a traffic data completion algorithm based on Tucker decomposition, which first represented traffic data as a dynamic tensor model [7]. Gong et al. also developed a data interpolation method based on de-trending and tensor decomposition, which fully considered the spatial-temporal characteristics of traffic flow and well filled the missing part of data [17].

The recovery of high dimensional traffic data presents the challenge of tensor completion. Recently, novel tensor decomposition methods have emerged, including tensor tubal rank decomposition [18], tensor train decomposition (TT) [19] and tensor ring decomposition (TR) [20, 21]. These methods, along with factory-based models and alternate least squares (ALS), have led to a proliferation of tensor completion models. Notable examples include Tubal-ALS with tube rank [22], TT-ALS with TT rank [23] and TR-ALS with tensor ring (TR) rank [24]. Long et al. took the lead in integrating TR rank into Bayesian networks and produced a Bayesian tensor ring decomposition method [25]. Despite their contributions, both the tube rank and the TT rank exhibit limitations. The utility of the tube rank is limited to third-order tensors and cannot handle higher-dimensional data. The rigidity of the TT rank in dealing with diverse data comes from its specific requirements for the first and last core tensor rank in the decomposition. The inherent ring structure of the TR rank mitigated this problem. Unfortunately, none of these methods take into account noise and outliers, which are unavoidable in actual traffic data. To solve this problem, Zhao et al. extended the traditional Bayesian CP rank tensor decomposition by introducing sparse tensor modelling and enhancing robustness through hierarchical Student-t distribution [26]. However, this method is not suitable for traffic data that has damaged temporal and spatial correlation. Therefore, Wang et al. proposed a Bayesian robust CP decomposition model (BRCP) for traffic data completion based on sparse tensor modelling [27]. Taking the above problems into consideration, in order to reduce the complexity and redundancy of the tensor, we propose to decompose high-dimensional traffic data using tensor rings. TR can explore more potential structures than TT due to its more balanced ring

structure and smaller rank [25]. Additionally, the presence of noise and outliers significantly impacts the performance and accuracy of data imputation. To address this challenge, we introduced a sparse tensor in the BRTRC model to represent outliers in the data and employed a Variational Bayesian (VB) inference method to estimate model parameters and latent variables.

The paper makes three main contributions: (1) It proposes a Bayesian robust tensor ring decomposition model for traffic data completion; (2) By fully considering the spatiotemporal characteristics and distribution features of traffic data, we perform a fully Bayesian treatment on the tensor TR decomposition. Gaussian-Wishart priors are placed on the hyperparameters of the horizontal and frontal slices of the core factor and sparse tensor modelling is introduced, which are intended to enhance imputation performance; (3) Experiments on real datasets demonstrate that the BRTRC model can produce accurate imputations for traffic data in both random missing scenarios and non-random scenarios.

The rest of this paper is divided into five sections. Section 1 explains some of the symbols and their definitions that are commonly used in the paper, as well as some preliminary knowledge of calculations and tensor ring decomposition that are commonly used in tensor learning. Section 2 describes the Bayesian robust tensor ring decomposition model. Section 3 details the derivation process of the model in the VB framework. Section 4 validates the model by putting it together with five other models on two real spatiotemporal traffic datasets and comparing the completion effects of several models under random missing (RM) and non-random missing (NM) scenarios. Section 5 summarises the paper.

2. NOTATIONS AND PRELIMINARIES

In this section, we first introduce some of the basic tensor symbols used in this paper and then give an algebraic framework for the tensor ring decomposition. More details on the tensor ring decomposition can be found in [25, 28, 29].

2.1 Notations

First, some symbols commonly used in tensor decomposition [8]. x, $\mathbf{x}$, X, X represents scalar, vector, matrix, tensor respectively. $p(\bullet)$ and $q(\bullet)$ represent probability density distributions. The fibre of a tensor refers to keeping one dimension changing and fixing the other dimensions. The column, row and tube fibres of the third-order tensor $\mathrm{Z}^{(n)} \in \mathrm{R}^{R_{n-1} \times I_n \times R_n}, n = 1,...,N$ are denoted respectively as $\mathrm{Z}^{(N)}(:, i_n, r_n)$, $\mathrm{Z}^{(N)}(r_{n-1}, :, r_n)$, $\mathrm{Z}^{(N)}(r_{n-1}, i_n, :)$. The slice of a tensor is equivalent to a two-dimensional matrix, which keeps two dimensions changing and fixes the other dimensions. The horizontal, lateral and frontal slices of the third-order tensor are denoted as: $\mathrm{Z}^{(N)}(r_{n-1}, :, :)$, $\mathrm{Z}^{(N)}(:, i_n, :)$, $\mathrm{Z}^{(N)}(:, :, r_n)$. $Tr(X)$ represents the trace operation of a square matrix, where $X \in \mathrm{R}^{I \times I}$.

$vec(\mathrm{X})$ represents the vectorisation of tensors. $\langle \mathrm{X}, \mathrm{Y} \rangle = \langle vec(\mathrm{X}), vec(\mathrm{Y}) \rangle$ represents the inner product of tensors X and Y. $\|\mathrm{X}\|_F = \langle \mathrm{X}, \mathrm{X} \rangle^{\frac{1}{2}}$ represents the Frobenius norm of tensor X, which is the square root of the sum of squares of absolute values of all elements in the vector, commonly used to compute similarity between tensors. $\mathrm{X} \otimes \mathrm{Y} \in \mathrm{R}^{I_1 J_1 \times \ldots \times I_N J_N}$ represents the Kronecker product operation between tensors X and Y, where $\mathrm{X} \in \mathrm{R}^{I_1 \times \ldots \times I_N}$, $\mathrm{Y} \in \mathrm{R}^{J_1 \times \ldots \times J_N}$. $\mathrm{Z} = \mathrm{X} \square \mathrm{Y}$ denotes the Hadamard product of two tensors, with its elements represented as $\mathrm{Z}_{i_1, i_2, \ldots i_N} = \mathrm{X}_{i_1, i_2, \ldots i_N} \mathrm{Y}_{i_1, i_2, \ldots i_N}$, where $\mathrm{X}_{i_1, i_2, \ldots i_N}$ and $\mathrm{Y}_{i_1, i_2, \ldots i_N}$ respectively denote entries of tensors X and Y.

2.2 Tensor ring decomposition

Definition 1 (Tensor Ring Decomposition (TR Decomposition)) [28]: For an N[th]-order tensor $\mathrm{Y} \in \mathrm{R}^{I_1 \times \ldots \times I_N}$, it can be decomposed by using the TR decomposition into a series of latent core factor tensors $\mathrm{Z}^{(n)} \in \mathrm{R}^{R_{n-1} \times I_n \times R_n}$. The decomposition model can be represented as follows:

$$\mathrm{Y}(i_1, i_2, \ldots i_N) = Tr(\mathrm{Z}^{(1)}(:, i_1, :)\ \mathrm{Z}^{(2)}(:, i_2, :)\ \ldots\ \mathrm{Z}^{(N)}(:, i_N, :)) \tag{1}$$

The tensor ring decomposition can be succinctly grasped as the trace of the product sequence of mode-2 slices of tensor $\mathrm{Z}^{(n)}(:, i_n, :)$. In this context, $\mathrm{Z}^{(n)}(:, i_n, :) \in \mathrm{R}^{R_{n-1} \times R_n}$ represents the i_n[th] horizontal slice of the core tensor

$Z^{(n)}$, $Tr(\bullet)$ signifies the trace operation of a matrix and $(R_0, R_1,...,R_N)$ denotes the TR rank, where $R_0 = R_N$. To enhance clarity, the aforementioned expression can be presented in indexed form:

$$\mathrm{Y}(i_1,i_2,...i_N) = \sum_{r_1,r_2,...r_N}^{i_1,i_2,...i_N} \prod_{n=1}^{N} Tr(\mathrm{Z}^{(n)}(r_{n-1}, i_n, r_n)) \tag{2}$$

where $1 \le r_n \le R_n$ is an index for latent dimensions and $1 \le i_n \le I_n$ is an index for data dimensions. It is worth noting that any two adjacent latent tensors, such as $\mathrm{Z}^{(n)}$ and $\mathrm{Z}^{(n+1)}$, have equivalent dimensions on their corresponding modes, denoted as r_n. For convenience, we abbreviate the TR decomposition as follows: $\mathrm{Y} = \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N))$

The tensor ring decomposition is shown schematically in *Figure 1*

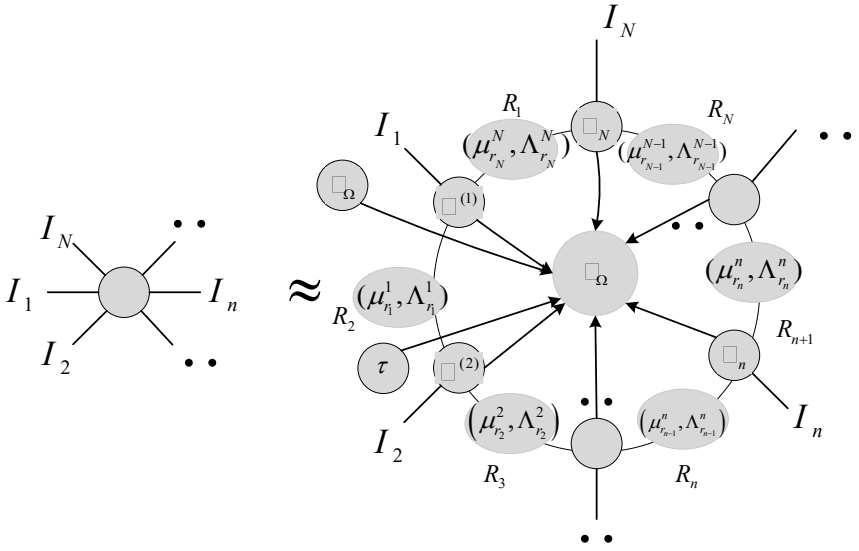

Figure 1 – Tensor ring decomposition

Definition 2 (Tensor Permutation) [25]: For an N[th]-order tensor $\mathrm{Y} \in \mathrm{R}^{I_1 \times ... \times I_N}$, a tensor permutation is defined as $\mathrm{Y}^{P_n} \in \mathrm{R}^{I_n \times ... \times I_N \times I_1 ... \times I_{n-1}}$:

$$\mathrm{Y}^{P_n}(i_n,i_2,...,i_N,i_1,i_2,...,i_{n-1}) = \mathrm{Y}(i_1,i_2,...i_N) \tag{3}$$

Theorem 1 (Cyclic Permutation Property) [28]: Based on the definitions of tensor permutation and TR decomposition, it can be concluded that the tensor permutation of tensor Y is equivalent to its factors circular shifting, as shown below:

$$\mathrm{Y}^{P_n}(i_n,i_2,...,i_N,i_1,i_2,...,i_{n-1}) = Tr(\mathrm{Z}^{(1)}(:,i_1,:)\ \mathrm{Z}^{(2)}(:,i_2,:)\ldots\mathrm{Z}^{(N)}(:,i_N,:)) = \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N)) \tag{4}$$

Definition 3 (Tensor Contraction Product (TCP)) [29]: The tensor contraction product of a third-order core factor tensor A is defined as follows:

$$\mathrm{Z} = TCP(\mathrm{Z}^{(1)}, \mathrm{Z}^{(2)} \ldots, \mathrm{Z}^{(N)}) \in \mathrm{R}^{R_0 \times (I_1 ... I_N) \times R_N} \tag{5}$$

and $\mathrm{Z}^{(\backslash n)}$, which is the tensor contraction product of a set of core factors excluding the n^{th} core tensor, is defined as follows:

$$\mathrm{Z}^{(\backslash n)} = TCP(\mathrm{Z}^{(n+1)}, \mathrm{Z}^{(N+2)} \ldots, \mathrm{Z}^{(N)}, \mathrm{Z}^{(1)}, \ldots, \mathrm{Z}^{(n-1)}) \in \mathrm{R}^{R_n \times (I_{n+1} ... I_N I_1 ... I_{n-1}) \times R_{n-1}} \tag{6}$$

3. MODEL DESCRIPTION

Given an incomplete tensor $\mathrm{Y}_{\Omega} \in \mathrm{R}^{I_1 \times \ldots \times I_N}$, let Y_{Ω} denote the incomplete tensor consisting of all observed entries, which are defined by $\{\mathrm{Y}_{(i_1,i_2,\ldots i_N)} \mid (i_1,i_2,\ldots i_N) \in \Omega\}$. Furthermore, we are given an indicator tensor O of the same size as Y_{Ω}, which equals 1 if the entry in $O_{i_1,i_2,\ldots i_N}$ coincides with the observed entry, i.e. $(i_1,i_2,\ldots i_N) \in \Omega$, otherwise it is 0. In the framework of Bayesian robust tensor model, the incomplete tensor Y_{Ω} can be expressed as follows:

$$\mathrm{Y}_{\Omega} = \mathrm{L} + \mathrm{S} + \varepsilon \tag{7}$$

where L represents the low-rank part that represents the global information generated by the TR decomposition, S is the sparse tensor that represents the local information and ε is the isotropic noise.

On the basis of the TR decomposition, assuming that all observed entries obey a Gaussian distribution, the generative model for the data is as follows:

$$\mathrm{Y}_{\Omega} \sim \mathrm{N}\left(\mathrm{R}\left(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2), \ldots, \mathrm{Z}^{(N)}(i_N)\right) + \mathrm{S}_{i_1,i_2,\ldots i_N}, \tau^{-1}\right)^{O_{i_1,i_2,\ldots i_N}} \tag{8}$$

where $\mathrm{N}(\bullet)$ denotes the Gaussian distribution, $\mathrm{S}_{i_1,i_2,\ldots i_N}$ is the sparse value corresponding to the sparsity on the observation entry and τ is the noise accuracy.

In order to model accurately, prior distributions are set on the core tensor, sparse term and noise accuracy. From *Figure 1* it can be seen that each core tensor $\mathrm{Z}^{(n)}$ in the tensor ring decomposition shares dimension r_{n-1} with its preceding core tensor $\mathrm{Z}^{(n-1)}$ and dimension r_n with its succeeding core tensor $\mathrm{Z}^{(n+1)}$, so two conjugate priors $\Lambda_{r_{n-1}}^{n-1}$ and $\Lambda_{r_n}^{n}$ are needed to control the distribution of the covariance matrices in different dimensions. Such a design allows the model to share information between different core tensors and to be constrained by common hyperparameters. Placing a multivariate Gaussian distribution on the core tensor $\{\mathrm{Z}^{(n)}\}_{n=1}^{N}$ and setting conjugate priors $\mu_{i_n}^{(n)}$ and $\Lambda_{r_{n-1}}^{n-1} \otimes \Lambda_{r_n}^{n}$ on the hyperparameters,

$$\begin{aligned} p(\mathrm{Z}^{(n)} \mid \mu_{i_n}^{(n)}, \Lambda_{r_{n-1}}^{n-1}, \Lambda_{r_n}^{n}) &= \prod_{i_n}^{I_n} \prod_{r_{n-1}}^{R_{n-1}} \prod_{r_n}^{R_n} \mathrm{N}\left(\mathrm{Z}^{(n)}(r_{n-1}, i_n, r_n) \mid \mu_{i_n}^{(n)}, (\Lambda_{r_{n-1}}^{n-1} \otimes \Lambda_{r_n}^{n})^{-1}\right) \\ &= \prod_{i_n}^{I_n} \mathrm{M}\,\mathrm{N}\left(vec(\mathrm{Z}^{(n)}(i_n)) \mid \mu_{i_n}^{(n)}, (\Lambda_{r_{n-1}}^{n-1} \otimes \Lambda_{r_n}^{n})^{-1}\right) \quad \forall n \in [1, N] \end{aligned} \tag{9}$$

where $vec(\bullet)$ denotes the vectorisation operator, $\otimes$ is the Kronecker product, $\mu_{i_n}^{(n)}$ is the mean of $\mathrm{Z}^{(n)}(i_n)$, and $\Lambda_{r_{n-1}}^{n-1} \in \mathrm{R}^{R_{n-1} \times R_{n-1}}$ and $\Lambda_{r_n}^{n} \in \mathrm{R}^{R_n \times R_n}$ are both covariance matrices controlling the r_{n-1} th horizontal and r_n th frontal slices of $\mathrm{Z}^{(n)}$, respectively, and their inverses are the accuracies in the Gaussian distribution. To enhance the robustness of the model and the updating of the hyper-parameters, we refer to the paper of Chen et al. [11] for the placement of Gaussian-Wishart distributions on the prior parameters of the horizontal and frontal slices.

$$\left(\mu_{r_{n-1}}^{n-1}, \Lambda_{r_{n-1}}^{n-1}\right) \sim \mathrm{N}\left(\mu_{r_{n-1}}^{n-1} \mid \mu_0, (\beta_0 \Lambda_{r_{n-1}}^{n-1})^{-1}\right) \times \mathrm{W}\left(\Lambda_{r_{n-1}}^{n-1} \mid W_0, v_0\right) \tag{10}$$

$$\left(\mu_{r_n}^{n}, \Lambda_{r_n}^{n}\right) \sim \mathrm{N}\left(\mu_{r_n}^{n} \mid \mu_1, (\beta_1 \Lambda_{r_n}^{n})^{-1}\right) \times \mathrm{W}\left(\Lambda_{r_n}^{n} \mid W_1, v_1\right) \tag{11}$$

where $\mathrm{W}(\bullet)$ denotes the Wishart distribution with probability density function (12)

$$\mathrm{W}\left(\Lambda \mid W, v\right) = \frac{1}{C} \mid \Lambda \mid^{\frac{1}{2}(v - I_N - 1)} \exp\left(-\frac{1}{2} Tr(W^{-1}\Lambda)\right) \tag{12}$$

where W is the scale matrix, $W_0 \in \mathrm{R}^{R_{n-1} \times R_{n-1}}$, $W_1 \in \mathrm{R}^{R_n \times R_n}$ are a positive definite matrix used to control the distribution of the precision ground, C is a normalisation constant, v is the degree of freedom used to control the convergence speed and the number of samples of the precision matrix, and I_N is the slice dimension.

Using a hierarchical induced prior, a Gaussian distribution with zero mean and accuracy η is placed over the index entries of the sparse tensor S_Ω, whose model is expressed as follows:

$$p(\mathrm{S}_\Omega \mid 0,\eta) = \prod_{i_1,i_2,\ldots i_N} \mathrm{N}\left(\mathrm{S}_{i_1,i_2,\ldots i_N} \mid 0,\eta^{-1}_{i_1,i_2,\ldots i_N}\right)^{O_{i_1 i_2 \cdots i_N}} \tag{13}$$

On the accuracy η we place the Gamma distribution:

$$p(\eta) = \prod_{i_1,i_2,\ldots i_N} \mathit{G}\mathrm{amma}\left(\eta_{i_1,i_2,\ldots i_N} \mid a_0^\eta, b_0^\eta\right) \tag{14}$$

Similarly, a Gamma distribution is placed on the noise precision τ in the model:

$$p(\tau) = \mathit{G}\mathrm{amma}\left(\tau \mid a_0^\tau, b_0^\tau\right) \tag{15}$$

Both a_0 and b_0 are hyperparameters placed on Gamma distributions and their probability density functions are given by $\mathit{G}\mathrm{amma}\left(\tau \mid a_0, b_0\right) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} \exp\left(-b\tau\right)$.

According to the mean-field approximation, the joint distribution of the Bayesian Robust Tensor Ring Decomposition model is as follows:

$$p(\mathrm{Y}_\Omega,\Theta) = p(\mathrm{Y}_\Omega \mid \{\mathrm{Z}^{(n)}\}_{n=1}^N, \mathrm{S}_\Omega, \tau) \times \prod_{n=1}^N p(\mathrm{Z}^{(n)} \mid \mu_{i_n}^{(n)}, \Lambda_{r_{n-1}}^{n-1}, \Lambda_{r_n}^n) p(\mu_{r_{n-1}}^{n-1}, \Lambda_{r_{n-1}}^{n-1}) p(\mu_{r_n}^n, \Lambda_{r_n}^n) p(\mathrm{S}_\Omega, \eta) p(\eta) p(\tau) \tag{16}$$

For simplicity, all parameters and hyperparameters are represented by Θ and unknown entries $\mathrm{Y}_{\backslash\Omega}$ are inferred by observing the index entries Y_Ω. The expression for the predictive distribution is given by

$$p(\mathrm{Y}_{\backslash\Omega} \mid \mathrm{Y}_\Omega) = \int p(\mathrm{Y}_{\backslash\Omega} \mid \Theta) p(\Theta \mid \mathrm{Y}_\Omega) d\Theta \tag{17}$$

where $p(\Theta \mid \mathrm{Y}_\Omega) = p(\Theta,\ \mathrm{Y}_\Omega) / \int (\Theta,\ \mathrm{Y}_\Omega) d\Theta$.

4. VARIATIONAL BAYESIAN POSTERIOR INFERENCE FOR MODELS

Variational Bayesian (VB) inference is a method used to approximate the posterior distribution of complex Bayesian models by approximating the intricate posterior distribution with a simpler parametric distribution. By minimising the Kullback-Leibler (KL) divergence, an appropriate distribution is found to approximate the true posterior distribution of the model. This process can be expressed as follows [30]:

$$KL(q(\Theta) \,||\, p(\Theta \mid \mathrm{Y}_\Omega)) = \int q(\Theta) \ln\{\frac{q(\Theta)}{p(\Theta \mid \mathrm{Y}_\Omega)}\} d\Theta = \ln p(\mathrm{Y}_\Omega) - \int q(\Theta) \ln\{\frac{p(\mathrm{Y}_\Omega,\Theta)}{q(\Theta)}\} d\Theta \tag{18}$$

where $\int q(\Theta) \ln\{\frac{p(\mathrm{Y}_\Omega,\Theta)}{q(\Theta)}\} d\Theta$ is the lower line of evidence for the model (ELBO), $\ln p(\mathrm{Y}_\Omega)$ is a constant about $q(\Theta)$. According to the mean-field approximation, the approximate posterior distribution is as follows:

$$q(\Theta) = \prod_{n=1}^N p(\mathrm{Z}^{(n)}) p(\mu_{r_{n-1}}^{n-1}, \Lambda_{r_{n-1}}^{n-1}) p(\mu_{r_n}^n, \Lambda_{r_n}^n) p(\mathrm{S}_\Omega, \eta) p(\eta) p(\tau) \tag{19}$$

This implies that each parameter in Θ is separable from the others, allowing for iterative adjustments of each parameter without affecting the others. Therefore, the optimisation of the model can be expressed as follows:

$$\ln q_j(\Theta_j) = E_{q(\Theta \backslash \Theta_j)}[\ln p(\mathrm{Y}_\Omega, \Theta)] + const \tag{20}$$

$E_{q(\Theta \backslash \Theta_j)}[\bullet]$ represents the posterior distribution of all parameters in Θ except for the variable Θ_j.

4.1 Posterior distribution of the core tensor $q(\mathrm{Z}^{(n)})$

Assume that the posterior distribution $q(\mathrm{Z}^{(n)})$ of the core tensor follows a Gaussian distribution, which is expressed as follows:

$$q(\mathrm{Z}^{(n)}) = \prod_{i_n=1}^{I_n} \mathrm{N}\left(vec\left(\mathrm{Z}^{(n)}(i_n)\right) \middle| \tilde{\mu}_{i_n}^{(n)}, \mathbf{V}_{i_n}^{n}\right) \tag{21}$$

Based on *Equation 20*, the variational posterior distribution of the core tensor, denoted as $q(\mathrm{Z}^{(n)})$, is derived as follows:

$$\begin{aligned} &\ln q(\mathrm{Z}^{(n)}) = E_{q(\Theta \setminus \mathrm{Z}^{(n)})}[\ln p(\mathrm{Y}_{\Omega}, \Theta)] + const \\ &= -\frac{1}{2}[(vec\left(\mathrm{Z}^{(n)}(i_n)\right)^{T} (E(\tau)E[(\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})^{T}\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})] + E[\Lambda_{r_{n-1}}^{n-1} \otimes \Lambda_{r_n}^{n}])vec\left(\mathrm{Z}^{(n)}(i_n)\right) \\ &-2vec\left(\mathrm{Z}^{(n)}(i_n)\right)^{T} (E(\tau)E[(\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})^{T}](\mathrm{Y}_{\mathrm{O}_{in}} - E[\mathrm{S}_{\mathrm{O}i_n}])) + const \end{aligned} \tag{22}$$

Therefore, the mean and covariance are updated as follows:

$$\tilde{\mu}_{i_n}^{(n)} = \mathbf{V}_{i_n}^{n} E(\tau)E[(\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})^{T}](\mathrm{Y}_{\mathrm{O}_{in}} - E[\mathrm{S}_{\mathrm{O}i_n}]) \tag{23}$$

$$\mathbf{V}_{i_n}^{n} = (E(\tau)E[(\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})^{T}\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)})] + E[\Lambda_{r_{n-1}}^{n-1} \otimes \Lambda_{r_n}^{n}])^{-1} \tag{24}$$

where $\mathrm{Y}_{\mathrm{O}_{in}} \in \mathrm{R}^{(I_{n+1}\ldots I_N I_1 \ldots I_{n-1})_{\mathrm{O}i_n} \times 1}$, $\mathrm{Z}_{\mathrm{O}_{in}}^{(/n)} \in \mathrm{R}^{(I_{n+1}\ldots I_N I_1 \ldots I_{n-1})_{\mathrm{O}i_n} \times R_{n-1}R_n}$. O_{in} denotes entries observed in $\mathrm{Y}_{\mathrm{O}_{in}}$.

4.2 Posterior distribution of hyperparameters $q(\Lambda^{(n)})$

Taking the posterior distribution of hyperparameter $\Lambda_{r_n}^{n}$ as an example, assuming the frontal slices of $\mathrm{Z}^{(n)}$ follow a Gaussian-Wishart distribution, it can be expressed using the following equation:

$$q\left(\mu_{r_n}^{n}, \Lambda_{r_n}^{n}\right) = \mathrm{N}\left(\mu_{r_n}^{n} \middle| \tilde{\mu}_{r_n}^{n}, (\tilde{\beta}_1 \Lambda_{r_n}^{n})^{-1}\right) \times \mathrm{W}\left(\Lambda_{r_n}^{n} \middle| \tilde{W}_1, \tilde{\nu}_1\right) \tag{25}$$

According to *Equation 20*, the variational posterior distribution of the hyperparameter $q(\Lambda_{r_n}^{n})$ is derived as follows:

$$\begin{aligned} &\ln q\left(\mu_{r_n}^{n}, \Lambda_{r_n}^{n}\right) = E_{q\left(\Theta \setminus \left(\mu_{r_n}^{n}, \Lambda_{r_n}^{n}\right)\right)}[\ln p(\mathrm{Y}_{\Omega}, \Theta)] + const \\ &= \frac{1}{2}\ln|\Lambda_{r_n}^{n}| - \frac{1}{2}(\mu_{r_n}^{n} - \frac{m\bar{u}_{r_n} + \beta_1\mu_1}{m+\beta_1})^{T}[(m+\beta_1)\Lambda_{r_n}^{n}](\mu_{r_n}^{n} - \frac{m\bar{u}_{r_n} + \beta_1\mu_1}{m+\beta_1}) \\ &+\frac{1}{2}(m + v_1 - I_N - 1)\ln|\Lambda_{r_n}^{n}| \\ &-\frac{1}{2}Tr((W_1^{-1} + \sum_{i=1}^{m}(E[\mathrm{Z}^{(n)}(r_{n-1}, i_n, :)] - E[\bar{u}_{r_n}])(E[\mathrm{Z}^{(n)}(r_{n-1}, i_n, :)] - E[\bar{u}_{r_n}])^{T} \\ &+\frac{m\beta_1}{m+\beta_1}(\bar{u}_{r_n} - \mu_1)(\bar{u}_{r_n} - \mu_1)^{T})\Lambda_{r_n}^{n}) + const \end{aligned} \tag{26}$$

where the parameters are updated as follows:

$$\tilde{W}_1 = (W_1^{-1} + \sum_{i=1}^{m}(E[\mathrm{Z}^{(n)}(r_{n-1}, i_n, :)] - E[\bar{u}_{r_n}])(E[\mathrm{Z}^{(n)}(r_{n-1}, i_n, :)] - E[\bar{u}_{r_n}])^{T} + \frac{m\beta_1}{m+\beta_1}(\bar{u}_{r_n} - \mu_1)(\bar{u}_{r_n} - \mu_1)^{T})^{-1} \tag{27}$$

$$\tilde{\mu}_{r_n}^{n} = \frac{m\bar{u}_{r_n} + \beta_1\mu_1}{m+\beta_1}, \quad \tilde{\beta}_1 = \beta_1 + m, \tilde{\nu}_1 = \nu_1 + m, \tilde{\Lambda}_{r_n}^{n} = (\beta_1 + m)\Lambda_{r_n}^{n} \tag{28}$$

where $\tilde{\mu}_{r_n}^n$ is the mean vector, $\bar{u}_{r_n} \in \mathrm{R}^{R_n}$ is the mean obtained by taking the mean of each column of the r_n th frontal slice of $\mathrm{Z}^{(n)}$, μ_1 is the set initial mean, which is usually a zero vector of the same size as $\bar{u}_{r_n}$, β_1 is used to control the size of elements in the precision matrix and is generally set to 1, m is the number of samples in the slice. The posterior distribution of hyperparameter $\Lambda_{r_n}^n$ can be derived using a similar approach.

4.3 Posterior distribution of sparse tensor $q(\mathrm{S})$

Assuming that the posterior distribution of the sparse tensor follows a Gaussian distribution, i.e.

$$q(\mathrm{S}) = \prod_{(i_1,i_2,\ldots i_N)\in\Omega} \mathrm{N}\left(\mathrm{S}_{i_1,i_2,\ldots i_N} \mid \tilde{\mathrm{S}}_{i_1,i_2,\ldots i_N}, \sigma^2_{i_1,i_2,\ldots i_N}\right) \tag{29}$$

the variational a posteriori $q(\mathrm{S})$ for the hyperparameters according to *Equation 22* is derived as follows:

$$\begin{aligned} &\ln q(\mathrm{S}) = E_{q(\Theta\setminus \mathrm{S})}[\ln p(\mathrm{Y}_\Omega, \Theta)] + const \\ &= -\frac{1}{2}((E[\tau] + E[\eta_{i_1,i_2,\ldots i_N}])\mathrm{S}^2_{i_1,i_2,\ldots i_N} + \mathrm{S}_{i_1,i_2,\ldots i_N} E[\tau](\mathrm{Y}_{i_1,i_2,\ldots i_N} - \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N)))) + const \end{aligned} \tag{30}$$

The posterior mean and variance are updated as follows:

$$\sigma^2_{i_1,i_2,\ldots i_N} = (E[\tau] + E[\eta_{i_1,i_2,\ldots i_N}])^{-1} \tag{31}$$

$$\tilde{\mathrm{S}}_{i_1,i_2,\ldots i_N} = \sigma^2_{i_1,i_2,\ldots i_N} E[\tau](\mathrm{Y}_{i_1,i_2,\ldots i_N} - \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N))) \tag{32}$$

4.4 Posterior distribution of sparse tensor hyperparameters $q(\eta)$

Assuming that the posterior distribution of the hyperparameters $q(\eta)$ for each entry of the sparse tensor follows a Gamma distribution, its Gamma distribution can be represented as follows:

$$q(\eta) = \prod_{(i_1,i_2,\ldots i_N)\in\Omega} \mathrm{Gamma}\left(\eta_{i_1,i_2,\ldots i_N} \mid a^\eta_{i_1,i_2,\ldots i_N}, b^\eta_{i_1,i_2,\ldots i_N}\right) \tag{33}$$

The variational a posteriori $q(\eta)$ for the hyperparameters according to *Equation 20* is derived as follows:

$$\begin{aligned} &\ln q(\eta) = E_{q(\Theta\setminus\eta)}[\ln p(\mathrm{Y}_\Omega, \Theta)] + const \\ &= -\frac{1}{2}((a^\eta_{i_1,i_2,\ldots i_N} - 1 + \frac{1}{2})\ln \eta_{i_1,i_2,\ldots i_N} - (b^\eta_{i_1,i_2,\ldots i_N} + \frac{1}{2}E[\mathrm{S}^2_{i_1,i_2,\ldots i_N}])\eta_{i_1,i_2,\ldots i_N}) + const \end{aligned} \tag{34}$$

Parameters $a^\eta_{i_1,i_2,\ldots i_N}$ and $b^\eta_{i_1,i_2,\ldots i_N}$ can be updated as follows:

$$\begin{aligned} &\tilde{a}^\eta_{i_1,i_2,\ldots i_N} = \frac{1}{2} + a^\eta_0 \\ &\tilde{b}^\eta_{i_1,i_2,\ldots i_N} = \frac{1}{2}((\tilde{\mathrm{S}}^2_{i_1,i_2,\ldots i_N} + \sigma^2_{i_1,i_2,\ldots i_N}) + b^\eta_0) \end{aligned} \tag{35}$$

4.5 Posterior distribution of noise precision $q(\tau)$

Assuming that the posterior distribution of the hyperparameters $q(\tau)$ also follows a Gamma distribution, its Gamma distribution can be represented as follows:

$$q(\tau) = G\text{amma}\left(\tau \mid a^{\tau}, b^{\tau}\right) \tag{36}$$

The variational posteriori $q(\tau)$ for the hyperparameters according to *Equation 22* is derived as follows:

$$\begin{aligned}
&\ln q(\tau) = E_{q(\Theta\setminus\tau)}[\ln p(\mathrm{Y}_{\Omega}, \Theta)] + const \\
&= (a_0^{\tau} - 1 + \frac{1}{2}\sum_{i_1,i_2,\ldots i_N} O_{i_1,i_2,\ldots i_N})\ln\tau \\
&-(b_0^{\tau} + \frac{1}{2}E[\left\| O_{i_1,i_2,\ldots i_N} \square (\mathrm{Y}_{i_1,i_2,\ldots i_N} - \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N)) - \mathrm{S}\right\|_F^2])\tau) + const
\end{aligned} \tag{37}$$

Parameters a^{τ} and b^{τ} can be updated as follows:

$$a^{\tau} = a_0^{\tau} + \frac{1}{2}\sum_{i_1,i_2,\ldots i_N} O_{i_1,i_2,\ldots i_N} \tag{38}$$

$$b^{\tau} = b_0^{\tau} + \frac{1}{2}E[\left\| O_{i_1,i_2,\ldots i_N} \square (\mathrm{Y}_{i_1,i_2,\ldots i_N} - \mathrm{R}(\mathrm{Z}^{(1)}(i_1), \mathrm{Z}^{(2)}(i_2) \ldots, \mathrm{Z}^{(N)}(i_N)) - \mathrm{S}\right\|_F^2] \tag{39}$$

The above constitutes the variational posterior derivation of the model. The model's Evidence Lower Bound (ELBO) can be expressed as follows:

$$\mathrm{L}(q) = E_{q(\Theta)}[\ln p(\mathrm{Y}_{\Omega}, \Theta)] + H(q(\Theta)) \tag{40}$$

where $E_{q(\Theta)}[\bullet]$ represents the expectation of the posterior distribution of all hyperparameters, $H(\bullet)$ is the entropy of the posterior distribution q. $\mathrm{L}(q)$ normally decreases with the number of iterations and finally tends to flatten, with which the convergence of the model can be judged.

4.6 Algorithm implementation

In the above, the posterior derivations for each parameter and hyperparameter update have been obtained through the VB framework. As a result, we can formulate the algorithm for implementing the Bayesian Gaussian Robust Tensor Ring Decomposition (BRTRC), as shown in *Algorithm 1*:

Algorithm 1 – BRTRC Algorithm

Inputs: a N^{th}-order incomplete tensor $\mathrm{Y}_{\Omega} \in \mathrm{R}^{I_1 \times \ldots \times I_N}$, an indicator tensor $O \in \mathrm{R}^{I_1 \times \ldots \times I_N}$

Initialisation: $\mathrm{Z}^{(n)}$, $\mathbf{V}^{(n)}$, $\Lambda_{r_{n-1}}^{n-1}$, $\Lambda_{r_n}^{n}$, $1 \le n \le N$, $1 \le r_{n-1} \le R_{n-1}$, $1 \le r_n \le R_n$, σ^2, $\tilde{\mathrm{S}}$, η, τ, top-level hyperparameter μ, β, W, v, a^{η}, b^{η}, a^{τ}, b^{τ}.

repeat

for n = 1to N do

Updating $q(\mathrm{Z}^{(n)})$ according to *Equation 23* and *Equation 24*

end for

for n = 1to N do

Updating $q(\Lambda^{(n)})$ according to *Equation 27* and *Equation 28*

end for

Updating $q(\mathrm{S})$ according to *Equation 31* and *Equation 32*

Updating $q(\eta)$ according to *Equation 35*

Updating $q(\tau)$ according to *Equation 38* and *Equation 39*

Calculate the lower limit of convergence for each parameter according to *Equation 40*

until convergence

Output: a low-rank complement tensor L and a sparse tensor S.

4.7 Initialisation parameters

To ensure a well-performing model, proper initialisation and initial values for hyperparameters are crucial. Therefore, for the initialisation of the core tensor $\mathcal{Z}^{(n)}$, we adopt the TR approximation [24]. When interpolating traffic data, it is assumed that the TR rank is initially R in all dimensions, i.e. set to $R * ones(1, N+1)$. Thus for the initial means $\mu = 0$, $W_1 = W_0 = eye(R)$, $v = R$, a^{η} , b^{η} , a^{τ} , b^{τ} in the Gaussian-Wishart distribution on the prior parameters, the initial values are all 10^{-6} and the sparse tensor $\mathcal{S}_{i_1, i_2, \ldots i_N}$ is sampled as a random sample of $N(0,1)$.

5. EXPERIMENTAL ANALYSIS

In order to evaluate the effectiveness of the algorithm in recovering missing traffic data in practical applications, experiments will be conducted on two real spatiotemporal traffic datasets. The algorithm in this paper will be compared with several state-of-the-art traffic data complementation algorithms, including Bayesian Robust CP Decomposition (BRCP) [27], Tensor Ring-Variable Bayesian Inference(TR-VBI)[25], Bayesian Augmented Tensor Factor Factorisation (BATF) [16], Bayesian Gaussian CP Decomposition (BGCP) [31], High-Precision Low-rank Tensor Completion (HaLRTC) [32], SVD-Combined Tensor Decomposition Model (STD) [33]. In order to consider the data loss situation in practice, two kinds of missing scenarios are set up in the experiments, which are random missing (RM) and non-random missing (NM). For the random missing scenario, in this paper, we just randomly delete some data items in the spatiotemporal traffic dataset, but in practice, the problem of missing data is often caused by a variety of reasons, for example, due to the breakage, attenuation, interruption or interference of optical fibre, which can lead to the inability of optical fibre to transmit data for a period of time, rather than randomly losing data. As a result, data lost from optical fibres tends to show some temporal continuity and localisation, i.e. all sensors on a certain roadway are missing data during a certain period of time. To simulate this situation, a non-random missing scenario, namely Fibber Optic Disruption, will be designed following the approach introduced by Chen et al [31]. This scenario aims to replicate the situation where data loss occurs in a continuous and localised manner due to fibre optic disruptions. This realistic missing data scenario better reflects the challenges encountered in real-world traffic data applications.

For model performance evaluation, this study utilises the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) as metrics. These metrics represent the relative error and absolute error, respectively, between the predicted values and actual values.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\mathrm{Y}_{\mathrm{O}i_n} - \tilde{\mathrm{Y}}_{\mathrm{O}i_n}|}{\mathrm{Y}_{\mathrm{O}i_n}} \tag{41}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{Y}_{\mathrm{O}i_n} - \tilde{\mathrm{Y}}_{\mathrm{O}i_n})^2} \tag{42}$$

where N denotes the total number of missing data, $\mathrm{Y}_{\mathrm{O}i_n}$ is the actual value of the missing data and $\tilde{\mathrm{Y}}_{\mathrm{O}i_n}$ is the value that has been model-complemented. In addition, all experiments are simulated on Matlab 2020b on a laptop with a 12th Gen Intel Core i7-12700H CPU and 16GB RAM.

The first step is to perform an experimental simulation on the Guangzhou large-scale spatiotemporal traffic speed dataset. This dataset consists of measured vehicle speed data from 214 road sections (mainly main roads and expressways) in Guangzhou, China, over a two-month period from 1 August 2016 to 30 September 2016. The dataset has a 10-minute time window, i.e. there are 144 time windows per day and the average vehicle speed of each road section is recorded for each time window, which can be represented by a size of 214×61×144. Three-dimensional tensor, where each element represents the value of vehicle speed at a spatiotemporal location, the first dimension is the road segment, the second dimension is the day and the third dimension is the time window, and about 1.29% of the dataset is missing. (https://doi.org/10.5281/zenodo.1205228). The three kinds of BRCP, BATF and BGCP are all tensor-based CP decomposition, they all decompose the potential information into a sum of several rank one quantities, and therefore, for better interpolation results, we set the rank of all three models to 80 and the number of iterations to 200. Tensor ring decomposition is easier for high rank or sparse tensors and less core tensor can be used to

approximate the original tensor. If the rank is set too high, it will lead to an increase in the dimension of the core tensor, especially for large-scale datasets such as the Guangzhou Traffic Speed dataset, which will lead to a situation where the size of the array exceeds the maximum memory available in the system in Matlab, which cannot be run due to equipment reasons. Therefore, the rank in BRTRC is set to 30 and TR-VBI is a model based on tensor ring decomposition, so the rank is set to 30 like BRTRC, and the two models HaLRTC and STD do not need too many parameter settings.

Since only 1.29% entries are not observed in the raw data, a certain number of entries needs to be manually deleted to constitute missing data and thus divide the raw data into two groups: the observed and the missing. For these "missing" entries, we also have the corresponding ground truth, which allows us to assess the imputation performance directly. The model performance is evaluated by imputing those missing entries. Five synthetic datasets with missing rates between 10% and 50% are created based on the Guangzhou traffic speed dataset [31].

Table 1 shows a comparison of the interpolation performance of several models under random missing scenarios. The best data are highlighted in bold. The five Bayesian network-based models clearly outperform the interpolation of the STD and HaLRTC, which are common decompositions and traditional low-rank tensor complementation methods. With the exception that the HaLRTC performs a little better than the three models BRCP, BATF and BGCP at low missing rates from 10% to 20%, but the interpolation performance of HaLRTC gets worse and worse as the missing rate of the data increases, while the five Bayesian network-based models are less sensitive to the missing rate relative to the other two models. This is because the Bayesian tensor decomposition framework can utilise a priori information to guide parameter estimation rather than relying solely on the observed data. This increases the robustness and generalisation ability of the model, especially in the case of sparse or noisy data. BRCP, BATF and BGCP are tensor-based CP decompositions, while BRTRC and TR-VBI is a model based on tensor-ring decomposition, which is more complex for multimodal or heterogeneous data and can flexibly portray the correlation or difference between different modes more adequately than CP decomposition. Therefore, in the same Bayesian network, the performance of the BRTRC and the TR-VBI is much better than the other three models. Especially when the deletion rate is as low as 10% to 20%, the BRTRC and the TR-VBI still perform better than the HaLRTC. The BRTRC has better robustness due to the addition of Gaussian-Wishart priors based on the distribution characteristics of traffic data. Combined with the experimental results, it can be proved that the recovery effect of the BCTRC is better than that of the TR-VBI.

Table 1 – Performance of models under random missing scenarios MAPE/RMSE

RM	10%	20%	30%	40%	50%
	MAPE/RMSE	**MAPE/RMSE**	**MAPE/RMSE**	**MAPE/RMSE**	**MAPE/RMSE**
BRTRC(30)	**0.0663/2.9029**	**0.0677/2.9614**	**0.0689/3.0395**	**0.0729/3.1893**	**0.0756/3.3152**
TR-VBI(30)	0.0672/2.9391	0.0701/3.0581	0.0731/3.1928	0.0776/3.4107	0.0816/3.5959
BRCP(80)	0.0832/3.5926	0.0836/3.5999	0.0838/3.6160	0.0837/3.6165	0.0844/3.6429
BATF(80)	0.0828/3.5787	0.0832/3.5863	0.0833/3.5955	0.0836/3.6105	0.0841/3.6290
BGCP(80)	0.0833/3.6000	0.0838/3.6060	0.0838/3.6186	0.0841/3.6268	0.0847/3.6546
HaLRTC	0.0777/3.1917	0.0815/3.3324	0.0850/3.4748	0.0887/3.6143	0.0931/3.7730
STD	0.0880/3.7708	0.0911/3.8308	0.0936/3.9286	0.0963/4.0265	0.0993/4.1253

Table 2 shows a comparison of the interpolation performance of several models in non-random missing scenarios, with the best data highlighted in bold. Used in the non-random missingness is the fibre missingness scenario designed by Chen et al. in order to simulate realistic time-correlated missingness scenarios in the fibre missingness experiment. In this scenario the temporal correlation in the original data is destroyed, so it is difficult to borrow information from other dimensions [16]. In the case of the same loss rate, the information lost by optical fibre loss is more than that lost by non-random loss. In order to avoid overfitting in data recovery, the TR rank of the BRTRC and the TR-VBI is set to 5, and the CP rank of the BRCP, BATF and BGCP is set to 20. From the table, it can be seen that the recovery performance of all the models is drastically reduced in

the non-random scenarios compared to the random missing scenarios. Since the previously mentioned temporal correlation is destroyed, it may not be possible to find similar or related data using different times or days on the same road section, so the data recovery for the same missing rate non-random missing scenario is more difficult. The error will be larger, but the tensor decomposition framework based on the Bayesian network is still better than the performance of the traditional tensor complementation models such as the STD and the HaLRTC. In the Bayesian network, the BRTRC still has the best performance.

Table 2 – Performance of models under non-random missing scenarios MAPE/RMSE

NM	10%	20%	30%	40%	50%
	MAPE/RMSE	**MAPE/RMSE**	**MAPE/RMSE**	**MAPE/RMSE**	**MAPE/RMSE**
BRTRC(5)	**0.0957/4.0524**	**0.0959/4.0524**	**0.0963/4.0709**	**0.0967/4.0783**	**0.0967/4.0812**
TR-VBI(5)	0.1019/4.3754	0.1026/4.3953	0.1032/4.4513	0.1049/4.5411	0.1078/4.6048
BRCP(20)	0.0982/4.1352	0.0980/4.1400	0.0989/4.2197	0.1003/4.2909	0.1022/4.4497
BATF(20)	0.0976/4.1252	0.0985/4.1718	0.0987/4.2133	0.1003/4.3912	0.1034/4.5607
BGCP(20)	0.0985/4.1452	0.0984/4.1569	0.0993/4.2357	0.1011/4.3643	0.1042/4.5579
HaLRTC	0.1033/4.1576	0.1046/4.2086	0.1062/4.2792	0.1088/4.3813	0.1131/4.5271
STD	0.1019/4.1881	0.1054/4.3300	0.1068/4.4029	0.1115/4.5573	0.1133/4.6391

Figure 2 illustrates a comparison of traffic data imputation for Road Segment #1 during the first two weeks of August. The figure depicts the contrast between the BRTRC, BRCP and the ground truth values under a 30% missing data scenario. Notably, this example showcases the impact on data imputation for a specific road segment and a specific time period. The graph reveals that the traffic speed data exhibit strong spatiotemporal correlations, with periodic fluctuations over time. Except for the second day, the peaks and valleys of the traffic data curves exhibit a consistent pattern across days. This temporal correlation is a key factor enabling effective data imputation. In the scenario of random missing data, both models exhibit well-fitting curves that closely match the ground truth curves. However, the BRTRC demonstrates superior fitting performance, particularly evident in accurately capturing fluctuations, especially at highly fluctuating peak points. This comparison indicates that the BRTRC outperforms the BRCP in terms of imputing missing traffic data, particularly when dealing with significant fluctuations in the data.

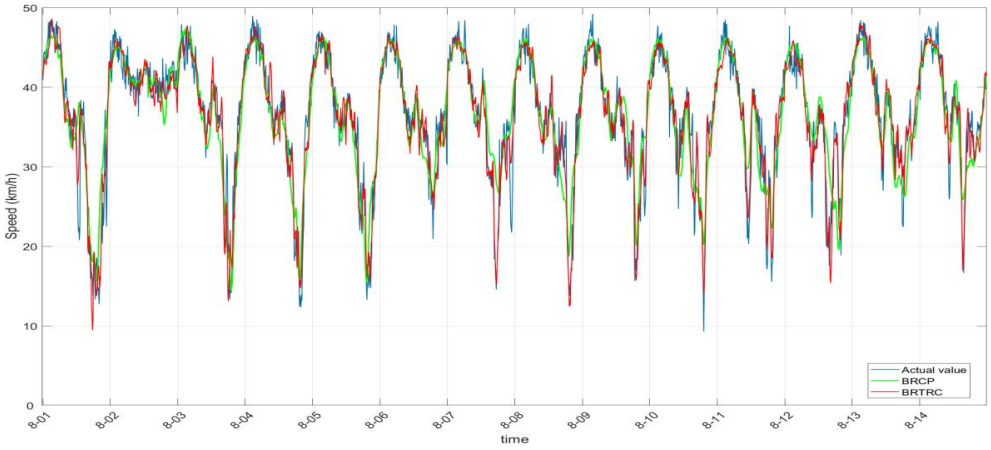

Figure 2 – Time series of the BRTRC under random missing scenario of section #1 from 1 to 14 August

Figures 3 and *Figures 4* respectively depict experiments in a non-random missing data scenario for Road Segment #1. The scenarios involve 50% missing data rates for the periods from 1 August to 14 August and 12 September to 25 September. These timeframes represent the traffic flow conditions at the beginning of August and the end of September. In these figures, the green rectangles represent the portions affected by fibre optic disruptions, while the white rectangles represent the observed values. Notably, in the non-random missing data

scenario, the observed speed data for an entire day is missing. Approximating the missing parts using adjacent dates becomes challenging in this context. As a result, the curves representing the recovered data by the models deviate more noticeably from the ground truth curves compared to the scenario of random missing data. Even in such challenging scenarios, the performance of the BRTRC model outshines other models in terms of data imputation. Taking 15 September as an example, the errors between the recovered values from the BRTRC model and the ground truth values are significantly smaller, especially at the peaks of the curves. These observations indicate that while the non-random missing data scenario presents difficulties in approximating the missing data, the BRTRC model still exhibits better data imputation performance compared to other models.

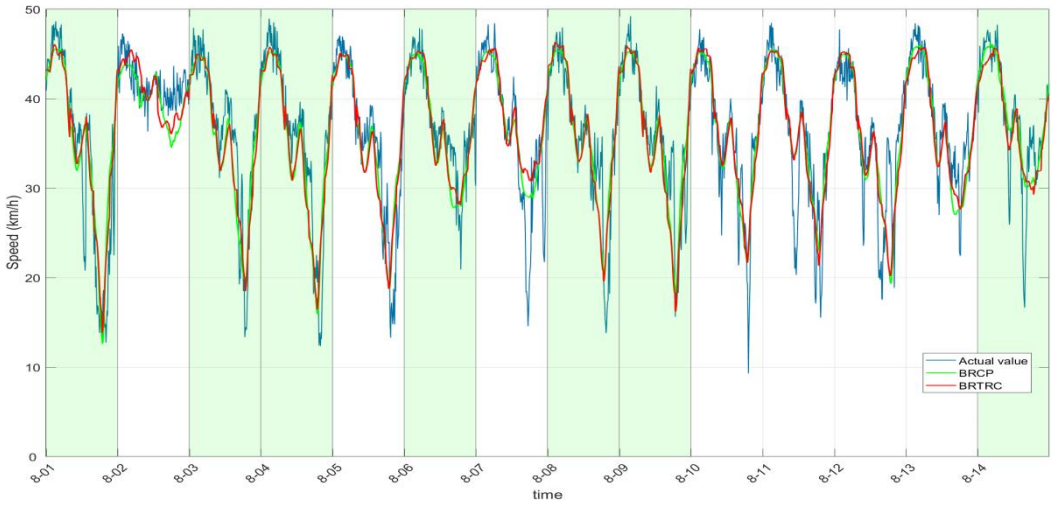

Figure 3 – Time series of the BRTRC under non-random missing scenario of section #1 from 1 to 14 August

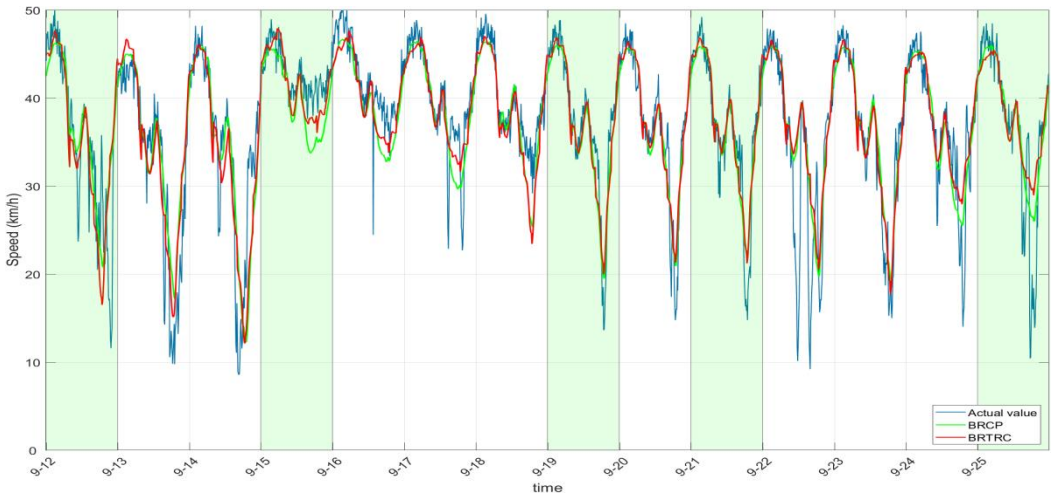

Figure 4 – Time series of the BRTRC under non-random missing scenario of section #1 from 12 to 25 September

To further validate the superiority of the model in this paper in terms of recovery performance for high-dimensional traffic data, the model is extended into a higher order fourth-order tensor and the traffic data is partitioned into 214 × 9 × 7 × 144, where each element represents the value of the vehicle speed in a spatiotemporal location: the first dimension is the road segment, the second dimension is the week, the third dimension is the day of week and the fourth dimension is the t time interval. Since the STD mainly solves the problem of non-convexity, it does not provide too much improvement in the fourth-order tensor, and therefore does not participate in the comparison. *Table 3* and *Table 4* show the comparisons of traffic data recovery performance in random missing scenarios and non-random missing scenarios after the model is extended to the fourth order, respectively. Due to the high computational cost of the fourth-order tensor, the ranks of the BRCP, the BATF and the BGCP are only set to 50 and the ranks of the BRTRC and the TR-VBI are still set to 30. Combined with the data in the content of the previous chapter for comparison, it is obvious to find that the performance of the recovery of traffic data in the case of the same rank is not as good as that in the case of

third-order under the expression of fourth-order tensor. This verifies Chen's result in his thesis that the recovery effect of traffic data is related to the expression of the data and the recovery effect of the data in the third-order expression under the same model is better than that of the fourth-order expression [31]. In addition, through the data comparison in the table, it can be clearly seen that the recovery effect of the BRTRC in high-dimensional traffic data in random missing scenarios is better than that of other models. In the non-random deletion mode, we can find that when the deletion rate is 10%, the recovery effect of the BAFT is better than that of other models. With the increasing of the deletion rate, the recovery effect of the BRTRC gradually exceeded that of other models. The BRTRC models with 20% and 30% deletion rates both had the highest MAPE, and when the deletion rates reached 40% and 50%, the recovery effect of the BRTRC was significantly better than that of other models.

Table 3 – Performance of the models in a fourth-order tensor random missing scenario

RM	10%	20%	30%	40%	50%
	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE
BRTRC(30)	**0.0693/3.0429**	**0.0697/3.1064**	**0.0723/3.1852**	**0.0762/3.2993**	**0.0786/3.4052**
TR-VBI(30)	0.0701/3.3118	0.0737/3.2654	0.0759/3.3041	0.0798/3.4802	0.0872/3.8151
BATF(50)	0.0867/3.7243	0.0871/3.7331	0.0874/3.7414	0.0876/3.7517	0.0876/3.7634
BRCP(50)	0.0872/3.6924	0.0886/3.7399	0.0878/3.8160	0.0885/3.8165	0.0897/3.8429
BGCP(50)	0.0896/3.8517	0.0900/3.8540	0.0898/3.8548	0.0900/3.8608	0.0903/3.8740
HaLRTC	0.0776/3.1716	0.0817/3.3231	0.0857/3.4802	0.0900/3.6397	0.0951/3.8259

Table 4 – Performance of models in non-random missing scenarios with fourth-order tensor

NM	10%	20%	30%	40%	50%
	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE
BRTRC(5)	0.1027/4.3083	**0.1006**/4.2362	**0.1013**/4.3117	**0.1018/4.3297**	**0.1039/4.4802**
TR-VBI(5)	0.1050/4.4993	0.1060/4.5553	0.1066/4.5670	0.1074/4.6501	0.1101/4.9191
BATF(20)	**0.1018/4.2397**	0.1025/4.2866	0.1029/4.3208	0.1033/4.5345	0.1055/4.7011
BRCP(20)	0.1022/4.2457	0.1014/**4.2016**	0.1013/**4.3019**	0.1027/4.4491	0.1053/4.5448
BGCP(20)	0.1019/4.2691	0.1018/4.2086	0.1014/4.3158	0.1025/4.4291	0.1062/4.6320
HaLRTC	0.1089/4.3367	0.1109/4.4173	0.1136/4.5303	0.1178/4.7054	0.1243/4.9463

The second set of experiments was conducted on the Birmingham Parking Lot Dataset. This dataset comprises parking occupancy data from 30 parking lots operated by the NCP Company, spanning a period of 77 days from 4 October to 29 December 2016. Data was collected at half-hour intervals between 8:00 a.m. and 5:00 p.m. each day, representing the occupancy status of the parking lots (i.e. the number of vehicles parked). The dataset can be represented as a tensor of the size 30x77x18. Approximately 14.89% of the data in this dataset is missing (source: https://doi.org/10.24432/C51K5Z). Similarly, synthetic data is created by artificially removing a certain number of items at random from the original data.

Table 5 provides a comparison of the recovery performance between the proposed model and the Bayesian Robust CP Decomposition (BRCP), Bayesian Augmented Tensor Factorisation (BATF), Bayesian Gaussian CP Decomposition (BGCP) and High-Accuracy Low-Rank Tensor Completion (HaLRTC) models in a random missing data scenario. The best-performing results are highlighted in bold. Given that the Birmingham Parking Lot Dataset is relatively smaller than the previous dataset, adjustments were made for accurate experimentation. In the BRTRC model, the TR rank was set to 10, while the CP rank for BRCP, BATF and BGCP was set to 30. The missing rate remained at 10% to 50%. Due to the dataset's relatively high missing

rate of 14.89% and the significant differences between data points, this can have a substantial impact on the evaluation metrics. From *Table 3*, it is evident that while the HaLRTC exhibits lower MAPE values compared to other models at 10% and 20% missing rates, its RMSE values are generally higher. This implies that the HaLRTC tends to have smaller relative errors but larger absolute errors in non-random missing data scenarios. The bolded data in the table highlights that the BRTRC model consistently achieves the lowest RMSE and relatively small relative errors across missing rates from 10% to 50%. Overall, the BRTRC model outperforms the other models in terms of recovery performance.

Table 5 – Performance of models under random missing scenarios MAPE/RMSE

RM	10%	20%	30%	40%	50%
	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE
BRTRC(10)	**0.0527**/**17.864**	0.0599/**18.643**	**0.0603**/**19.527**	0.0673/**19.979**	0.0737/**21.393**
BRCP(30)	0.0631/19.052	0.0621/19.585	0.0625/20.127	**0.0632**/20.957	**0.0648**/22.053
BATF(30)	0.0662/18.909	0.0645/19.628	0.0613/20.410	0.0693/20.973	0.0789/26.385
BGCP(30)	0.0754/19.942	0.0706/21.016	0.0652/21.717	0.0726/22.321	0.0754/24.300
HaLRTC	0.0547/18.753	**0.0598**/19.810	0.0637/24.284	0.0780/28.857	0.0910/31.230

Table 6 presents a comparison of the imputation performance of various models in a non-random missing data scenario. Overall, the imputation performance of the models significantly declines in the non-random missing data scenario. This reduction is particularly notable due to the dataset's inherent high missing rate and substantial differences between data points, resulting in larger discrepancies between evaluation metrics across models. Notably, the performance of the HaLRTC model is severely affected in this scenario due to its sensitivity to missing rates and noise. Both relative and absolute errors increase dramatically with higher missing rates. Conversely, the BRTRC model demonstrates robustness in the face of such challenges. Its MAPE values exhibit minimal variation and its RMSE values consistently remain below 100, outperforming the other models. This highlights the strong robustness of the BRTRC model even in the presence of non-random missing data and significant variations between data points.

Table 6 – Performance of models under non-random missing scenarios MAPE/RMSE

NM	10%	20%	30%	40%	50%
	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE	MAPE/RMSE
BRTRC(5)	0.1160/**38.891**	0.1134/**47.664**	0.1173/**48.508**	**0.1188**/**69.653**	**0.1212**/**82.073**
BRCP(10)	0.1524/41.376	0.1423/47.293	0.1787/56.003	0.1731/80.943	0.1724/101.323
BATF(10)	0.1172/50.572	0.1276/56.644	0.1163/56.146	0.1369/94.795	0.1674/103.789
BGCP(10)	0.1254/76.724	0.1181/64.984	0.1202/55.779	0.1285/75.236	0.1743/164.961
HaLRTC	**0.0707**/60.053	**0.0929**/72.859	**0.1166**/117.86	0.1395/163.35	0.1898/221.84

Figure 5 is selected as a line graph of data recovery experiments conducted for parking lot #1 under random missing scenario, in a two-week time period from 6 December to 19 December, with 30% missing rate, where the blue line is the real value in the original dataset and the red line is the value recovered by BRTRC model. From the figure, it can be seen that the parking lot occupancy data shows periodicity over time, with peaks occurring at similar times of the day. It can be clearly seen that the data recovered by the BRTRC model under the random missing scenario and the original dataset fit very well, and the BRTRC has a very good performance.

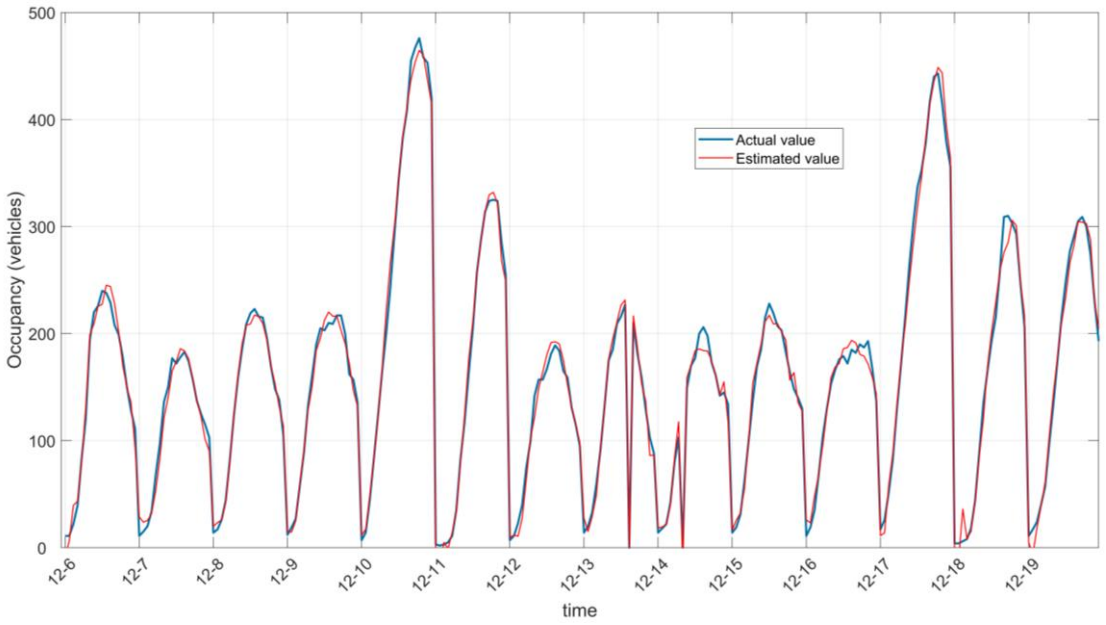

Figure 5 – Occupancy rate of the BRTRC in parking lot #1 under random missing scenario from 6 to 19 December

Figure 6 and *Figure 7* depict data recovery experiments conducted on the Parking Lot data under a non-random missing data scenario. Specifically, these figures represent the recovery experiments for two-week periods with a 30% missing data rate: 4 to 17 October (*Figure 6*) and 4 to 17 December (*Figure 7*). Similar to the fibre optic missing data in the traffic data context, the non-random missing data scenario in the parking lot data disrupts the time correlation. In both figures, the green rectangles represent the portions affected by the missing data, while the white rectangles depict the regions without missing values. The considerable variations between data points and the disruption of time correlation contribute to the reduced recovery performance in the non-random missing data scenario compared to the random missing data scenario. Consequently, there is a noticeable disparity between the imputed values and the ground truth values, particularly evident at peak points. For instance, in *Figures 7*, on 13 December, the trend of the imputed curve deviates significantly from the trend of the actual values. In summary, these figures underscore the challenge posed by non-random missing data scenarios, where disruptions in time correlation and substantial variations between data points can lead to decreased recovery performance and larger discrepancies between imputed and actual values.

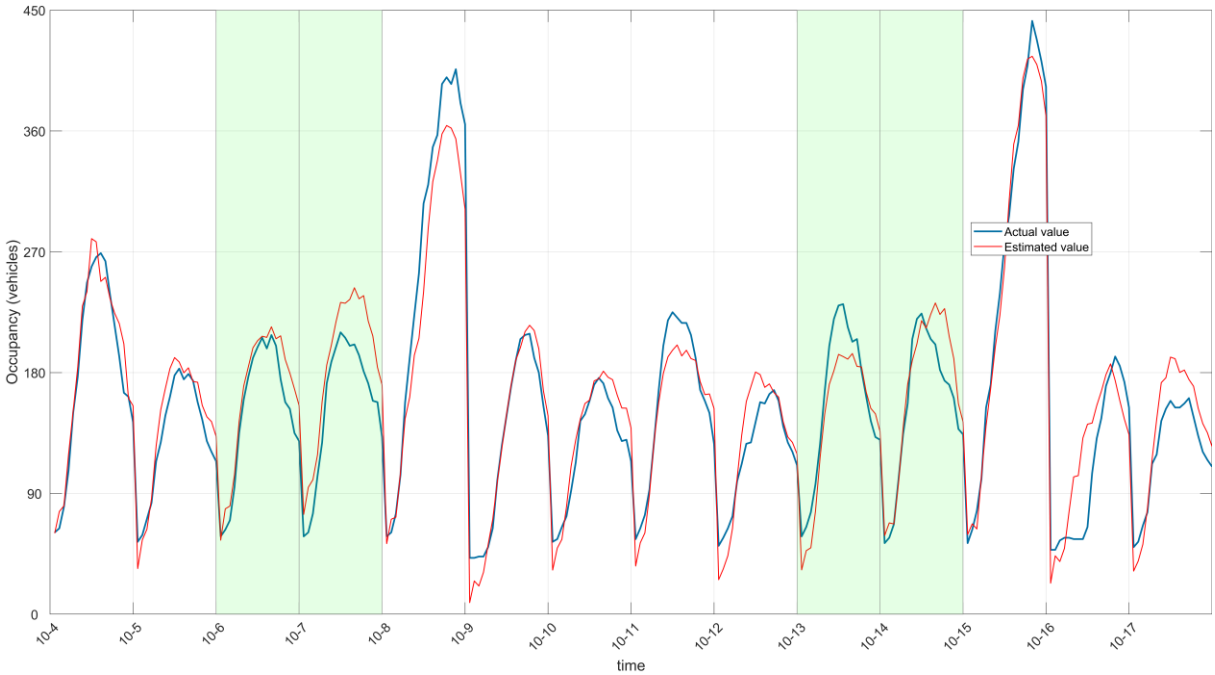

Figure 6 – Occupancy rate of the BRTRC in parking lot #1 under non-random missing scenario from 4 to 17 October

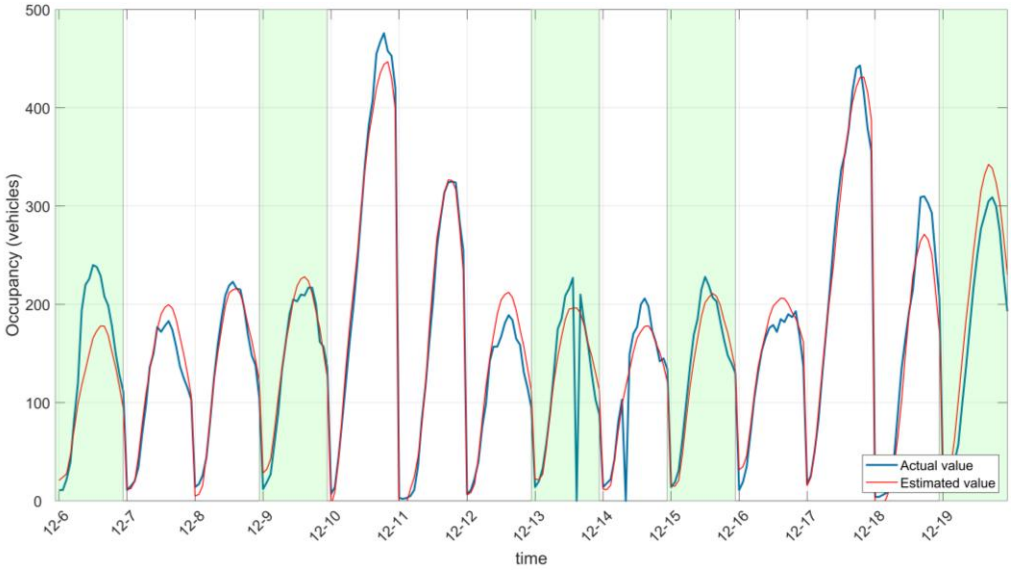

Figure 7 – Occupancy rate of the BRTRC in parking lot #1 under non-random missing scenario from 6 to 19 December

6. CONCLUSION

In this paper, we propose a Bayesian robust tensor ring decomposition algorithm applied to traffic data complementation. It has good performance in recovering missing data and strong robustness to noise and handling outliers. This should be the first application of the TR rank decomposition to recover spatio-temporal traffic data in Bayesian networks. The distributional characteristics of the traffic data improve the prior distribution method and the conjugate prior placed on the core tensor via the Gaussian-Wishart distribution gives a good recovery performance on both random missing scenarios and non-random missing scenarios. Our proposed BRTRC model has been compared to the BRCP, TR-VBI, BATF, BGCP, HaLRTC, STD, which are several state-of-the-art models in spatio-temporal traffic data recovery, on two real datasets: the Guangzhou large-scale traffic speed dataset and the Birmingham parking lot dataset. The experimental results of our model outperform the above models, and by extending the model to the fourth order, it is verified that the model has a better recovery effect for the high-dimensional characteristics of traffic data compared to other models.

ACKNOWLEDGEMENTS

This work was financially supported by the National Natural Science Foundation of China (No.12071104), the Natural Science Foundation of the Zhejiang Province (No. LD19A010002), the Science and Technology Project of the Jiangxi Provincial Education Department (No. GJJ201409) and the Key Research and Development Programs of the Jiangxi Province, China (No.20203BBF63040).

REFERENCES

[1] Tali I, Nassrullah Z, Abduljaleel L. A case study on reducing traffic congestion–proposals to improve current conditions. *Civil Engineering Journal*, 2023;9(10):2456-2466. DOI:10.28991/cej-2023-09-10-07.

[2] XU JR, LI XY, YI HJ. Short-term traffic flow forecasting model under missing data. *Journal of Computer Applications*, 2010;30(4):1117.

[3] Sun T, et al. Traffic missing data imputation: A selective overview of temporal theories and algorithms. *Mathematics*, 2022;10(14). DOI:10.3390/math10142544

[4] Li Y, Li Z, Li L, Missing traffic data: comparison of imputation methods. *IET Intelligent Transport Systems*, 2014;8(1):51-57. DOI:10.1049/iet-its.2013.0052.

[5] Ni D, et al. Multiple imputation scheme for overcoming the missing values and variability issues in ITS data. *Journal of transportation engineering*, 2005;131(12):931-938.

[6] Chen X, et al., Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting. *International Journal of Distributed Sensor Networks*, 2017;13(6):1550147717713376.

[7] Tan H. et al. Short-term traffic prediction based on dynamic tensor completion. *IEEE Transactions on Intelligent Transportation Systems*, 2016;17(8):2123-2133. DOI:10.1109/tits.2015.2513411.

[8] Kolda TG, Bader BW, Tensor decompositions and applications. *SIAM review*, 2009;51(3):455-500.

[9] Qiu Y, et al. Canonical polyadic decomposition (CPD) of big tensors with low multilinear rank. *Multimedia Tools Applications*, 2021;80(15):22987-23007.

[10] Long Z, et al. Low rank tensor completion for multiway visual data. *Signal processing*, 2019;155:301-316.

[11] Zhu Y. et al. Infrared object detection via patch-tensor model and image denoising based on weighted truncated Schatten-p norm minimization. *IET image processing*, 2023.

[12] Zhou G, et al. Efficient nonnegative tucker decompositions: Algorithms and uniqueness. *IEEE Transactions on Image Processing*, 2015;24(12):4990-5003.

[13] Qiu Y, et al. Approximately orthogonal nonnegative Tucker decomposition for flexible multiway clustering. *Science China Technological Sciences*, 2021(064-009).

[14] Qiu Y, et al. A generalized graph regularized non-negative Tucker decomposition framework for tensor data representation. *IEEE Transactions on Cybernetics,* 2022;52(1):594-607.

[15] Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. *In Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008)*, Helsinki, Finland, June 5-9, 2008.

[16] Chen X, et al. Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model. *Transportation Research*, 2019;104(JUL.):66-77.

[17] Gong C, Zhang Y. Urban traffic data imputation with detrending and tensor decomposition. *IEEE Access*, 2020;8:11124-11137.

[18] Zhang Z, et al. Novel methods for multilinear data completion and de-noising based on tensor-SVD. *In Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014.

[19] He J, et al. Low-rank tensor completion based on tensor train rank with partially overlapped sub-blocks. *Signal Processing*, 2022;190:108339.

[20] Yu J, et al. Low tensor-ring rank completion by parallel matrix factorization. *IEEE Trans Neural Netw Learn Systems*, 2021;32(7):3020-3033. DOI:10.1109/TNNLS.2020.3009210.

[21] Huang HY, et al. Provable tensor ring completion. *Signal Processing*, 2020;171:107486. DOI:10.1016/j.sigpro.2020.107486.

[22] Liu XY, et al. Low-tubal-rank tensor completion using alternating minimization. *IEEE Transactions on Information Theory,* 2020;66(3):1714-1737. DOI:10.1109/tit.2019.2959980.

[23] Grasedyck L, Kluge M, Krämer S. Variants of alternating least squares tensor completion in the tensor train format. *Siam Journal on Scientific Computing*, 2015;37(5):A2424-A2450. DOI:10.1137/130942401.

[24] Wang W, Aggarwal V, Aeron S. Efficient low rank tensor ring completion. *In Proceedings of the IEEE International Conference on Computer Vision*. 2017.

[25] Long Z, et al. Bayesian low rank tensor ring for image recovery. *IEEE Trans Image Processing*, 2021;30:3568-3580. DOI:10.1109/TIP.2021.3062195.

[26] Zhao Q, et al. Bayesian robust tensor factorization for incomplete multiway data. *IEEE transactions on neural networks learning systems*, 2016;27(4):736-48. DOI:10.1109/TNNLS.2015.2423694.

[27] Zhu Y, et al. A Bayesian robust CP decomposition approach for missing traffic data imputation. *Multimedia Tools and Applications*, 2022;81(23):33171-33184. DOI:10.1007/s11042-022-13069-7.

[28] Zhao Q, et al. Tensor ring decomposition. arXiv:1606.05535, 2016.

[29] Cichocki A. Era of big data processing: A new approach via tensor networks and tensor decompositions. arXiv:1403.2048, 2014.

[30] Tipping ME. Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research*, 2001;1(Jun):211-244. DOI:10.1162/15324430152748236.

[31] Chen XY, He ZC, Sun LJ. A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transportation Research Part C-Emerging Technologies*, 2019;98:73-84. DOI:10.1016/j.trc.2018.11.003.

[32] Liu J, et al. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis machine intelligence*, 2012;35(1):208-220.

[33] Chen X, He Z, Wang J. Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition. *Transportation Research Part C: Emerging Technologies*, 2018;86:59-77.