



BACP: Bayesian Augmented CP Factorization for Traffic Data Imputation

Rongping Huang^{1,2}, Wenwu Gong¹, Jiaxin Lu¹, Zhejun Huang¹, and Lili Yang^{1,2(✉)}

¹ Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Shenzhen, China

yangl1@sustech.edu.cn

² Southern University of Science and Technology, Shenzhen 518055, China

Abstract. Traffic data possesses spatiotemporal characteristics and encounters missing value problems due to sensor failure in real-world scenarios. Addressing this challenge requires a fast and efficient traffic data imputation method capable of leveraging spatiotemporal information. This paper proposes a Bayesian Augmented CP factorization (BACP) model for the traffic data imputation, which combines the Multiplicative Gamma Process (MGP) with the CP factorization to address the CP rank estimation. Extensive experiment results demonstrate that the BACP model has superior imputation accuracy. Additionally, it offers explicit interpretation of traffic patterns and exhibits lower computational complexity than other Bayesian methods.

Keywords: Traffic Data Imputation · Variational Inference · Augmented CP Factorization · MGP Shrinkage · CP Rank

1 Introduction

With advancements in sensor technologies, the increasing development of Intelligent Transportation Systems (ITS) has led to more multidimensional traffic data collection. Traffic data encompasses temporal and spatial attributes since the data is acquired at specific times and locations [1, 2]. On the one hand, real-world traffic data may suffer from missing data problems due to sensor failures, weather conditions, and occlusions caused by environmental obstacles [3]. The missing traffic data imputation is one of the most critical research questions in traffic data analysis [4] since accurate and reliable imputation can help various applications in ITS, such as traffic forecasting, traffic control, and traffic management. On the other hand, the tensor-based method demonstrates the capacity to capture the spatiotemporal characteristics of traffic data, proving the data in a third-order tensor form [5]. The significance of handling missing traffic data based on low-rank tensor decomposition (LRTD) has become increasingly prominent in traffic data imputation [6].

The current LRTD models mainly have two typical aspects: Tucker decomposition [7] and CP (CANDECOMP/PARAFAC) factorization [8]. Ran et al. [9] proposed a low multilinear rank tensor decomposition model to explore spatiotemporal information and

estimate the missing traffic volumes. Goulart et al. [10] demonstrated that Tucker-based imputation methods are effective when processing missing speed data. Chen et al. [11] introduced a framework consisting of three processes that utilize Tucker decomposition to predict missing values by considering the spatiotemporal features of traffic data. Gong et al. [12] proposed a novel spatiotemporal regularized Tucker decomposition model for addressing the traffic data imputation, which adopts a perspective emphasizing tensor sparsity without requiring a predefined Tucker rank. However, those methods based on Tucker decomposition involve large matrix operations, which bring high time consumption and cause inefficiency in the solution algorithm.

Numerous efficient and interpretability models have emerged in the context of CP-based traffic data imputation methods, including those that incorporate spatiotemporal characteristics. Xu et al. [13] devised an interpretable and adaptable Spatiotemporal Bayesian Multivariate Adaptive-Regression Splines (ST-BMARS) model for traffic data analysis. In addition, Chen et al. [14] proposed a Bayesian Gaussian CP decomposition model to solve the problem of temporal dependencies in traffic data by using the Bayesian probability framework. However, the current CP-based models necessitate the manual specification of the tensor rank, which presents a challenge in determining the appropriate CP rank.

It is crucial to highlight the limitations of CP rank determination, which has been proven NP-hard. Previous studies have divided the CP rank estimation into rank optimization [15] and Bayesian inference [16, 17]. However, it often takes much time to achieve this goal. Zhao et al. [18] developed a deterministic Bayesian inference model, Automatic Rank Determination (ARD), which solves the CP rank estimation but scales linearly with tensor size. Takayama et al. [19] applied the Multiplicative Gamma Process (MGP) Shrinkage Prior to ARD, achieving better rank accuracy. It has been observed that the performance of the Gaussian-Gamma models is not satisfactory when dealing with tensors of high rank and/or low signal-to-noise ratios. To address these limitations, Cheng et al. [20] proposed the integration of a Generalized Hyperbolic (GH) prior into the probabilistic CP model. Zhang et al. [21] investigated general-purpose Bayesian tensor learning, which determines the CP rank automatically and can quantify the uncertainty of results. Those enhanced prior exhibit flexible model building and superiority performance varying in different missing scenarios.

In summary, many off-the-shelf CP-based models struggle with rank determination and modeling capacity. Furthermore, a fast and efficient traffic data imputation method that can utilize spatiotemporal information remains. In this paper, we apply the MGP shrinkage prior to the proposed augmented CP factorization model under the variational inference framework, which solves the problem of rank estimation and interprets the traffic patterns.

2 Preliminaries

The following section reviews the concepts related to CP factorization. Table 1 shows all the notations used in this paper.

Table 1. Notations.

Symbol	Description
$\mathbf{a}, \mathbf{A}, \mathcal{A}$	A vector, matrix, and tensor, respectively
$\text{rank}(\mathcal{A})$	The minimum number of rank-1 tensors
$\langle \mathcal{A}, \mathcal{B} \rangle$	The inner product of tensors \mathcal{A} and \mathcal{B}
$\ \mathcal{A}\ _F^2$	The squared Frobenius norm of tensor \mathcal{A}
$\mathbf{A} \circledast \mathbf{B}$	The Hadamard product of matrices \mathbf{A} and \mathbf{B}
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product of matrices \mathbf{A} and \mathbf{B}
$\mathbf{A} \odot \mathbf{B}$	The Khatri-Rao product of matrices \mathbf{A} and \mathbf{B}
$\mathbf{A}^{(n)}$	$\odot_{k \neq n} \mathbf{A}^{(k)} = \mathbf{A}^{(N)} \odot \dots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \dots \odot \mathbf{A}^{(1)}$

3 Proposed Model: BACP

This paper proposes a novel CP factorization model, Bayesian augmented CP factorization (BACP), for traffic data imputation and semantic traffic patterns interpretation. The proposed model combines the Multiplicative Gamma Process (MGP) shrinkage prior to the CP factorization to facilitate the automatic determination of the CP-rank.

3.1 Model Analysis

Assumed that \mathcal{Y} consists of independent Gaussian distribution:

$$y_{i_1 i_2 i_3} \sim \mathcal{N}(\mu + \phi_{i_1} + \theta_{i_2} + \eta_{i_3} + \sum_{r=1}^R \lambda_r a_{i_1 r} a_{i_2 r} a_{i_3 r}, \tau_c^{-1}), \forall (i_1, i_2, i_3), \quad (1)$$

$$\mu, \phi_{i_1}, \theta_{i_2}, \eta_{i_3} \sim \mathcal{N}(\mu_0, \tau_0^{-1}), \forall (i_1, i_2, i_3), \mathbf{a}_{i_n}^{(n)} \sim \mathcal{N}(0, \Lambda^{-1}), \forall n \in [1, 3],$$

$$\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_R], \boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda}), \tau_c \sim \text{Gamma}(a_0, b_0).$$

where $\mathcal{N}(\cdot)$ denotes Gaussian distribution, and τ_c is the precision parameter which holds a universal value applicable to each element within the tensor. Besides, $\mu \in \mathbb{R}$ denotes the pervasive impact on tensor elements, $\phi \in \mathbb{R}^{I_1}$, $\theta \in \mathbb{R}^{I_2}$, $\eta \in \mathbb{R}^{I_3}$ capture the information of dimensions, $\mathbf{A}^{(1)} \in \mathbb{R}^{I_1 \times R}$, $\mathbf{A}^{(2)} \in \mathbb{R}^{I_2 \times R}$, $\mathbf{A}^{(3)} \in \mathbb{R}^{I_3 \times R}$ are factor matrices governing interactions across distinct dimensions, and $\boldsymbol{\lambda} \in \mathbb{R}^{R \times R}$ corresponds to the weight of each rank-1 tensor.

The proposed BACP model is driven by the MGP prior in Bayesian framework, which is represented by Eq. (1). The MGP prior, as described in Eq. (2), applied to the precision of the Gaussian distribution for λ_r , shrinks the λ_r towards zero as r increases. Figure 1 provides a graphical representation of the BACP model.

$$\lambda_r \sim \text{Gamma}(c_0, \tau_r), 1 \leq r \leq R, \tau_r = \prod_{l=1}^r \delta_l (0 < \delta_r < 1), \delta_r \sim \text{Gamma}(e_0, f_0). \quad (2)$$

Remark. In the augmented CP factorization model, the parameter μ aims to approximate the average value of tensor elements. Conditional upon μ , biases are introduced along each dimension to encapsulate explicit tensor patterns or characteristics. For the traffic data imputation, it proves advantageous to interpret the traffic patterns that account for biases in spatial and temporal attributes.

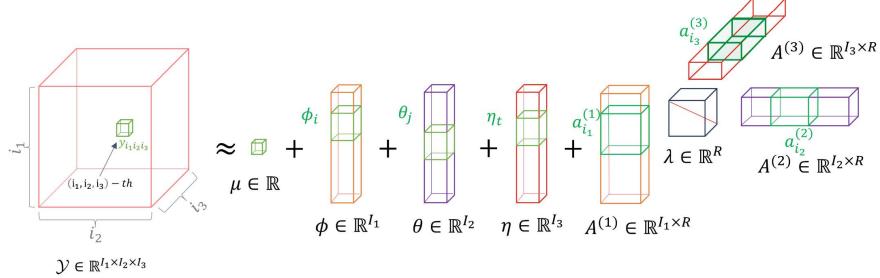


Fig. 1. Visualization of the proposed BACP model.

3.2 Variational Inference of BACP

Let $\Theta = \{\mu, \phi, \theta, \eta, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \boldsymbol{\lambda}, \delta, \tau_c\}$ for the convenience, this paper employs variational inference to learn parameters from the data tensor \mathcal{Y} , which involves identifying a distribution $q(\Theta)$ that acts as an approximation to the posterior distribution $p(\Theta|\mathcal{Y}_\Omega)$. We present the procedure of solving algorithm for the BACP model.

We first calculate the joint distribution by Eq. (3)

$$\begin{aligned} p(\mathcal{Y}_\Omega, \Theta) &= p(\mathcal{Y}_\Omega|\Theta) \times p(\mu) \times \prod_{i_1=1}^{I_1} p(\phi_{i_1}) \prod_{i_2=1}^{I_2} p(\theta_{i_2}) \prod_{i_3=1}^{I_3} p(\eta_{i_3}) \prod_{n=1}^3 p(\mathbf{A}^{(n)}|\boldsymbol{\lambda}) \\ &\quad \times \prod_{r=1}^R p(\lambda_r|\tau_r) p(\delta_r) \times p(\tau_c). \end{aligned} \quad (3)$$

Then we derive $q(\Theta)$ in a manner that minimizes the KL divergence

$$\text{argmin}_q \text{KL}(q(\Theta)||p(\Theta|\mathcal{Y}_\Omega)) = \text{argmax}_q \mathcal{L}(q), \quad (4)$$

where $\mathcal{L}(q) = \int q(\Theta) \ln \{p(\mathcal{Y}_\Omega, \Theta)/q(\Theta)\} d\Theta$ denotes the lower bound. Thus, the mean-field approximation for $q(\Theta)$ is denoted as

$$q(\Theta) = q(\mu) \times \prod_{i_1=1}^{I_1} q(\phi_{i_1}) \prod_{i_2=1}^{I_2} q(\theta_{i_2}) \prod_{i_3=1}^{I_3} q(\eta_{i_3}) \prod_{n=1}^3 q(\mathbf{A}^{(n)}|\boldsymbol{\lambda}) \times \prod_{r=1}^R q(\lambda_r|\tau_r) q(\delta_r) \times q(\tau_c). \quad (5)$$

The j -th factor optimization is obtained through maximizing $\mathcal{L}(q)$

$$\ln q(\Theta_j) = \mathbb{E}_{q(\Theta \setminus \Theta_j)} [\ln p(\mathcal{Y}_\Omega, \Theta)] + \text{const}, \quad (6)$$

where the $\mathbb{E}_{q(\Theta \setminus \Theta_j)}[\cdot]$ represents the expectation of overall variables except Θ_j .

Posterior Distribution of Global Parameter μ

We can get the posterior of μ as Gaussian distribution $q(\mu) = \mathcal{N}(\tilde{\mu}_0, \tilde{\tau}_0^{-1})$ with

$$\tilde{\mu}_0 = \tilde{\tau}_0^{-1} (\mathbb{E}[\tau_c] \sum_{(i_1, i_2, i_3) \in \Omega} \mathbb{E}[z_{i_1 i_2 i_3}] + \tau_0 \mu_0), \quad \tilde{\tau}_0 = \mathbb{E}[\tau_c] \sum_{(i_1, i_2, i_3) \in \Omega} \mathcal{O}_{i_1 i_2 i_3} + \tau_0, \quad (7)$$

where $z_{i_1 i_2 i_3} = y_{i_1 i_2 i_3} - \phi_{i_1} - \theta_{i_2} - \eta_{i_3} - \sum_{r=1}^R a_{i_1 r} a_{i_2 r} a_{i_3 r}$.

Posterior Distribution of Bias Vectors $\{\phi, \theta, \eta\}$

The variational posterior of ϕ_{i_1} is Gaussian distribution $q(\phi_{i_1}) = N(\tilde{\mu}_\phi, \tilde{\tau}_\phi^{-1})$ with

$$\tilde{\mu}_\phi = \tilde{\tau}_\phi^{-1} (\mathbb{E}[\tau_c] \sum_{(i_1, i_2, i_3) \in \Omega} \mathbb{E}[f_{i_1 i_2 i_3}] + \tau_0 \mu_0), \quad \tilde{\tau}_\phi = \mathbb{E}[\tau_c] \sum_{(i_2, i_3) \in \Omega} o_{i_1 i_2 i_3} + \tau_0, \quad (8)$$

where $f_{i_1 i_2 i_3} = y_{i_1 i_2 i_3} - \mu - \theta_{i_2} - \eta_{i_3} - \sum_{r=1}^R a_{i_1 r} a_{i_2 r} a_{i_3 r}$. Upon obtaining the variational posterior distribution $q(\phi_{i_1})$, the variational posterior distributions $q(\theta_{i_2})$ and $q(\eta_{i_3})$ can be derived through the similar procedures.

Posterior Distribution of Factor Matrices $[\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]$

The posterior distributions of the factor matrices have been demonstrated to be decomposable into independent distributions of their rows, which exhibit Gaussian characteristics. The posterior distribution of $[\mathbf{A}^{(n)}]$ is given by

$$q(\mathbf{A}^{(n)}) = \prod_{i_n=1}^{I_n} \mathcal{N}(\mathbf{a}_{i_n}^{(n)} | \tilde{\mathbf{a}}_{i_n}^{(n)}, \mathbf{V}_{i_n}^{(n)}), \quad \forall n \in [1, 3], \quad (9)$$

and we update the posterior parameters through

$$\tilde{\mathbf{a}}_{i_n}^{(n)} = \mathbb{E}[\tau_c] \mathbf{V}_{i_n}^{(n)} \mathbb{E}[\mathbf{A}_{i_n}^{(n)T}] \text{vec}(\mathcal{W}_{\mathbb{I}(\mathcal{O}_{i_n}=1)}), \quad \mathbf{V}_{i_n}^{(n)} = (\mathbb{E}[\tau_c] \mathbb{E}[\mathbf{A}_{i_n}^{(n)T} \mathbf{A}_{i_n}^{(n)}] + \mathbb{E}[\Lambda])^{-1}, \quad (10)$$

where the operation $\text{vec}(\cdot)$ denotes the process of vectorization performed column-wise, $w_{i_1 i_2 i_3} = y_{i_1 i_2 i_3} - \mu - \phi_{i_1} - \theta_{i_2} - \eta_{i_3}$ and $\mathbb{I}(\cdot)$ denotes the indicator function.

Posterior Distribution of MGP Prior

Considering the weight parameter λ , combining with MGP prior, we have

$$q(\lambda_r) = \text{Ga}(\lambda_r | c_M^r, d_M^r), \quad (11)$$

and its parameters are

$$c_M^r = c_0 + \frac{1}{2} \sum_{n=1}^3 I_n, \quad d_M^r = \mathbb{E}[\tau_r] + \frac{1}{2} \sum_{n=1}^3 \mathbb{E}[\mathbf{a}_{.,r}^{(n)T} \mathbf{a}_{.,r}^{(n)}]. \quad (12)$$

The posterior distribution of δ and its parameters are as follows.

$$q(\delta_r) = \text{Ga}(\delta_r | e_M^r, f_M^r), \quad (13)$$

$$e_M^r = (R - r + 1)c_0 + e_0, f_M^r = \sum_{h=r}^R \mathbb{E}_q[\lambda_h] \prod_{l=1, l \neq r}^h \mathbb{E}_q[\delta_l] + f_0. \quad (14)$$

The precision parameter $\tau \in \mathbb{R}$, where the variational posterior is derived as Gamma distribution $q(\tau_c) = \text{Ga}(\tau_c | a_M, b_M)$, and its parameters are as follows.

$$\begin{aligned} a_M &= a_0 + \frac{1}{2} \sum_{i_1, i_2, i_3} \mathcal{O}_{i_1, i_2, i_3}, b_M \\ &= b_0 + \frac{1}{2} \mathbb{E}[\mathcal{O} \circledast (\mathcal{Y} - \mathcal{G})_F^2] = b_0 + \sum_{(i_1, i_2, i_3) \in \Omega} \mathbb{E}[(y_{i_1 i_2 i_3} - g_{i_1 i_2 i_3})^2]. \end{aligned} \quad (15)$$

3.3 Algorithm Analysis

Convergence

The assessment of algorithmic convergence is facilitated by monitoring the lower bound value. $\mathcal{L}(q)$ exhibit a monotonic increase at each epoch, serving as an indicator of algorithmic convergence. We can get the lower bound as follows:

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_q[\ln p(\mathcal{Y}_\Omega | \Theta)] + \mathbb{E}_q[\ln p(\mu)] + \mathbb{E}_q[\ln p(\phi)] + \mathbb{E}_q[\ln p(\theta)] + \mathbb{E}_q[\ln p(\eta)] \\ &\quad + \sum_{n=1}^3 \mathbb{E}_q[\ln p(\mathbf{A}^{(n)} | \lambda)] + \mathbb{E}_q[\ln p(\tau_c)] + \mathbb{E}_q[\ln p(\delta)] + \mathbb{E}_q[\ln p(\lambda)] \\ &\quad - \mathbb{E}_q[\ln q(\mu)] - \mathbb{E}_q[\ln q(\phi)] - \mathbb{E}_q[\ln q(\theta)] - \mathbb{E}_q[\ln q(\eta)] \\ &\quad - \sum_{n=1}^3 \mathbb{E}_q[\ln q(\mathbf{A}^{(n)} | \lambda)] - \mathbb{E}_q[\tau_c] - \mathbb{E}_q[\delta] - \mathbb{E}_q[\lambda]. \end{aligned} \quad (16)$$

The algorithm converges when the condition holds:

$$\left| \frac{\mathcal{L}(q)^{(t)} - \mathcal{L}(q)^{(t-1)}}{\mathcal{L}(q)^{(2)}} \right| < \epsilon, \quad (17)$$

where ϵ is the adjustable parameter that acts as the convergence threshold.

Algorithm 1: Bayesian augmented CP factorization using MGP shrinkage prior(MGP-BACP)

```

input : Observation data tensor  $\mathcal{Y}^{I_1 \times I_2 \times I_3}$ , indicator tensor  $\mathcal{O}^{I_1 \times I_2 \times I_3}$ ,  $\mu, \phi, \theta, \eta$  and
        factor matrices  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}$ .
output: Estimated tensor  $\hat{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , and updated  $\mu, \phi, \theta, \eta$ , and  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}$ .
1 Initialize:  $\mathbf{V}_{i_n}^{(n)}, \forall i_n \in [1, I_n], \forall n \in [1, N], a_0, b_0, c_0, d_0, e_0, f_0, \mu_0 = 0, \tau_0 = 1, \tau_c = \frac{a_0}{b_0}, \lambda_r =$ 
 $\frac{c_0}{\tau_r}, \delta_r = \frac{e_0}{f_0}, \forall r \in [1, R]$ .
2 repeat
3     Update the posterior distribution  $q(\mu)$  using Eq.(7).
4     for  $i_1 = 1$  to  $I_1$  do
5         Update the posterior distribution  $q(\phi_{i_1})$  using Eq.(8).
6         Update the posterior distribution  $q(\mathbf{a}_{i_1}^{(1)})$  using Eq.(10).
7     end
8     for  $i_2 = 1$  to  $I_2$  do
9         Update the posterior distribution  $q(\theta_{i_2})$  using Eq.(8).
10         Update the posterior distribution  $q(\mathbf{a}_{i_2}^{(2)})$  using Eq.(10).
11     end
12     for  $i_3 = 1$  to  $I_3$  do
13         Update the posterior distribution  $q(\eta_{i_3})$  using Eq.(8).
14         Update the posterior distribution  $q(\mathbf{a}_{i_3}^{(3)})$  using Eq.(10).
15     end
16     Update the posterior distribution  $q(\boldsymbol{\delta})$  using Eq.(14).
17     Update the posterior distribution  $q(\boldsymbol{\lambda})$  using Eq.(12).
18     Update the posterior distribution  $q(\tau_c)$  using Eq.(15).
19     Calculate the variational lower bound  $\mathcal{L}(q)$  using Eq.(16).
20     Reduce the rank R by removing the 0 component of  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}$ .
21 until  $\mathcal{L}(q)$  converges by checking Eq.(17);
```

Computation Complexity

The factor matrices are the most computationally intensive part of our algorithm with Eq. (10) which cost $O(NR^2M + R^3 \sum_n I_n)$, where N denotes the order of tensor, $\sum_n I_n$ is the sum of values of each dimension, R is tensor rank and M is the number of observations. The computation cost of global parameter λ is $O(R^2 \sum_n I_n)$. δ costs $O(R^2)$, τ costs $O(R^2M)$, and μ, ϕ, θ, η cost less than $O(\sum_n I_n)$. Therefore, the complexity of BACP is $O(NR^2M + R^3)$. When employing MGP priors, our rank estimation tends to be conservative [19], enabling faster computation with smaller rank.

4 Experiments

In this section, we conduct experiments on three open urban traffic datasets to compare the BACP model with baselines in different missing scenarios and demonstrate the proposed model's interpretability on one data set.

4.1 Settings

Dataset. We use the following three open urban traffic datasets for our experiment and form them as 3rd-order tensors for traffic data imputation problems.

- **(G):** It was collected over two months, specifically from August 1, 2016, to September 30, 2016. The speed data was recorded every 10 min. It was expressed as a third-order tensor with size $216 \times 61 \times 144$ (road segment \times day \times time interval).
- **(S):** It contains highway speeds for 323 loop detectors in Seattle, USA, throughout 2015. A subset (from February 1 to February 28) is selected and structured as a third-order tensor with size $323 \times 28 \times 288$ (road segment \times day \times time interval).
- **(PeMS):** The traffic dataset in California, USA, for 358 data sets collected in the Gulf of California for 5 min for three months (September 1 to November 30, 2018). It contains a subset of 24 detectors for one month and forms a third-order tensor with size $24 \times 30 \times 288$ (road segment \times day \times time interval).

Performance Metrics. For evaluating the imputation performance, the Mean absolute percentage error (MAPE) and root mean square error (RMSE) are adapted,

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \times 100, \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (18)$$

where N represents the overall count of missing values, y_i and \hat{y}_i denote the real value of a missing element and its corresponding imputation result.

Experimental Settings. Initially, we set different missing rates in the two scenarios while keeping the rank fixed. Subsequently, we compare the BATF model to demonstrate our automated rank determination. Finally, we showcase the interpretability of our model using datasets. The maximum epoch is set to 200; the max rank is set to 100, and the convergence threshold ϵ is set to 1e-4.

Baselines. For comparison, we select three Bayesian CP factorization models: the Bayesian augmented tensor factorization (BATF, Chen et al. [26]), the Bayesian CP factorization (BCPF, Zhao et al. [18]), and the Bayesian CP factorization with ARD using MGP shrinkage prior (MGP-ARD, Takayama et al. [19]). The comparison of each model is shown in Table 2.

Table 2. The difference between baselines and BACP.

	BCPF	MGP-ARD	BATF	BACP
Interpretability			✓	✓
Automatic CP Rank	✓	✓		✓

4.2 Results

Performances with Given Rank

We first compare the proposed BACP model with BCPF and MGP-ARD methods with the given CP rank. The proposed BACP model outperforms the other two baselines in

Table 3. (MAPE/RMSE) of two missing scenarios on three datasets with given rank.

Dataset	Missing Scenario	BCPF	MGP-ARD	BACP
(G)	EM40%(r = 30)	8.50/3.66	8.52/3.67	8.43/3.63
	EM50%(r = 30)	8.54/3.68	8.56/3.69	8.46/3.64
	FM40%(r = 30)	10.58/4.71	10.72/4.76	10.13/4.32
	FM50%(r = 30)	10.70/4.75	10.85/4.84	10.36/4.40
(S)	EM60%(r = 30)	6.86/4.92	6.88/4.96	6.80/4.85
	EM70%(r = 30)	6.95/5.08	6.96/5.09	6.88/5.02
	FM60%(r = 30)	8.49/5.68	8.68/5.91	8.40/5.53
	FM70%(r = 30)	9.18/7.16	9.50/7.36	9.11/7.11
(PeMs)	EM60%(r = 30)	7.80/23.99	7.85/24.04	7.79/23.97
	EM70%(r = 30)	7.92/24.63	7.97/24.35	7.90/24.40
	FM40%(r = 30)	12.50/37.00	12.62/36.98	12.40/36.43
	FM50%(r = 30)	14.02/38.85	13.06/38.04	12.43/37.54

terms of accuracy, as evidenced by diverse missing rates and missing scenarios. Table 3 demonstrates the performances, with the best results highlighted in bold fonts.

Performances of CP Rank Estimation

Table 4 shows the overall imputation performance of BATF and BACP on the three datasets under different missing scenarios. We select the rank value commonly assigned in the refer paper for the BATF rank. For BACP, we set the initial max rank to 100 and then use MGP prior to make it automatically shrink the rank. In contrast to the semantic model BATF, the proposed BACP model exhibits superior performance under 10% missing rates, with a marginal loss of accuracy within 1%, yet achieving a significantly faster runtime with 58.7% enhancement. The results of the BACP model also demonstrate that is non sensitivity to different missing rates.

Semantic Interpretations

We present the semantic representation of the proposed BACP model on dataset S. The global parameters in Fig. 2 represent the average speed curve, which is close to the average speed of the real data. Figure 3 shows the bias of the time intervals and road segment dimensions from the global parameters, where a positive bias value indicates good road conditions, and a negative one indicates poor road conditions. These parameters represent the explicit patterns of the model and have good practical significance.

In Fig. 3, the first graph shows the biases between the global average speed and the average speed of each road so we can find which road segment is in good condition. We sort the roads according to the number and select road number #210 for analysis from the time dimension in the second graph. We can find two days a week in bad condition, from a practical point of view, which is the local weekend, caused by more vehicles, as true as the facts. In addition, on a specific day, night bias is often positive, and day bias

Table 4. (MAPE/RMSE) of manually selected rank and automatically determined rank.

Dataset	Missing Scenario	BATF	BACP
(G)	EM10%	8.25/3.57 (r = 80)	8.34/3.64 (r = 64)
	EM30%	8.41/3.63 (r = 80)	8.59/3.74 (r = 55)
	EM50%	10.20/4.19 (r = 80)	10.42/4.40 (r = 49)
	FM10%	9.76/4.13 (r = 20)	9.64/4.08 (r = 28)
	FM30%	9.95/4.23 (r = 15)	9.98/4.25 (r = 13)
	FM50%	10.29/4.36 (r = 10)	10.33/4.40 (r = 7)
	Average Runtime(s)	2137	883
(S)	EM10%	5.84/4.11 (r = 30)	5.88/4.15 (r = 34)
	EM30%	6.15/4.38 (r = 30)	6.24/4.46 (r = 22)
	EM50%	6.48/4.62 (r = 30)	6.60/4.67 (r = 19)
	FM10%	7.26/4.94 (r = 30)	6.74/4.64 (r = 24)
	FM30%	7.34/5.03 (r = 30)	7.28/4.95 (r = 20)
	FM50%	7.69/5.21 (r = 30)	7.24/4.50 (r = 18)
	Average Runtime(s)	3293	1482
(PeMs)	EM10%	6.98/23.05 (r = 30)	6.98/23.04 (r = 27)
	EM30%	7.49/23.19 (r = 30)	7.61/23.37 (r = 21)
	EM50%	7.69/23.69 (r = 30)	7.78/23.87 (r = 19)
	FM10%	8.57/30.72 (r = 30)	8.61/31.01 (r = 23)
	FM30%	9.50/34.03 (r = 30)	9.78/34.96 (r = 18)
	FM50%	12.19/37.12 (r = 30)	12.53/37.79 (r = 15)
	Average Runtime(s)	524	240

is often negative. Therefore, our model has realistic interpretability for traffic data, and we conclude those as follows.

- **Global explainability:** the average value of the overall velocity.
- **Interpretability in time dimension:** judge day/night or working day/weekend by the positive and negative values of bias.
- **Interpretability in spatial dimension:** judge the road conditions of specific roads by the positive and negative values of bias.

Imputation Example

Here, we also show the imputing example of the BACP model on dataset G with a 50% fiber missing scenario in Fig. 4. It indicates that BACP imputes missing values more accurately.

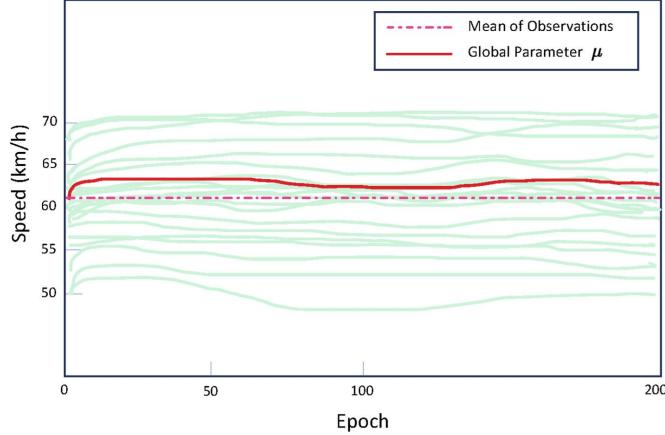


Fig. 2. Global parameter at each epoch.

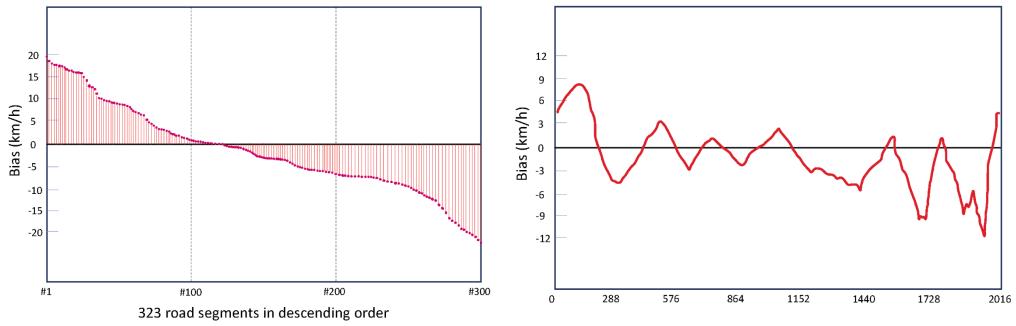


Fig. 3. Biases of 323 road segments and time intervals of a week.

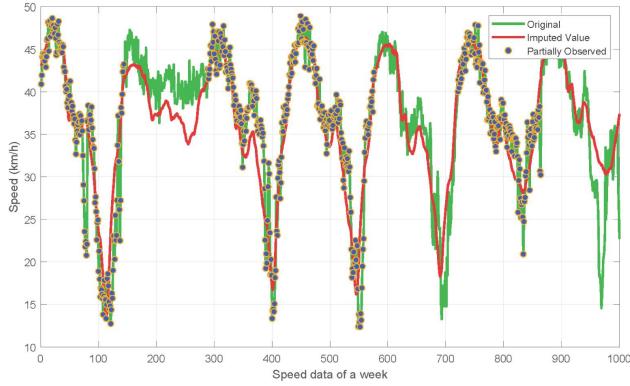


Fig. 4. Imputation results of traffic speed data (km/h) with 50% fiber missing scenario.

5 Conclusions and Future Work

The missing traffic data imputation is one of the most critical research questions in traffic data analysis. This paper proposes an interpretable CP factorization model named Bayesian Augmented CP factorization (BACP) for traffic data imputation. BACP addresses the CP rank determination using the Multiplicative Gamma Process (MGP) prior and derives closed-form update rules within the variational inference framework.

Besides, the BACP model consists of latent factors, which exhibit excellent interpretability of traffic data patterns. Numeric results under various missing scenarios indicate that BACP outperforms other rank-self-adjusting baselines and is faster than the BATF model. The future work includes extending it to other tensor completion tasks, such as image inpainting, and employing different priors based on dataset characteristics to achieve superior results.

Acknowledgments. The authors express their gratitude to the anonymous referees and editors for their valuable insights, which significantly contributed to the enhancement of this paper. Additionally, heartfelt thanks are extended to the authors who shared their codes on websites. This research is supported by the Shenzhen Science and Technology Program (Grant No. ZDSYS20210623092007023), the Educational Commission of Guangdong Province (Grant No. 2021ZDZX1069), and Guangdong Province Universities and Colleges Key Areas of Special Projects (Grant No. 2021222012).

References

1. Asif, M. T., Kannan, S., Dauwels, J., Jaillet, P.: Data compression techniques for urban traffic data. In: 2013 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS), pp. 44–49. IEEE, Singapore (2013)
2. Li, L., Li, Y., Li, Z.: Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. *Transp. Res. Part C: Emerg. Technol.* **34**, 108–120 (2013)
3. Chen, X., Lei, M., Saunier, N., Sun, L.: Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Trans. Intell. Transp. Syst.* **23**(8), 12301–12310 (2021)
4. Ran, B., Tan, H., Wu, Y., Jin, P.J.: Tensor based missing traffic data completion with spatial-temporal correlation. *Phys. A* **446**, 54–63 (2016)
5. Silva-Ramírez, E.L., Pino-Mejías, R., López-Coello, M., Cubiles-de-la-Vega, M.D.: Missing value imputation on missing completely at random data using multilayer perceptrons. *Neural Netw.* **24**(1), 121–129 (2011)
6. Gong, W., Huang, Z., Yang, L.: LSPTD: low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), pp. 460–465. IEEE, Bilbao, Spain (2023)
7. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**(3), 279–311 (1966)
8. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika* **35**(3), 283–319 (1970)
9. Ran, B., Tan, H., Feng, J., Wang, W., Cheng, Y., Jin, P.: Estimating missing traffic volume using low multilinear rank tensor completion. *Journal of Intelligent Transportation Systems* **20**(2), 152–161 (2016)
10. Goulart, J.D.M., Kibangou, A.Y., Favier, G.: Traffic data imputation via tensor completion based on soft thresholding of Tucker core. *Transp. Res. Part C: Emerg. Technol.* **85**, 348–362 (2017)
11. Chen, X., He, Z., Wang, J.: Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition. *Transp. Res. Part C: Emerg. Technol.* **86**, 59–77 (2018)

12. Gong, W., Huang, Z., Yang, L.: Spatiotemporal regularized tucker decomposition approach for traffic data imputation. arXiv preprint [arXiv:2305.06563](https://arxiv.org/abs/2305.06563) (2023)
13. Xu, Y., Kong, Q.J., Klette, R., Liu, Y.: Accurate and interpretable bayesian mars for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **15**(6), 2457–2469 (2014)
14. Chen, X., Sun, L.: Bayesian temporal factorization for multidimensional time series prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(9), 4659–4673 (2021)
15. Acar, E., Dunlavy, D.M., Kolda, T.G., Mørup, M.: Scalable tensor factorizations for incomplete data. *Chemom. Intell. Lab. Syst.* **106**(1), 41–56 (2011)
16. Sheng, G.A.O., Denoyer, L., Gallinari, P., Jun, G.U.O.: Probabilistic latent tensor factorization model for link pattern prediction in multi-relational networks. *J. China Univ. Posts Telecommun.* **19**, 172–181 (2012)
17. Xiong, L., Chen, X., Huang, T. K., Schneider, J., Carbonell, J. G.: Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In: Proceedings of the 2010 SIAM International Conference on Data Mining (SDM), pp. 211–222. Society for Industrial and Applied Mathematics, Columbus, USA (2010)
18. Zhao, Q., Zhang, L., Cichocki, A.: Bayesian CP factorization of incomplete tensors with automatic rank determination. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(9), 1751–1763 (2015)
19. Takayama, H., Zhao, Q., Hontani, H., Yokota, T.: Bayesian tensor completion and decomposition with automatic CP rank determination using MGP shrinkage prior. *SN Comput. Sci.* **3**(3), 225 (2022)
20. Cheng, L., Chen, Z., Shi, Q., Wu, Y.C., Theodoridis, S.: Towards flexible sparsity-aware modeling: automatic tensor rank learning using the generalized hyperbolic prior. *IEEE Trans. Signal Process.* **70**, 1834–1849 (2022)
21. Zhang, K., Hawkins, C., Zhang, Z.: General-purpose Bayesian tensor learning with automatic rank determination and uncertainty quantification. *Front. Artific. Intell.* **4**, 668353 (2022)
22. Chen, X., He, Z., Chen, Y., Lu, Y., Wang, J.: Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model. *Transp. Res. Part C: Emerg. Technol.* **104**, 66–77 (2019)