Mechanical Systems and Signal Processing 158 (2021) 107734

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Incremental Bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting

Pu Ren [a], Xinyu Chen [b], Lijun Sun [b], Hao Sun [a,c,*]

[a] Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA
[b] Department of Civil Engineering, McGill University, Montreal, Quebec H3A0G4, Canada
[c] Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 16 July 2020
Received in revised form 22 January 2021
Accepted 8 February 2021

Keywords:
Tensor decomposition
Bayesian inference
Incremental learning
Data imputation
Response forecasting
Structural health monitoring

ABSTRACT

There has been increased interest in missing sensor data imputation, which is ubiquitous in the field of structural health monitoring (SHM) due to discontinuous sensing caused by sensor malfunction. Recent development in Bayesian temporal factorization models for high-dimensional time series analysis has provided an effective tool to solve both imputation and prediction problems. However, for large datasets, the default Bayesian temporal factorization model becomes inefficient since the model has to be fully retrained when new data arrives. A potential solution is to train the model using a short time window covering only most recent data; however, by doing so, we may miss some critical dynamics and long-term dependencies which can only be identified from a longer time window. To address this fundamental issue in temporal factorization models, this paper presents an incremental Bayesian matrix/tensor learning scheme to achieve efficient imputation and prediction of structural response in long-term SHM. In particular, a spatiotemporal tensor is first constructed followed by Bayesian tensor factorization that extracts latent features for missing data imputation. To enable structural response forecasting based on long-term and incomplete sensing data, we develop an incremental learning scheme to effectively update the Bayesian temporal factorization model. The performance of the proposed approach is validated on continuous field-sensing data (including strain and temperature records) of a concrete bridge, based on the assumption that strain time histories are highly correlated to temperature recordings. The results indicate that the proposed probabilistic tensor learning framework is accurate and robust even in the presence of large rates of random missing, structured missing and their combination. The effect of rank selection on the imputation and prediction performance is also investigated. The results show that a better estimation accuracy can be achieved with a higher rank for random missing whereas a lower rank for structured missing.

1. Introduction

High-quality data plays a pivotal role in structural health monitoring (SHM) for condition assessment, damage detection, and decision making. However, during long-term monitoring, it is inevitable for imperfect and corrupted sensor measurements, especially in a harsh and noisy environment, which calls for effective approaches for imputation/recovery missing

* Corresponding author at: Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA.
E-mail addresses: ren.pu@northeastern.edu (P. Ren), lijun.sun@mcgill.ca (L. Sun), h.sun@northeastern.edu (H. Sun).

https://doi.org/10.1016/j.ymssp.2021.107734
0888-3270/

and noisy data. Furthermore, in order to conduct real-time early-warning of structural deterioration or even disastrous failure, forecasting/prediction of structural response has also received considerable attention. The general idea of time series analysis, in the context of imputation and forecasting, is to find key dynamic patterns from observations and establish a mapping function between the historical records and the estimation. Nevertheless, these tasks are rather challenging on account of complex spatiotemporal dependencies and inherent difficulty in large-scale and nonlinear characteristics of SHM data, especially in practical applications.

There have been a number of attempts made to solve the data imputation and forecasting problems in the SHM community. On one hand, in the missing data recovery research, compressive sensing is one common and typical approach to rebuild the entire temporal signals based on the sparsity assumption of the data in certain feature spaces [1–5]. Another interesting stream for data imputation is the use of probability methods (e.g., Gaussian process (GP) thanks to its great interpretation capacity for nonlinear dynamic processes), which has been comprehensively studied in outlier detection [6,7], model calibration/updating [8–16] and system identification [17–22]. For instance, Wan et al. [23] employed Bayesian multi-task learning with multi-dimensional GP priors to recover SHM data. Chen et al. [24] explored the possibility of probability density function estimation for data loss compensation with warping transformations. Some recent surveys have reported the great potential in data imputation by considering both the sensor information and time series, which is usually conceptualized as spatiotemporal. Yang et al. [25] developed a low-rank matrix completion method with ℓ_1-norm and a nuclear norm for imputation of random missing data. This approach is powerful but has limitations due to an ideal assumption that the data is randomly missing, which is less common in practical SHM (e.g., data might be missing for a continuous duration). Chen et al. [26] investigated the inter-sensor relationship of stochastic structural responses with non-parametric copulas, which flexibly captured the spatial dependency for strain data. Moreover, the sequential broad learning (SBL) approach was recently presented for efficiently reconstructing structural response [27], which is however short for spatial consideration. On the other hand, for the sake of data-driven structural response forecasting, the majority of existing research focus on the time-dependent response approximation based on high-quality collected data (e.g., data missing is not considered). In particular, the widely-accepted and well-studied methods are based on the linear combination of previous observations, for example, dynamic linear models [28–30] and autoregressive (AR) models [31–33]. Distinctively, Wan and Ni [34] examined the capability of a GP-based Bayesian approach for underlying nonlinear dynamic system response prediction from a statistic perspective. Besides, deep learning techniques, such as the convolutional neural network (CNN) [35–37], the long-short term memory (LSTM) network [38,39], and the variational autoencoder (VAE) [40], have also been proven to be a decent alternative for extracting spatial features for dynamic response reconstruction and prediction.

Despite the rapid development of data science in SHM, there still remain three representative challenges for the specific aim of data imputation and response forecasting. Firstly, very little work has been devoted to the spatial dependency and correlation in the time series analysis. The second is the lack of consideration on vast and continuous missing scenarios (e.g., data missing for a long continuous period such as one day or consecutive days). Lastly, almost all of the present studies on response forecasting are based on high-quality data instead of imperfect measurements with missing values. To this end, in light of the recent renaissance in tensor learning [41–43], which has already greatly contributed to image processing [44–49], recommender systems [50–52], and traffic data analysis [53–63]. In the context of SHM, we can naturally consider the data as multivariate time-series matrix and then apply temporal factorization models (e.g., [61]) where the low-rank representation can effectively characterize the complex spatial and temporal dependencies rooted in the data.However, in order to keep the model up-to-date, retraining is typically needed given streaming data: the same model deeds to be retrained based on the complete recorded data whenever new data arrives to the system, resulting in computational inefficiency. The growing data size demands extensive computation especially for large-scale long-term datasets. This poses a critical challenge for SHM which requires efficient models to account for continuous monitoring. To address these issues, in this paper, we inherit the modeling idea for matrix/tensor factors based on Chen and Sun [61] and propose an incremental Bayesian learning scheme that enables imputation of SHM data and forecasting of structural response in a long-term horizon for temporal/continuous SHM, which is the limitation of [61]. In addition, this research sheds new light on integrating physics into the tensor model inspired by the strong correlation between strain data and temperature data [64–67,30], resulting in an interpretative low-rank data structure. Instead of training on the full data ever recorded, we propose an incremental updating scheme leveraging locally streaming data, resulting in more efficient and accurate imputation/prediction for long-term SHM data. In particular, we employ the proposed learning approach for (1) reconstruction of spatiotemporal missing data in SHM and (2) forecasting of structural response under the scenario of missing/incomplete data. It is worthy to mention that, different from [25], tensor factorization in the context of Bayesian inference [61] provides a principled selection mechanism for suitable likelihood models and allows for uncertainty quantification in parameter estimation and prediction [68].

The main contribution of this paper can be summarized as follows. Firstly, to the best of our knowledge, it is the first time to realize response forecasting with incomplete data in SHM applications, based on reliable latent features instead of directly using the corrupted data. Secondly, by constructing one-dimensional time series data into a second- or third-order tensor structure (e.g., sensor locations $\times$ time stamps), we can easily capture the spatiotemporal features of the data for accurate imputation and forecasting. Thirdly, the physical relationship between strain and temperature is introduced to optimize the tensor structure. Fourthly, we propose an incremental learning scheme to tackle practical continuous monitoring problems and speed up the tensor factorization process through efficient updating. We further validate the proposed approach on a concrete bridge with multi-year recordings of strain and temperature time histories.

The rest of the paper is organized as follows, in addition to this Introduction section. Section 2 begins by laying out the theoretical dimensions of this work, and is concerned with the proposed methodology. In Section 2.1, we describe the problem definition and general principle of data imputation and response forecasting under the data missing scenarios. In Sections 2.2, 2.3 and 2.4, we circumstantially present the Bayesian generation and inference procedure, as well as the autoregressive process for temporal feature modeling. In Section 2.5, we introduce the incremental Bayesian tensor learning architecture and two adaptive factor updating stages. In Section 2.6, the extension to Bayesian high-order tensor learning is discussed. Section 3 elaborates the experimental validation results of the proposed method, focusing on three key themes: imputation and forecasting performance with respect to different missing rates, uncertainty quantification and rank analysis. Section 4 concludes the current work and the outlook of future directions.

2. Methodology

In this section, we formulate the problem of SHM data imputation and response forecasting in the context of incremental Bayesian tensor learning, and present the spatiotemporal dependency modeling procedure via matrix factorization.

2.1. Problem description

The goal of continuous/steaming SHM data imputation and forecasting is to estimate the missing values and predict the future structural response given partially observed data collected from a sensor network. The multidimensional time series data, with missing values, can be represented by the second-order tensor $\mathbf{Y} \in \mathbb{R}^{M\times T}$, where M denotes the number of sensor locations and T is the number of time stamps for a certain continuously monitoring period, or the third-order tensor $\mathbf{Y} \in \mathbb{R}^{M_1\times M_2\times T}$, where M_1 and M_2 might represent the number of sensors in two different spatial directions. Let us take the second-order tensor (matrix) for illustration. The imputation process aims to firstly learn a factorized spatial feature $\mathbf{U}$ and a temporal feature $\mathbf{X}$ based on the observed data $\mathbf{Y}$, and then reconstruct the response with imputed values. Afterwards, given $\mathbf{y}_{:,t}$ signifying the multivariate data at time t, the course of response forecasting utilizes the well-trained spatial factor $\mathbf{U}$ and the updated temporal factor $\mathbf{X}^*$ to map L ($\geqslant 1$) historical sensing data to future $\mathcal{T}$ ($\geqslant 1$) structural responses, given by

$$\left[\mathbf{y}_{:,t-L+1},\cdots,\mathbf{y}_{:,t}\right] \xrightarrow[\mathbf{U}]{\mathbf{X}^*} \left[\mathbf{y}_{:,t+1},\cdots,\mathbf{y}_{:,t+\mathcal{T}}\right] \tag{1}$$

which essentially establishes a temporal forecasting process.

2.2. Hierarchical Bayesian modeling for matrix decomposition

Naturally, spatiotemporal SHM data observed from M sensor locations with T time stamps can be constructed in the form of a two-dimensional tensor, $\mathbf{Y} \in \mathbb{R}^{M\times T}$. Due to inevitable data missing in practical applications, we define an indicator set for the observed elements in $\mathbf{Y}$ as $\Omega = \{(i,t)|y_{i,t}\text{isobserved}\}$. To characterize the spatiotemporal dependencies, we employ the general idea of second-order tensor (matrix) decomposition to approximate the multidimensional data through two factor matrices, namely,

$$\mathbf{Y} \approx \mathbf{U}^\top\mathbf{X}, \tag{2}$$

where $\mathbf{U} \in \mathbb{R}^{K\times M}$ is the spatial latent factor whose columns are $\mathbf{u}_i$'s, and $\mathbf{X} \in \mathbb{R}^{K\times T}$ is the temporal latent feature whose columns are $\mathbf{x}_t$'s. Here, K is a positive integer referring to the tensor rank. Element-wise, $y_{i,t}$ is estimated by the inner product of $\mathbf{u}_i$ and $\mathbf{x}_t$, where $\mathbf{u}_i \in \mathbb{R}^K$ represents the latent spatial feature at sensor i and $\mathbf{x}_t \in \mathbb{R}^K$ is the latent temporal embedding at time t, expressed as

$$y_{i,t} \approx \mathbf{u}_i^\top\mathbf{x}_t. \tag{3}$$

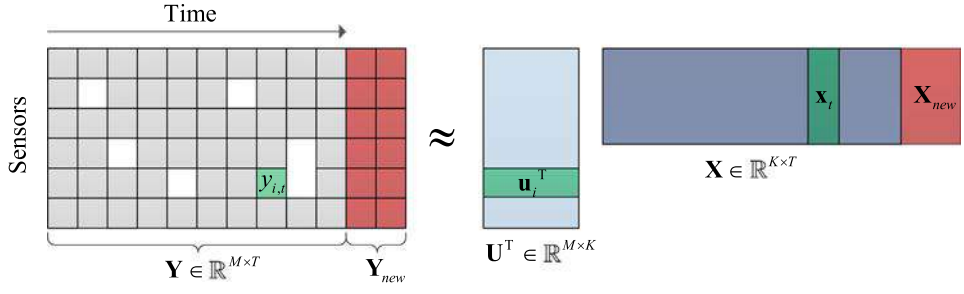


Fig. 1. A graphic illustration of matrix factorization. Note that the white boxes represent the missing values while the grey boxes denote the observed data.

The basic concept of matrix factorization is illustrated in Fig. 1.

Next, we introduce the fully Bayesian method for tensor learning [61]. To begin this process, the likelihood of the observed SHM data $y_{i,t}$ is given by:

$$y_{i,t} \sim \mathcal{N}\left(\mathbf{u}_i^\top \mathbf{x}_t, \tau_\epsilon^{-1}\right), \tag{4}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution with mean $\mathbf{u}_i^\top \mathbf{x}_t$ and precision τ_ϵ. Secondly, to model the spatial factor, the prior distribution over the spatial feature vectors (i.e., $\mathbf{u}_i$) is assumed to be multivariate Gaussian, viz.,

$$\mathbf{u}_i \sim \mathcal{N}\left(\boldsymbol{\mu}_u, \Lambda_u^{-1}\right). \tag{5}$$

We further place conjugate Gaussian-Wishart priors on the spatial feature parameters $\boldsymbol{\Theta}_u = \{\boldsymbol{\mu}_u, \boldsymbol{\Lambda}_u\}$, i.e.,mean vector $\boldsymbol{\mu}_u \in \mathbb{R}^K$ andprecision matrix $\boldsymbol{\Lambda}_u \in \mathbb{R}^{K\times K}$, written as [61]

$$p(\boldsymbol{\Theta}_u|\boldsymbol{\mu}_0, \beta_0, \mathbf{W}_0, \nu_0) = p(\boldsymbol{\mu}_u|\boldsymbol{\Lambda}_u)p(\boldsymbol{\Lambda}_u) = \mathcal{N}\left(\boldsymbol{\mu}_u|\boldsymbol{\mu}_0, (\beta_0\boldsymbol{\Lambda}_u)^{-1}\right)\mathcal{W}(\boldsymbol{\Lambda}_u|\mathbf{W}_0, \nu_0). \tag{6}$$

Here, $\boldsymbol{\mu}_0, \beta_0, \mathbf{W}_0, \nu_0$ are hyper-parameters; $\mathcal{W}(\cdot)$ denotes the Wishart distribution with ν_0 degrees of freedom and a $K \times K$ scale matrix $\mathbf{W}_0$, namely,

$$\mathcal{W}(\boldsymbol{\Lambda}_u|\mathbf{W}_0, \nu_0) = \frac{1}{C}|\boldsymbol{\Lambda}_u|^{\frac{\nu_0-K-1}{2}} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{W}_0^{-1}\boldsymbol{\Lambda}_u\right)\right), \tag{7}$$

where C is the normalizing constant and $\mathrm{Tr}(\cdot)$ denotes the matrix trace defined as the sum of all the elements on the main diagonal of the matrix.

The second stage of model generation is to deal with the precision parameter τ_ϵ as shown in Eq. (4). In particular, a conjugate Gamma prior over τ_ϵ is introduced to make the generative model robust in consideration of the indeterminate noise effect in SHM data:

$$\tau_\epsilon \sim \mathrm{Gamma}(a_0, b_0). \tag{8}$$

Here, we define $\boldsymbol{\Theta}_\tau = \{a_0, b_0\}$ where a_0 and b_0 represent the shape parameter and the rate parameter, respectively. The PDF of τ_ϵ has the form as follows:

$$p(\tau_\epsilon|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}\tau_\epsilon^{a_0-1}\exp(-b_0\tau_\epsilon). \tag{9}$$

Although probabilistic modeling of the spatial factors is straightforward, it is tricky to capture the time-evolving patterns and predict the dynamic trends in the Bayesian learning. Here, we consider incorporating the AR process into the matrix/tensor factorization model for describing the temporal dependencies [61]. Generally, an AR model is characterized by a time lag set and a weight parameter vector. However, different from the traditional AR model which is more applicable for low-dimensional data, we make two modifications to handle the multi-dimensional time-series issue. The first distinction is that we introduce a flexible AR structure on time lags $\mathcal{L}$ [53]. Instead of applying a small-size lag set (e.g., $\mathcal{L} = \{1\}$) which only learns the simple temporal patterns (e.g., daily similarity), we try to use more complex time lags to infer seasonal or yearly trends for long-term forecasting. The second alteration is changing the tensor structure of AR model parameters for convenience. Let the time lags set be $\mathcal{L} = \{l_1, l_2, \ldots, l_d\}$, where d is the order of the AR model. In our case, the weight parameter $\mathbf{A}_j$ ($j \in \{1, 2, \ldots, d\}$) should be a $K \times K$ matrix since the elements in the AR model are formed as column vectors (i.e., $\mathbf{x}_t \in \mathbb{R}^{K\times 1}$) in the temporal feature matrix $\mathbf{X}$. The graphic illustration is shown in Fig. 2 with the example of $\mathcal{L} = \{1, 3\}$.

In addition, there is no diagonal restriction on $\mathbf{A}_j$ due to the complicated causal relationship between factors. Thus, the reorganized formulation of the AR process can be written as:

$$\mathbf{x}_{t+1} \approx \mathbf{A}_1\mathbf{x}_{t+1-l_1} + \mathbf{A}_2\mathbf{x}_{t+1-l_2} + \cdots + \mathbf{A}_d\mathbf{x}_{t+1-l_d} = \underbrace{[\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_d]}_{\mathbf{A}^\top}\underbrace{\left[\mathbf{x}_{t+1-l_1}^\top, \mathbf{x}_{t+1-l_2}^\top, \ldots, \mathbf{x}_{t+1-l_d}^\top\right]^\top}_{\mathbf{z}_{t+1}}. \tag{10}$$

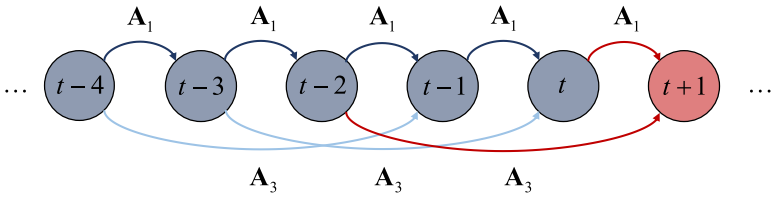


Fig. 2. Auto-regressive model for temporal dependencies.

For simplicity, we define a time-invariant matrix $\mathbf{A} \in \mathbb{R}^{(Kd)\times K}$ and a historical observation vector $\mathbf{z}_{t+1} \in \mathbb{R}^{(Kd)\times 1}$ shown in Eq. (10). As a result, by assuming the prior distribution for the temporal factor $\mathbf{x}_t$ as multivariate Gaussian, we have the mean vector as $\mathbf{A}^\top \mathbf{z}_t$ for the forecasting process. Therefore, the piecewise modeling of the temporal feature matrix is summarized as:

$$\mathbf{x}_t \sim \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_x, \tilde{\boldsymbol{\Sigma}}_x\right) \sim \begin{cases} \mathcal{N}(\mathbf{0}, \mathbf{I}_x), & \text{if } t \in \{1, 2, \ldots, l_d\}, \\ \mathcal{N}\left(\mathbf{A}^\top \mathbf{z}_t, \boldsymbol{\Sigma}\right), & \text{otherwise}, \end{cases} \tag{11}$$

where $\mathbf{0} \in \mathbb{R}^{K\times 1}$ is a zero vector and $\mathbf{I}_x \in \mathbb{R}^{K\times K}$ is an identity matrix.

Likewise, a conjugate Matrix Normal Inverse Wishart prior is applied to the hyper-parameters $\boldsymbol{\Theta}_x = \{\mathbf{A}, \boldsymbol{\Sigma}\}$ in the forecasting process [61]:

$$p(\boldsymbol{\Theta}_x|\boldsymbol{\Lambda}_0, \mathbf{V}_0, \boldsymbol{\Psi}_0, v_0) = p(\mathbf{A}|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}) = \mathcal{MN}(\mathbf{A}|\boldsymbol{\Lambda}_0, \mathbf{V}_0, \boldsymbol{\Sigma})\mathcal{IW}(\boldsymbol{\Sigma}|\boldsymbol{\Psi}_0, v_0), \tag{12}$$

where $\mathcal{MN}(\cdot)$ is Matrix Normal distribution and $\mathcal{IW}(\cdot)$ denotes Inverse Wishart function.Herein, the Inverse-Wishart distribution $\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Psi}_0, v_0)$ is equivalent to $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\boldsymbol{\Psi}_0^{-1}, v_0)$. Besides, the probability density function (PDF) for $\mathbf{A}$ is given by

$$p(\mathbf{A}|\boldsymbol{\Lambda}_0, \mathbf{V}_0, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{K^2 d}{2}} |\mathbf{V}_0|^{-\frac{K}{2}} |\boldsymbol{\Sigma}|^{-\frac{Kd}{2}} \exp\left(-\frac{1}{2}\left[\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{A}-\boldsymbol{\Lambda}_0)^\top \mathbf{V}_0 (\mathbf{A}-\boldsymbol{\Lambda}_0)\right)\right]\right), \tag{13}$$

in which $\boldsymbol{\Lambda}_0 \in \mathbb{R}^{(Kd)\times K}$ is the mean matrix parameter, $\mathbf{V}_0 \in \mathbb{R}^{(Kd)\times(Kd)}$ represents the row-variance matrix, and $\boldsymbol{\Sigma} \in \mathbb{R}^{K\times K}$ denotes the column-variance matrix parameter.

The graphic model representing the generative Bayesian tensor learning described above is depicted in Fig. 3. The grey node $y_{i,t}$ $((i,t) \in \Omega)$ is the observed SHM data, while $\mathbf{u}_i, \mathbf{x}_t$ and τ_ϵ are the parameters in the likelihood distribution 4. In our experiments, we initialize the scalars as: $\beta_0 = 1, v_0 = K, a_0 = b_0 = 1 \times 10^{-6}$. The vector $\boldsymbol{\mu}_0$ and the matrix $\boldsymbol{\Lambda}_0$ are set as a zeros. The remaining matrices $\{\mathbf{W}_0, \mathbf{V}_0, \boldsymbol{\Psi}_0\}$ are all set to be identity matrix but with different dimensions.

It is notable that the proposed Bayesian matrix/tensor learning approach shares similarity with the linear Gaussian state space model (LGSSM) [69,70] and the dynamic linear model (DLM) [29,71,72], especially in the context of temporal feature modeling.However, the Bayesian matrix/tensor learning builds the temporal dependency over the continuous latent features $\mathbf{X}$ as shown in Eq. (2), while LGSSM and DLM operate directly on the raw incomplete data through latent learning. Thus, satisfactory flexibility and scalability are two distinctions that differentiate our method from LGSSM. Firstly, the selection of time lags of our method is more flexible compared with LGSSM and DLM. For example, the time lags of our model can be multiple and discontinuous (e.g., $\{1, 2, 144\}$), more capable of representing the periodicity or seasonality. Secondly, the traditional DLM [71,72] has always encountered the computational inefficiency issue when handling high-dimensional data [53]. Specifically, the classical DLM based on the Kalman filtering scheme exhibits $\mathcal{O}\left(LM^2T + L^3T\right)$ complexity, where L is the number of latent parameters and usually larger than the number of time series M [73]. However, our method can reduce the number of trainable parameters, reduce computational complexity, and offer a scalable choice for data imputation and forecasting, based on a low rank structure and the temporal latent factor modeling scheme (e.g., since the rank K is generally smaller or much smaller than M, the number of resulting latent parameters of our method is also less than L). Thirdly, when

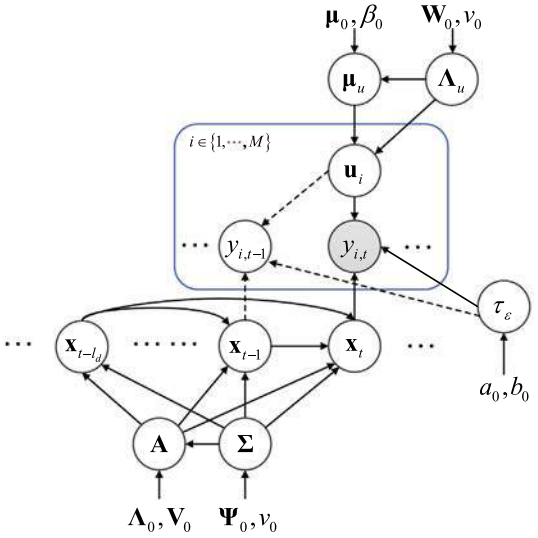


Fig. 3. Probabilistic graphic model for Bayesian matrix/tensor learning [61].

there exists missing data at a high ratio (e.g., $\geqslant 10\%$), challenges arise for LGSSM and DLM to reach satisfactory imputation/-forecasting accuracy since the temporal modeling is directly performed over the raw data under missing conditions. One the contrary, our method tackles this issue via modeling the temporal dependency based on the decomposed latent features which are continuous.

2.3. Missing data imputation

Following the Bayesian modeling formulation in [44,74], we infer the predictive distribution over missing entries below:

$$p\left(y_{i,t}^{*}|\mathbf{Y},\boldsymbol{\Theta}_0^u,\boldsymbol{\Theta}_0^x\right)=\iint p\left(y_{i,t}^{*}|\mathbf{u}_i,\mathbf{x}_t,\tau_\epsilon\right)p(\mathbf{U},\mathbf{X},\tau_\epsilon|\mathbf{Y},\boldsymbol{\Theta}_u,\boldsymbol{\Theta}_x,\boldsymbol{\Theta}_\tau)p(\boldsymbol{\Theta}_u,\boldsymbol{\Theta}_x,\boldsymbol{\Theta}_\tau|\boldsymbol{\Theta}_0^u,\boldsymbol{\Theta}_0^x)\mathrm{d}\{\mathbf{U},\mathbf{X},\tau_\epsilon\}\mathrm{d}\{\boldsymbol{\Theta}_u,\boldsymbol{\Theta}_x,\boldsymbol{\Theta}_\tau\} \tag{14}$$

where $\boldsymbol{\Theta}_0^u=\{\boldsymbol{\mu}_0,\beta_0,\mathbf{W}_0,\nu_0\}$ and $\boldsymbol{\Theta}_0^x=\{\boldsymbol{\Lambda}_0,\mathbf{V}_0,\boldsymbol{\Psi}_0,\nu_0\}$ are assemblies of the hyper-parameters. Noteworthy, the exact solution of Eq. (14) cannot be obtained analytically due to the intricate integration over all the latent variables and hyper-parameters. Therefore, we seek to use Markov Chain Monte Carlo (MCMC) sampling [75] to estimate the distributions. The underlying logic of MCMC sampling is that we can draw dependent sequences of samples representing the posterior distribution. Thus, we can describe the predictive distribution in Eq. (14) as:

$$p\left(y_{i,t}^{*}|\mathbf{Y},\boldsymbol{\Theta}_0^u,\boldsymbol{\Theta}_0^x\right)\approx\frac{1}{N}\sum_{n=1}^{N}p\left(y_{i,t}^{*}|\mathbf{u}_i^{(n)},\mathbf{x}_t^{(n)},\tau_\epsilon^{(n)}\right), \tag{15}$$

where $\left\{\mathbf{u}_i^{(n)},\mathbf{x}_t^{(n)},\tau_\epsilon^{(n)}\right\}$ denote the n^{th} simulated sample from the posterior distribution of interest. Herein, we introduce the Gibbs sampling [76] to generate the posterior samples, which is a sequential sampling approach by sweeping through each variable to sample from its conditional distribution with the remaining variables fixed to their current values. In addition, thanks to the use of conjugate priors in the Bayesian model generation, we can easily derive the conditional distributions since the posterior distribution is in the same probability distribution family as the prior distribution. The Gibbs sampling procedure for all the parameters and hyper-parameters are described below.

2.3.1. Sampling spatial features

We sample the spatial hyper-parameters $\boldsymbol{\Theta}_u$ first. Considering the likelihood in Eq. (5) and the prior in Eq. (6), the posterior distribution is given by a Gaussian-Wishart distribution [61]:

$$p(\boldsymbol{\mu}_u,\boldsymbol{\Lambda}_u|\mathbf{U},\boldsymbol{\Theta}_0^u)=\mathcal{N}\left(\boldsymbol{\mu}_u|\boldsymbol{\mu}_0^{*},\left(\beta_0^{*}\boldsymbol{\Lambda}_u\right)^{-1}\right)\mathcal{W}(\boldsymbol{\Lambda}_u|\mathbf{W}_0^{*},\nu_0^{*})\propto\prod_{i=1}^{M}\mathcal{N}\left(\mathbf{u}_i|\boldsymbol{\mu}_u,\boldsymbol{\Lambda}_u^{-1}\right)\times\mathcal{N}\left(\boldsymbol{\mu}_u|\boldsymbol{\mu}_0,\left(\beta_0\boldsymbol{\Lambda}_u\right)^{-1}\right)\times\mathcal{W}(\boldsymbol{\Lambda}_u|\mathbf{W}_0,\nu_0) \tag{16}$$

where

$$\boldsymbol{\mu}_0^{*}=\frac{\beta_0\boldsymbol{\mu}_0+M\bar{\mathbf{u}}}{\beta_0+M},\beta_0^{*}=\beta_0+M,\nu_0^{*}=\nu_0+M,\left(\mathbf{W}_0^{*}\right)^{-1}=\mathbf{W}_0^{-1}+M\bar{\mathbf{S}}+\frac{\beta_0 M}{\beta_0+M}\left(\boldsymbol{\mu}_0-\bar{\mathbf{u}}\right)\left(\boldsymbol{\mu}_0-\bar{\mathbf{u}}\right)^{\top}. \tag{17}$$

Here, $\bar{\mathbf{u}}$ and $\bar{\mathbf{S}}$ are two statistical parameters defined as:

$$\bar{\mathbf{u}}=\frac{1}{M}\sum_{i=1}^{M}\mathbf{u}_i,\ \bar{\mathbf{S}}=\frac{1}{M}\sum_{i=1}^{M}(\mathbf{u}_i-\bar{\mathbf{u}})(\mathbf{u}_i-\bar{\mathbf{u}})^{\top}. \tag{18}$$

The conditional distribution over spatial features $\mathbf{u}_i$, conditioned on temporal features $\mathbf{X}$, partially observed sensor data $\mathbf{Y}$, precision τ_ϵ and all other hyper-parameters of interest can be obtained [61]:

$$p(\mathbf{u}_i|\mathbf{Y},\mathbf{X},\boldsymbol{\Theta}_u,\tau_\epsilon)=\mathcal{N}\left(\mathbf{u}_i|\boldsymbol{\mu}_u^{*},\left(\boldsymbol{\Lambda}_u^{*}\right)^{-1}\right)\propto\prod_{t=1}^{T}\mathcal{N}\left(y_{i,t}|\mathbf{u}_i^{\top}\mathbf{x}_t,\tau_\epsilon\right)\times\mathcal{N}\left(\mathbf{u}_i|\boldsymbol{\mu}_u,\left(\boldsymbol{\Lambda}_u\right)^{-1}\right), \tag{19}$$

where

$$\boldsymbol{\Lambda}_u^{*}=\boldsymbol{\Lambda}_u+\tau_\epsilon\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^{\top},\boldsymbol{\mu}_u^{*}=\left(\boldsymbol{\Lambda}_u^{*}\right)^{-1}\left(\tau_\epsilon\sum_{t=1}^{T}\mathbf{x}_t y_{i,t}+\boldsymbol{\Lambda}_u\boldsymbol{\mu}_u\right),\ (i,t)\in\Omega. \tag{20}$$

2.3.2. Sampling temporal features

Following the sampling procedure for spatial features, we infer the conditional distribution of the hyper-parameters $\boldsymbol{\Theta}_x$ with the likelihood in Eq. (11) and the prior in Eq. (12), namely,

$$p(\mathbf{A},\boldsymbol{\Sigma}|\mathbf{X},\boldsymbol{\Theta}_0^x)=\mathcal{MN}(\mathbf{A}|\boldsymbol{\Lambda}_0^{*},\mathbf{V}_0^{*},\boldsymbol{\Sigma})\mathcal{IW}(\boldsymbol{\Sigma}|\boldsymbol{\Psi}_0^{*},\nu_0^{*})\propto\prod_{t=1}^{T}\mathcal{N}\left(\mathbf{x}_t|\tilde{\boldsymbol{\mu}}_x,\tilde{\boldsymbol{\Sigma}}_x\right)\times\mathcal{MN}(\mathbf{A}|\boldsymbol{\Lambda}_0,\mathbf{V}_0,\boldsymbol{\Sigma})\times\mathcal{IW}(\boldsymbol{\Sigma}|\boldsymbol{\Psi}_0,\nu_0). \tag{21}$$

Matching the coefficients of the hyper-parameters in Eq. (21), we can obtain the updated parameters as follows [61]:

$$\mathbf{V}_0^* = \left(\mathbf{V}_0^{-1} + \mathbf{Q}^\top \mathbf{Q}\right)^{-1}, \boldsymbol{\Lambda}_0^* = \mathbf{V}_0^* \left(\mathbf{V}_0^{-1} \boldsymbol{\Lambda}_0 + \mathbf{Q}^\top \mathbf{P}\right), v_0^* = v_0 + T - l_d, \boldsymbol{\Psi}_0^* = \boldsymbol{\Psi}_0 + \mathbf{P}^\top \mathbf{P} + \boldsymbol{\Lambda}_0^\top \mathbf{V}_0^{-1} \boldsymbol{\Lambda}_0 - \left(\boldsymbol{\Lambda}_0^*\right)^\top \left(\mathbf{V}_0^*\right)^{-1} \boldsymbol{\Lambda}_0^*. \tag{22}$$

These two matrices $\mathbf{P} \in \mathbb{R}^{(T-l_d)\times K}$ and $\mathbf{Q} \in \mathbb{R}^{(T-d)\times(Kd)}$ are defined for simplicity and convenience, expressed as

$$\mathbf{P} = \left[\mathbf{x}_{l_d+1}^\top, \ldots, \mathbf{x}_T^\top\right]^\top, \mathbf{Q} = \left[\mathbf{z}_{l_d+1}^\top, \ldots, \mathbf{z}_T^\top\right]^\top. \tag{23}$$

After sampling the hyper-parameters, we further derive the conditional distribution of the temporal factor $\mathbf{x}_t$, whose posterior distribution follows Gaussian, given by

$$p(\mathbf{x}_t|\mathbf{Y}, \mathbf{U}, \boldsymbol{\Theta}_x, \tau_\epsilon) = \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_x^*, \boldsymbol{\Sigma}_x^*\right) \propto \prod_{i=1}^{M} \mathcal{N}\left(y_{i,t}|\mathbf{u}_i^\top \mathbf{x}_t, \tau_\epsilon\right) \times \mathcal{N}\left(\mathbf{x}_t|\widetilde{\boldsymbol{\mu}}_x, \widetilde{\boldsymbol{\Sigma}}_x\right). \tag{24}$$

Nevertheless, sampling $\mathbf{x}_t$ is complicated due to the piecewise Bayesian modeling on the temporal feature parameters. Here, we introduce four auxiliary variables $\{\mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}\}$ considering the function of the AR process. The general updating formulation can thus be written as [61]

$$\boldsymbol{\Sigma}_x^* = \left(\tau_\epsilon \sum_{i=1}^{M} \mathbf{u}_i \mathbf{u}_i^\top + \mathbf{C} + \mathbf{D}\right)^{-1}, \boldsymbol{\mu}_x^* = \boldsymbol{\Sigma}_x^* \left(\tau_\epsilon \sum_{i=1}^{M} \mathbf{u}_i y_{i,t} + \mathbf{E} + \mathbf{F}\right), (i,t) \in \Omega, \tag{25}$$

where the variables $\mathbf{C}$ and $\mathbf{E}$ are given by

$$\mathbf{C} = \begin{cases} \sum_{j=1, l_d < t+l_j \leqslant T}^{d} \mathbf{A}_j^\top \boldsymbol{\Sigma}^{-1} \mathbf{A}_j, & \text{if } t \in \{1, 2, \ldots, T - l_1\}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{26}$$

$$\mathbf{E} = \begin{cases} \sum_{j=1, l_d < t+l_j \leqslant T}^{d} \mathbf{A}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}_{t+l_j}, & \text{if } t \in \{1, 2, \ldots, T - l_1\}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{27}$$

with $\boldsymbol{\phi}_{t+l_j}$ being defined as

$$\boldsymbol{\phi}_{t+l_j} = \mathbf{x}_{t+l_j} - \sum_{p=1, p\neq j}^{d} \mathbf{A}_p \mathbf{x}_{t+l_j-l_p}. \tag{28}$$

In addition, the variables $\mathbf{D}$ and $\mathbf{F}$ can be written as

$$\mathbf{D} = \begin{cases} \mathbf{I}_x, & \text{if } t \in \{1, 2, \ldots, l_d\}, \\ \boldsymbol{\Sigma}^{-1}, & \text{otherwise}, \end{cases} \tag{29}$$

$$\mathbf{F} = \begin{cases} \mathbf{0}, & \text{if } t \in \{1, 2, \ldots, l_d\}, \\ (\boldsymbol{\Sigma}^{-1}) \sum_{p=1}^{d} \mathbf{A}_p \mathbf{x}_{t-l_p}, & \text{otherwise}. \end{cases} \tag{30}$$

2.3.3. Sampling precision

With the combination of the likelihood in Eq. (4) and the prior in Eq. (8), the posterior distribution of precision τ_ϵ can be represented by a Gamma distribution [61], namely,

$$p(\tau_\epsilon|\mathbf{Y}, \mathbf{U}, \mathbf{X}, \boldsymbol{\Theta}_\tau) = \text{Gamma}\left(a_0^*, b_0^*\right) \propto \prod_{i=1}^{M} \prod_{t=1}^{T} \mathcal{N}\left(y_{i,t}|\mathbf{u}_i^\top \mathbf{x}_t, \tau_\epsilon\right) \times \text{Gamma}(\tau_\epsilon|a_0, b_0), \tag{31}$$

where the hyper-parameters a_0^* and b_0^* can be expressed as

$$a_0^* = \frac{1}{2} \sum_{(i,t)\in\Omega} s_{i,t} + a_0, b_0^* = \frac{1}{2} \sum_{(i,t)\in\Omega} \left(y_{i,t} - \mathbf{u}_i^\top \mathbf{x}_t\right)^2 + b_0. \tag{32}$$

Note that $s_{i,t}$ is 1 if $(i,t) \in \Omega$ and 0 otherwise.

2.4. Structural response forecasting

We predict the future structural response $y_{i,t+1}$ based on both incrementally updated spatial feature $\mathbf{U}$ (see Fig. 4) and temporal feature $\mathbf{x}_{t+1}$ (see Fig. 1), but set a periodical updating constraint on forecasting the spatial attribute considering computational efficiency. Namely, after getting well-trained parameters from the imputation process, we keep $\{\mathbf{X}, \mathbf{A}\}$ unchanged for the forecasting step and only view $\{\mathbf{U}, \mathbf{\Sigma}, \mathbf{x}_{t+1}, \tau_\epsilon\}$ as the updated targets. Moreover, to predict $y_{i,t+2}$ sequentially, we provide the observed $y_{i,t+1}$ as an input and conduct the above procedure iteratively. The general philosophy of Bayesian forecasting can be illustrated by two steps as follows.

The first step is to learn $\{\mathbf{U}, \mathbf{\Sigma}, \mathbf{x}_t, \tau_\epsilon\}$ from the historical observation $\mathbf{y}_{:,t}$. The model generative formulations are expressed as:

$$y_{i,t} \sim \mathcal{N}\left(\mathbf{u}_i^\top \mathbf{x}_t, \tau_\epsilon^{-1}\right), \mathbf{x}_t \sim \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_x, \tilde{\mathbf{\Sigma}}_x\right), \tilde{\mathbf{\Sigma}}_x \sim \mathcal{IW}(\mathbf{\Psi}_0, \nu_0), \tau_\epsilon \sim \text{Gamma}(a_0, b_0), \tag{33}$$

where $\tilde{\boldsymbol{\mu}}_x$ is a known parameter denoted as $\mathbf{A}^\top \mathbf{z}_t$. Note that there is possible missing values in $\mathbf{y}_{:,t}$. The model inference using Gibbs sampling for this step is divided into three parts. To begin with, if t is at the end of batch window, we need to sample the spatial factor and its hyper-parameters referring to Eqs. (16) and (19). Secondly, for the temporal feature, we do sampling on the hyper-parameter $\left(\tilde{\mathbf{\Sigma}}_x\right)^{-1} \sim \mathcal{W}\left(\left(\mathbf{\Psi}_0^*\right)^{-1}, \nu_0 + 1\right)$ where

$$\mathbf{\Psi}_0^* = \mathbf{\Psi}_0 + \left(\mathbf{x}_t - \mathbf{A}^\top \mathbf{z}_t\right)\left(\mathbf{x}_t - \mathbf{A}^\top \mathbf{z}_t\right)^\top. \tag{34}$$

Then we sample the future temporal factor $\mathbf{x}_t \sim \mathcal{N}\left(\left(\tilde{\boldsymbol{\mu}}_x\right)^*, \left(\tilde{\mathbf{\Sigma}}_x\right)^*\right)$ with

$$\tilde{\mathbf{\Sigma}}_x^* = \left(\tau_\epsilon \sum_{i=1}^{M} \mathbf{u}_i \mathbf{u}_i^\top + \tilde{\mathbf{\Sigma}}_x^{-1}\right)^{-1}, \tilde{\boldsymbol{\mu}}_x^* = \tilde{\mathbf{\Sigma}}_x^* \left(\tau_\epsilon \sum_{i=1}^{M} \mathbf{u}_i y_{i,t} + \tilde{\mathbf{\Sigma}}_x^{-1} \mathbf{A}^\top \mathbf{z}_t\right). \tag{35}$$

The third part is the sampling of the precision parameter τ_ϵ using Eq. (32).

After generating samples of $\mathbf{x}_t$, the second step for forecasting is that we run Gibbs sampling on the prediction for multiple iterations based on $\mathbf{y}_{:,t+1} \approx \mathbf{U}^\top \left(\mathbf{A}^\top \mathbf{z}_{t+1}\right)$, and get the average of these samples in the burn-in period as output. This is an efficient strategy for forecasting, especially for large-scale problems.

2.5. Incremental learning scheme

For continuous SHM, data streams over time where imputation and forecasting should be ideally done in a real-time manner accounting for new records. This typically requires online learning with model re-training involved, resulting in significant computational burden especially when large-scale data analysis is performed. To this end, inspired by the work in [58], we present an incremental learning scheme as illustrated in Fig. 4. Instead of retraining the model when new data arrives, in the incremental scheme we only take the up-to-date information from sensors within a certain number of time stamps for intermittent training/updating. The benefits of this proposed scheme are twofold: (1) possessing efficiency and alleviating the computational burden induced by online model re-training for every time stamps, and (2) maintaining satisfactory accuracy thanks to the use of streaming sensing data.

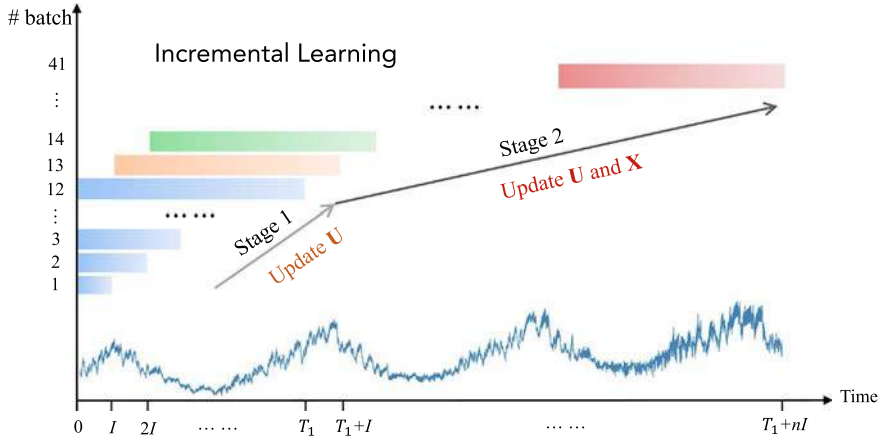


Fig. 4. The proposed incremental learning scheme. Note that the dynamic batch window size is $I, 2I, \ldots, T_1$ for each short-time updating period respectively, while the long-period fixed batch window remains constant (e.g. T_1) after reaching a critical point (e.g., one year).

Specifically, there are two updating stages for data imputation and response forecasting in continuous SHM: (1) short-time dynamic batch window and (2) long-period fixed batch window. The dynamic batch window stage learns the latent spatial attribute $\mathbf{U}$ from previous records during the time interval $[0, I]$ when I is small (e.g., one month), where I denotes the length of data for forward imputation period. Then we fix $\mathbf{U}$ for response forecasting within $[I, 2I]$ and execute the next spatial information updating using records during $[0, 2I]$. The rest of the first dynamic tensor learning stage will follow the same manner until the total imputation time reaches one critical time stamp T_1 (e.g., one year), where $\mathbf{U}$ is incrementally updated. In the second stage, we update the spatial attribute $\mathbf{U}$ and the temporal factor $\mathbf{X}$ simultaneously due to the constant temporal dimension of the fixed batch window. For instance, $\mathbf{U}$ and $\mathbf{X}$ will be updated in the time period $[I, T_1 + I]$ for imputation, then we still keep $\mathbf{U}$ unchanged and perform forecasting within $[T_1 + I, T_1 + 2I]$. Subsequently, we repeat the imputation and forecasting procedure with the fixed batch window as the continuous monitoring proceeds. The basic concept of the proposed incremental learning scheme for semi-online tensor learning is presented in Fig. 4. Such a scheme enables recovery of missing data incrementally (e.g., every I-unit increment) and forecasting of structural response on the fly for long-term SHM. The process can be realized through a Bayesian tensor learning approach [61], which is introduced above.

The pseudo code for the proposed incremental Bayesian tensor learning for missing SHM data imputation and structural response forecasting is summarized in Algorithms 1 and 2. Empirically, the general guideline for selecting the Gibbs sampling parameters is that we can use small number of samples (e.g., 50) for both chain length $N_1^{\text{mc}}, N_2^{\text{mc}}$ and burn-in period $N_1^{\text{b}}, N_2^{\text{b}}$ at low missing rates (e.g., no more than 10%). For a larger missing rate, it is suitable to proportionally increase the number of samples for $\left\{N_1^{\text{mc}}, N_2^{\text{mc}}, N_1^{\text{b}}, N_2^{\text{b}}\right\}$. Regarding the setting of time lags, we can define it as the combination of different time slots (e.g., $\{1, 2, 144\}$) to account for different periodic patterns. The forecasting length and the forward batch window can be set as one month while the critical point can be defined as one year. The selection of rank varies significantly and is analyzed in Section 3.5.

Algorithm 1: Incremental tensor learning

Input: the SHM data tensor $\mathbf{Y}$, the indicator tensor $\boldsymbol{\Omega}$, the chain length for imputation N_1^{mc}, the burn-in period for imputation N_1^{b}, the chain length for forecasting N_2^{mc}, the burn-in period for forecasting N_2^{b}, tensor rank K, time lags $\mathcal{L}$, forecasting length $\mathcal{T}$, forward batch length I and critical point T_1.

Output: the chains of samples for the total imputed tensor $\widehat{\mathbf{Y}}_1$ and the forecasted tensor $\widehat{\mathbf{Y}}_2$.

Initialize: random $\mathbf{U}$;
$N_{\text{s}} = T_1/I$, $N_{\text{total}} = T_{\text{total}}/I$;
Define a counting tensor for averaging imputations $\mathbf{C} \leftarrow \mathbf{0}$;
for $w = 1, \cdots, N_{\text{total}}$ **do**
    `// Short time dynamic batch window`
    if $w \leqslant N_s$ **then**
        $\mathbf{U}^{(w)} \leftarrow$ updated $\mathbf{U}$;
        $\mathbf{X}^{(w)} \leftarrow$ randomly initialized $\mathbf{X}$;
        $\mathbf{U}, \widehat{\mathbf{Y}}_1^{(w)}, \widehat{\mathbf{Y}}_2^{(w)} \leftarrow$ impute and forecast with $\mathbf{U}^{(w)}, \mathbf{X}^{(w)}$ in $[0, wI]$ (Algorithm 2);
        $\mathbf{C}^{(w)} \leftarrow \mathbf{1}$;
    end
    `// Long-term fixed batch window`
    else
        $\mathbf{U}^{(w)} \leftarrow$ updated $\mathbf{U}$;
        $\mathbf{X}^{(w)} \leftarrow$ updated $\mathbf{X}$;
        $\mathbf{U}, \mathbf{X}, \widehat{\mathbf{Y}}_1^{(w)}, \widehat{\mathbf{Y}}_2^{(w)} \leftarrow$ impute and forecast with $\mathbf{U}^{(w)}, \mathbf{X}^{(w)}$ in $[(w - N_{\text{s}})I, wI]$ (Algorithm 2);
        $\mathbf{C}^{(w)} \leftarrow \mathbf{1}$;
    end
end
Collect and average the imputations $\widehat{\mathbf{Y}}_1 \leftarrow \text{sum}(\widehat{\mathbf{Y}}_1(1 : N_{\text{total}}))/\text{sum}(\mathbf{C}(1 : N_{\text{total}}))$;
Collect the forecasting $\widehat{\mathbf{Y}}_2 \leftarrow \text{sum}(\widehat{\mathbf{Y}}_2(1 : N_{\text{total}}))$;

Algorithm 2: Gibbs sampling for Bayesian tensor learning

Input: the SHM data tensor $\mathbf{Y}^{(w)}$, the indicator tensor $\mathbf{\Omega}^{(w)}$, the chain length for imputation N_1^{mc}, the burn-in period for imputation N_1^{b}, the chain length for forecasting N_2^{mc}, the burn-in period for forecasting N_2^{b}, tensor rank K, time lags $\mathcal{L}$ and forecasting length $\mathcal{T}$.

Output: updated $\mathbf{U}$, updated $\mathbf{X}$, the chains of samples for the estimated tensor $\widehat{\mathbf{Y}}_1^{(w)}$ and the predicted tensor $\widehat{\mathbf{Y}}_2^{(w)}$.

Initialize: random $\mathbf{U}^{(w)}$, random $\mathbf{X}^{(w)}$, random $\mathbf{A}$, $\mathbf{\Theta}_0^u$, $\mathbf{\Theta}_0^x$ and $\mathbf{\Theta}_\tau$;
`// The imputation process`
for $n_1 = 1, \cdots, N_1^{\mathrm{mc}}$ **do**
  Sample the hyperparameter $\mathbf{\Theta}_u$ (Eq. 16);
  $\mathbf{\Theta}_u \sim p(\mathbf{\Theta}_u|\mathbf{U}^{(w)}, \mathbf{\Theta}_0^u)$;
  for $i = 1, \cdots, M$ **do**
    Sample the spatial feature $\mathbf{u}_i$ (Eq. 19);
    $\mathbf{u}_i \sim p(\mathbf{u}_i|\mathbf{Y}^{(w)}, \mathbf{X}^{(w)}, \mathbf{\Theta}_u, \tau_\epsilon)$;
  end
  Sample the hyperparameter $\mathbf{\Theta}_x$ (Eq. 21);
  $\mathbf{\Theta}_x \sim p(\mathbf{\Theta}_x|\mathbf{X}^{(w)}, \mathbf{\Theta}_0^x)$;
  for $t = 1, \cdots, T$ **do**
    Sample the temporal feature $\mathbf{x}_t$ (Eq. 24);
    $\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{Y}^{(w)}, \mathbf{U}^{(w)}, \mathbf{\Theta}_x, \tau_\epsilon)$;
  end
  Sample the precision parameter τ_ϵ (Eq. 31);
  $\tau_\epsilon \sim p(\tau_\epsilon|\mathbf{Y}^{(w)}, \mathbf{U}^{(w)}, \mathbf{X}^{(w)}, \mathbf{\Theta}_\tau)$;
  if $n_1 \geqslant N_1^{\mathrm{b}}$ **then**
    Compute and collect the sample $\widehat{\mathbf{Y}}_1^{(w)} = [\mathbf{U}^{(w)}]^\top \mathbf{X}^{(w)}$;
  end
end
`// The forecasting process`
for $s = 1, \cdots, \mathcal{T}$ **do**
  for $n_2 = 1, \cdots, N_2^{\mathrm{mc}}$ **do**
    if $s \bmod I == 0$ **then**
      Sample the hyperparameter $\mathbf{\Theta}_u$ (Eq. 16);
      Sample the spatial feature $\mathbf{U}$ (Eq. 19);
    end
    Sample the hyperparameter $\widetilde{\mathbf{\Sigma}}_x$ (Eq. 34);
    Sample the temporal feature $\mathbf{x}_t$ (Eq. 35);
    Sample the precision parameter τ_ϵ (Eq. 31);
  end
end

2.6. Extension to Bayesian high-order tensor learning

Furthermore, we extend the data format to a third-order tensor $\mathbf{Y} \in \mathbb{R}^{M \times N \times T}$. It can be a versatile tensor with data representation of "sensors × sensor types × time stamps", "transverse sensors × longitudinal sensors × time stamps" or "sensors × days × time stamps per day", depending on specific sensor types/locations. In particular, we consider the tensor format of "sensors × days × time stamps per day" for data imputation due to the lack of rich spatial distribution of sensors in this work. We leverage the CANDECOMP/PARAFAC (CP) decomposition [41] to extract the factors, viz., $\mathbf{Y} \approx \sum_{p=1}^{P} \mathbf{u}_p \circ \mathbf{v}_p \circ \mathbf{x}_p$. Here, P is a positive integer; $\mathbf{u}_p \in \mathbb{R}^M, \mathbf{v}_p \in \mathbb{R}^N$, and $\mathbf{x}_p \in \mathbb{R}^T$ are the rank-1 tensors of the pth column factor matrices. Intuitively, we can extend presented Bayesian matrix factorization to probabilistic high-order tensor learning, whose element-wise form is given by:

$$y_{i,j,t} \sim \mathcal{N}\left(\sum_{p=1}^{P} u_{i,p} v_{j,p} x_{t,p},\ \tau_{\epsilon}^{-1}\right), (i,j,t) \in \Omega, \tag{36}$$

where Ω is a third-order indicator set for observed elements. Similarly, the hierarchical Bayesian modeling of $\mathbf{u}_i, \mathbf{x}_t$ and τ_ϵ is consistent to that of the Bayesian matrix factorization using on Eqs. (5), (11) and (8) respectively. For the factor $\mathbf{v}_j$, we also assume its prior distribution as multivariate Gaussian, namely,

$$\mathbf{v}_j \sim \mathcal{N}\left(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v^{-1}\right). \tag{37}$$

The same conjugate Gaussian-Wishart priors are applied on the hyper-parameters of $\mathbf{u}_i$ (i.e., $\boldsymbol{\Theta}_u = \{\boldsymbol{\mu}_u, \boldsymbol{\Lambda}_u\}$) and those of $\mathbf{v}_j$ (i.e., $\boldsymbol{\Theta}_v = \{\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v\}$) as shown in Eq. (6), while the Matrix Normal Inverse Wishart prior is employed for the hyper-parameters of $\mathbf{x}_t$ (i.e., $\boldsymbol{\Theta}_x = \{\mathbf{A}, \boldsymbol{\Sigma}\}$) as shown in Eq. (12).

The model inference can be realized via the holistic sampling procedure analogical to that of Bayesian matrix factorization. The posterior distributions of $\boldsymbol{\Theta}_u$ and $\boldsymbol{\Theta}_v$ have the same formulations as described in Eqs. (16) and (17). The posterior distribution of $\boldsymbol{\Theta}_x$ has the identical form shown in Eqs. (21) and (22). Nevertheless, slight difference is present on the posterior distributions of $\mathbf{u}_i, \mathbf{v}_j$ and $\mathbf{x}_t$ between Bayesian tensor decomposition and matrix factorization. In particular, the conditional distribution over $\mathbf{u}_i$ is given by $\mathcal{N}\left(\mathbf{u}_i | \boldsymbol{\mu}_u^*, \left(\boldsymbol{\Lambda}_u^*\right)^{-1}\right)$ with

$$\boldsymbol{\Lambda}_u^* = \boldsymbol{\Lambda}_u + \tau_\epsilon \sum_{j,t} \mathbf{w}_{j,t} \mathbf{w}_{j,t}^\top, \boldsymbol{\mu}_u^* = \left(\boldsymbol{\Lambda}_u^*\right)^{-1} \left(\tau_\epsilon \sum_{j,t} \mathbf{w}_{j,t} y_{i,j,t} + \boldsymbol{\Lambda}_u \boldsymbol{\mu}_u\right), \mathbf{w}_{j,t} = \mathbf{v}_j \odot \mathbf{x}_t, (i,j,t) \in \Omega. \tag{38}$$

The conditional distribution on $\mathbf{v}_j$ share the same probabilistic inference as Eq. (38). The conditional distribution over $\mathbf{x}_t$ can be described by $\mathcal{N}\left(\mathbf{x}_t | \boldsymbol{\mu}_x^*, \boldsymbol{\Sigma}_x^*\right)$ with

$$\boldsymbol{\Sigma}_x^* = \left(\tau_\epsilon \sum_{i,j} \mathbf{w}_{i,j} \mathbf{w}_{i,j}^\top + \mathbf{C} + \mathbf{D}\right)^{-1}, \boldsymbol{\mu}_x^* = \boldsymbol{\Sigma}_x^* \left(\tau_\epsilon \sum_{i,j} \mathbf{w}_{i,j} y_{i,j,t} + \mathbf{E} + \mathbf{F}\right), \mathbf{w}_{i,j} = \mathbf{u}_i \odot \mathbf{v}_j, (i,j,t) \in \Omega. \tag{39}$$

Here, $\{\mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}\}$ are the same variables defined in Eqs. (26), (29), (27) and (30), respectively. For the precision, the posterior distribution of τ_ϵ is expressed as $\text{Gamma}\left(a_0^*, b_0^*\right)$ where

$$\begin{aligned} a_0^* &= \tfrac{1}{2} \sum_{(i,j,t) \in \Omega} s_{i,j,t} + a_0, \\ b_0^* &= \tfrac{1}{2} \sum_{(i,j,t) \in \Omega} \left(y_{i,j,t} - \sum_{r=1}^{K} \mathbf{u}_{i,r} \mathbf{v}_{j,r} \mathbf{x}_{t,r}\right)^2 + b_0. \end{aligned} \tag{40}$$

Note that $s_{i,j,t}$ is 1 if $(i,j,t) \in \Omega$ and 0 otherwise.

3. Experimental validation

In this section, we test the imputation and forecasting performance of the proposed Bayesian tensor learning method under data missing scenarios, using long-term field-monitoring data of a concrete bridge (e.g., strain and temperature records). In particular, we impute and forecast the strain time histories of the bridge. Inspired by the strong correlation between strain and temperature, we formulate the tensor data structure by combining both strain and temperature along the sensor dimension. We also conduct a series of analyses of uncertainty quantification and rank selection for tensor factorization. The numerical analyses were performed on a standard workstation with 28 Intel Core i9-7940X CPUs.

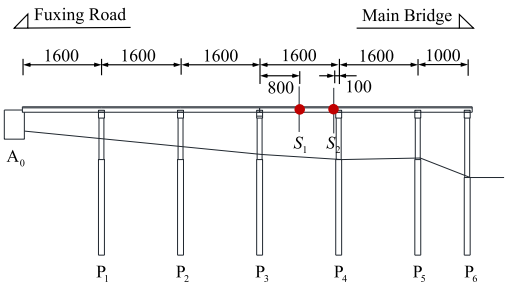
3.1. Bridge description

The instrumented concrete bridge (see Fig. 5(a)) considered herein is a connection bridge located in the old section of Wanzhou District, Chongqing, China. It has the total length of 94.015 m, whose span composition is 5×16 m + 10 m (see Fig. 5(c)). The superstructure of this bridge consists of continuous hollow slab beams constructed of reinforced concrete. As shown in Fig. 5(c), we name the bridge abutment A_0 and the piers $\{P_1, P_2, P_3, P_4, P_5, P_6\}$ orderly according to the vehicle moving direction to the main bridge. Two sections are monitored, marked as S_1 and S_2 in the mid-span of the fourth span and near the top of pier P_4, respectively. Fig. 5(d) shows that each monitoring section $S_i (i \in \{1,2\})$ has five strain sensors installed on the bottom of the hollow slab beam. Vibrational chord strain gauges are installed which facilitate monitoring of both strain response of the bridge and the corresponding operation temperature (see Fig. 5(b)).

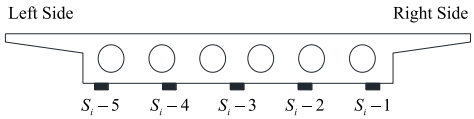
The dataset collected from the above SHM system contains strain and temperature time histories recorded from June 1, 2015 to October 14, 2018. We resample the data at the rate of 10 min interval. Thus, it can be organized as a two-dimensional tensor with both strain data and temperature data (with a size of $20 \times 177{,}408$, representing sensors $\times$ time stamps). The salient feature behind this data arrangement is that the tensor structure with both strain and temperature

(a) Elevation view

(b) Vibrational chord strain gauge

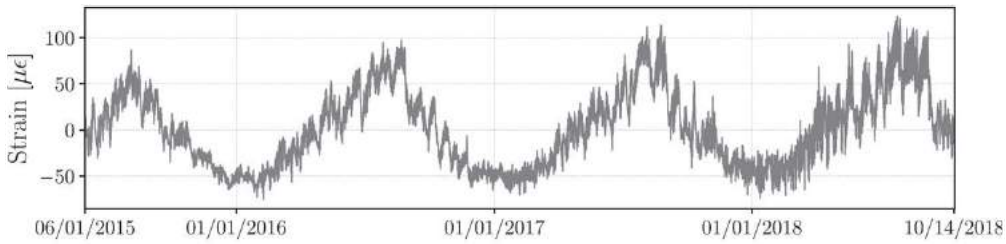


(c) Monitoring sections of strain sensors

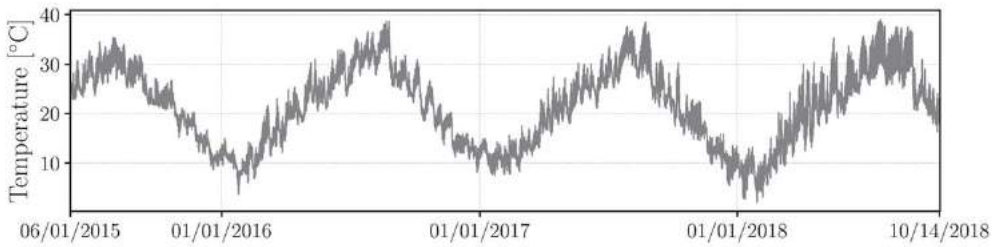


(d) Strain sensor locations at a typical section

Fig. 5. The instrumented concrete bridge.



(a) Measured strain data



(b) Measured temperature data

Fig. 6. Time series of measurements at sensor $S_1 - 1$.

can capture a lower rank compared to the tensor structure with only strain data, due to the strong correlation between strain and temperature. Fig. 6 illustrates the recorded strain and temperature time series for over three years from a typical sensor (e.g., $S_1 - 1$ as shown in Fig. 5(c) and (d)).

3.2. Scenarios of missing data

First of all, to evaluate the proposed model for imputation and forecasting, we only set data missing on the strain recordings while keeping the temperature data fixed/known. Namely, the missing rate η (e.g., 20%) is introduced for the strain data in the validation experiments, which is computed as the ratio of the amount of missing data to the total amount of measurements. Secondly, to simulate the real-world missing conditions during monitoring period, we define three primary missing scenarios for the two-dimensional tensor data considered herein, following a similar experimental design procedure for higher-dimensional tensors discussed in [59]. The first scenario is called "random missing" (RM) which presents discrete and arbitrary lack of data in the time histories. Each strain entry in the data matrix is dropped randomly (e.g., following a uniform random distribution). The second scenario is termed as "structured missing" (SM) where there is data missing occurs continuously for certain periods (e.g., one day or consecutive days). It is a common scenario in practical SHM applications due to sensor malfunctioning, but more challenging and less investigated in literature. In particular, we structurally remove the strain data by selecting multiple days randomly and dropping the corresponding data to simulate a practical missing condition. The last scenario is named "mixed missing" (MM) which combines random missing and structured missing at different rates.

After setting the different missing scenarios, we define a sparse binary matrix $\mathbf{B} \in \mathbb{R}^{M\times T}(b_{i,t} = 1 \text{ if } (i,t) \in \Omega \text{ and 0 otherwise})$ to record the missing positions for the subsequent comparison between imputation results and the ground truth. The target dense tensor without data missing is named $\mathbf{Y}_{\mathrm{d}}$, and the partially observed tensor $\mathbf{Y}$ can be calculated element-wisely by $\mathbf{B} \odot \mathbf{Y}_{\mathrm{d}}$, where $\odot$ denotes the Hadamard product. The imputation/-forecasting accuracy ρ is defined as the root mean square error (RMSE) between the reconstructed/predicted data and the corresponding ground truth, normalized by the root mean square (RMS) of the target values:

$$\rho = \left(1 - \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}}\right) \times 100\%. \tag{41}$$

where y_i and y_i^* denote the ground truth and the estimated value at the same missing position i, and n is the total number of missing entries.

3.3. Results

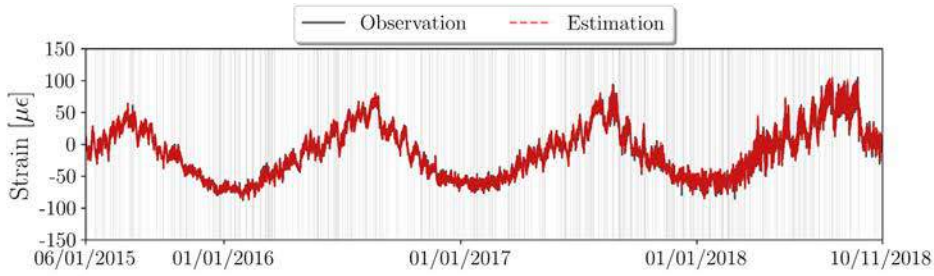
We test the overall performance of the proposed method and identify its limit of capacity for data imputation and response forecasting under various missing settings with different missing rates. We assume that the time lags $\mathcal{L}$ is $\{1, 2, 144\}$. In the incremental learning, the forward batch window length I is defined as 30 days with 4,320 data points, while the critical time stamps T_1 is one year (i.e., 12×30). The dataset for imputation ranges from June 1, 2015 to October 11, 2018 with 1,230 days in total, and the forecasting data is from July 1, 2015 to October 11, 2018 with thirty-days data ahead of the imputation dataset. To begin with, we first consider the missing rate of 10%, for both random and structured missing scenarios, while setting the tensor rank of eight. In addition, keeping the tensor rank fixed, we also set two mixed missing cases: Case 1 for 10% structured and 20% random missing occurring at the same time, while Case 2 for 20% structured and 30% random missing simultaneously. Here, sensor $S_2 - 4$ is selected to showcase the result.

Figs. 7 and 8 show the corresponding imputation and forecasting result obtained by the proposed incremental Bayesian tensor learning model. It can be seen that the predicted time series match well with the ground truth. In particular, the imputed data possess excellent agreement with the ground truth (see Fig. 7), while the forecasted response has relatively larger errors especially for the mixed missing cases with overall large missing rates (e.g., Case 1 and Case 2) as shown in Fig. 8(c) and (d). In general, the spatiotemporal dependencies of the data are well learned by the proposed model. Besides, we provide three representative segments (zoomed view) of the predicted response by choosing one-month strain (March 2016, December 2017 and August 2018) for showcase of imputation (see Fig. 9) and forecasting (see Fig. 10).

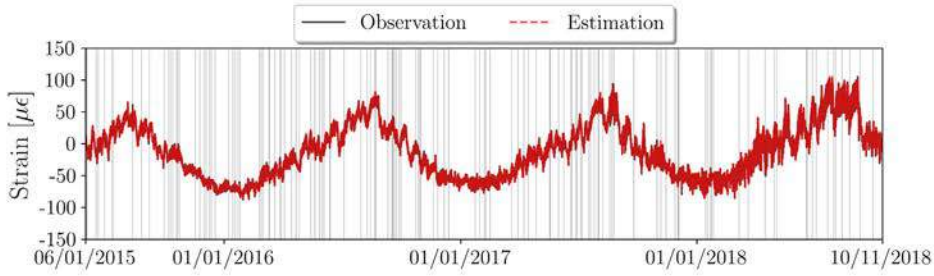
It is notable that, despite large missing rates, the imputation is very robust and produces excellent estimation as shown in Fig. 9. Though the forecasted responses exhibit noisy oscillations depicted in Fig. 10, the overall trend is well captured (especially for relatively smaller missing rates, e.g., 10%).

To further investigate the performance of the proposed approach, we conduct uncertainty quantification of the prediction. Thanks to the incremental learning architecture, we can achieve the convergence with fewer Monte Carlo samples (e.g., $10 \sim 50$) for small missing rates (e.g., 10%). The probabilistic imputation and forecasting results are summarized in Figs. 11 and 12 respectively, where the predicted mean and three standard deviations of the strain response are shown for Sensor $S_2 - 4$ in comparison with the ground truth records on March 10, 2016, December 15, 2017 and August 22, 2018.

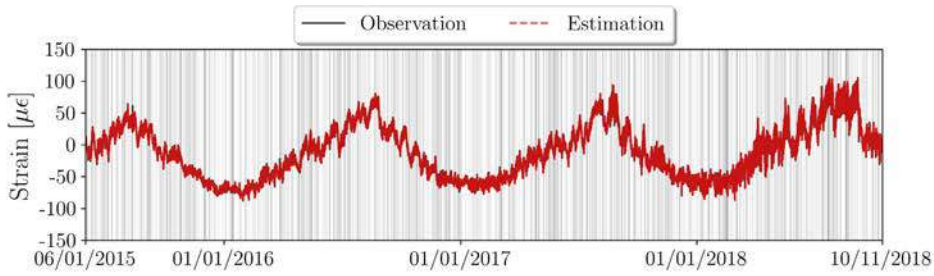
For all these four missing scenarios, the imputation uncertainty is prominently smaller than the forecasting uncertainty. In the forecasting cases, it is observed that the missing data cause prediction fluctuations, which leads to deviation from the



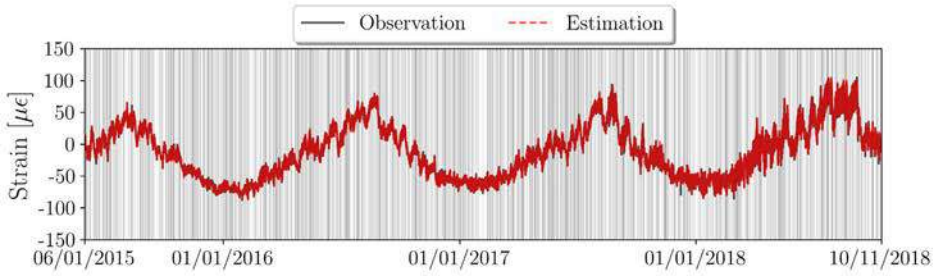
(a) Random missing scenario



(b) Structured missing scenario



(c) Mixed missing scenario (Case 1)



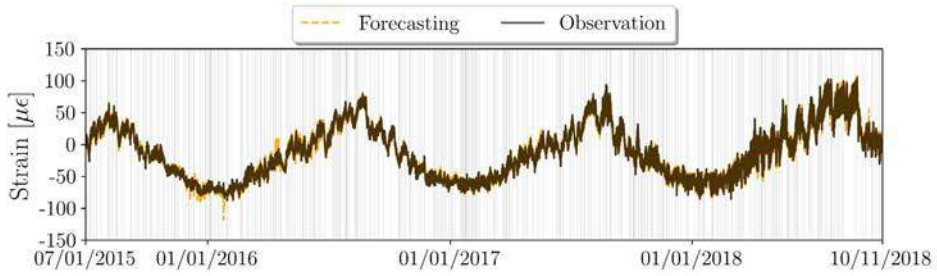
(d) Mixed missing scenario (Case 2)

Fig. 7. The imputation result for four missing cases of Sensor $S_2 - 4$. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The dataset ranges from June 1, 2015 to October 11, 2018 including 41 months.

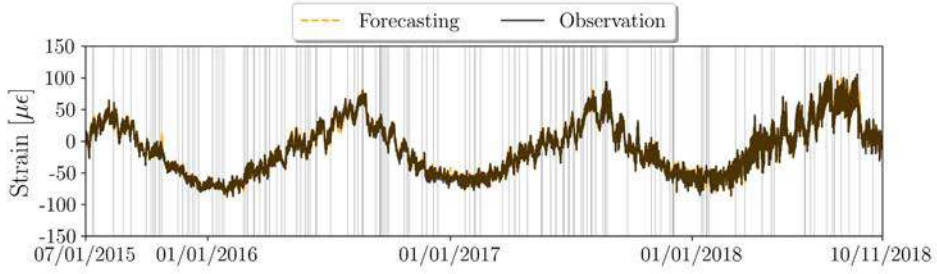
ground truth. Moreover, the forecasting uncertainty tends to be more unstable and larger when the missing rate becomes larger as shown in Fig. 12(g)–(i) and (j)–(l).

3.4. Missing rate effect

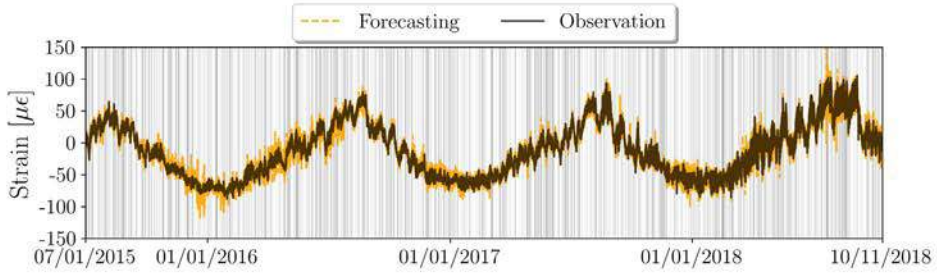
In addition, we perform parametric studies on the influence of data missing rate on the accuracy of imputation and forecasting. The test experiments arrange the first 80% portion of the recorded data for missing data recovery and the rest 20% for



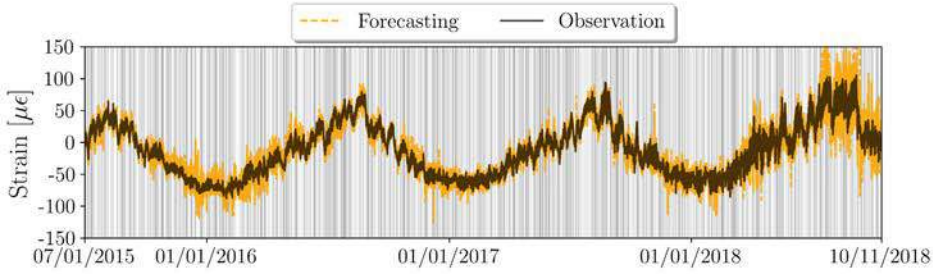
(a) Random missing scenario



(b) Structured missing scenario



(c) Mixed missing scenario (Case 1)



(d) Mixed missing scenario (Case 2)

Fig. 8. The forecasting result for four missing cases of Sensor $S_2 - 4$. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The forecasting dataset is from July 1, 2015 to October 11, 2018 with 40 months.

response forecasting. Here, the rank K is defined as eight while the time lags $\mathcal{L}$ is $\{1, 2, 144, 145\}$. In total, we run 200 iterations for MCMC sampling, where the first 100 are taken as burn-in and the last 100 are used to estimate the missing entries. Fig. 13 summarizes the parametric study result. For the random missing scenario shown in Fig. 13(a), the proposed method presents outstanding accuracy (over 95.0% for both imputation and forecasting) given the missing rate η up to 70%. The extreme case we consider here is the missing condition with $\eta = 80\%$. Nevertheless, the proposed approach still achieves over 86.0% missing data recovery accuracy and more than 92.0% forecasting accuracy. For the structured missing scenario

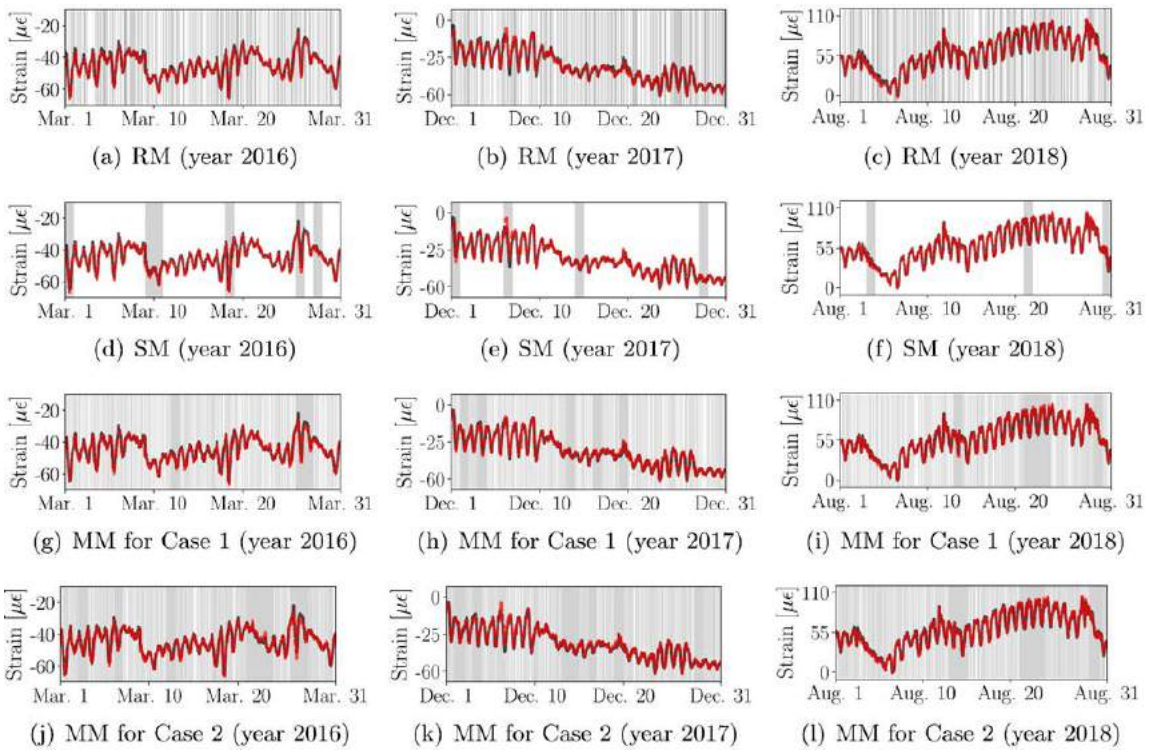


Fig. 9. The zoomed view of the imputed strain time series in Fig. 7. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The black lines and the red dashed lines depict the one-month field measurement and the imputed time histories, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(more practical and commonly seen in real-world applications), it is naturally more challenging to recover the missing data and forecast the response compared with the ideal random missing. As is seen in Fig. 13(b), the capacity limit of the proposed Bayesian tensor learning method shows to be $\eta = 40\%$ where the imputation accuracy surpasses 91.0% while the forecasting has over 88.0% accuracy. Interestingly, the mixed missing scenarios demonstrate quite perfect imputation and forecasting accuracy, namely, 99.8% for imputation and 98.4% for forecasting in Case 1% for imputation and 98.0% for forecasting in Case 2. This result is closely related to optimal tensor rank selection which is discussed in Section 3.5.

3.5. Rank analysis

There exist many recent researches attempting to reveal the effect of imperfect data on tensor representation [77–80]. According to [79], it is believed that clean datasets exhibit correlations across time and modalities while the imperfect data with incomplete values break these natural correlations and lead to the requirement of a higher rank. Inspired by this study, we also quantitatively investigate the prediction performance of the proposed Bayesian tensor learning method with different ranks (e.g., 4, 8 and 12) under different data missing scenarios (e.g., random and structured), and summarize the result in Fig. 14. In particular, we test the imputation capability. As shown in Fig. 14(a), with the increasing rank, the estimation achieves a better accuracy in the random missing scenario, which agree with the observation in [79]. In other words, random missing destroys the spatiotemporal correlations so that we should increase the tensor rank for a more accurate result when dealing with this type of imperfect data condition.

For the structured missing scenario (see Fig. 14(b)), it is surprising to see that we get a higher accuracy of missing data recovery with a lower rank under different missing rates. We empirically extrapolate that continuous element missing helps to build a more correlated tensor structure.

3.6. Comparison with robust principle component analysis

We also compare the performance of data imputation between the proposed method and the Robust Principle Component Analysis (RPCA) approach [81] which has also been implemented for missing data recovery in SHM [25]. The dataset considered herein contains one year (360 days) observations starting from June 1, 2015. We conduct the comparison exper-

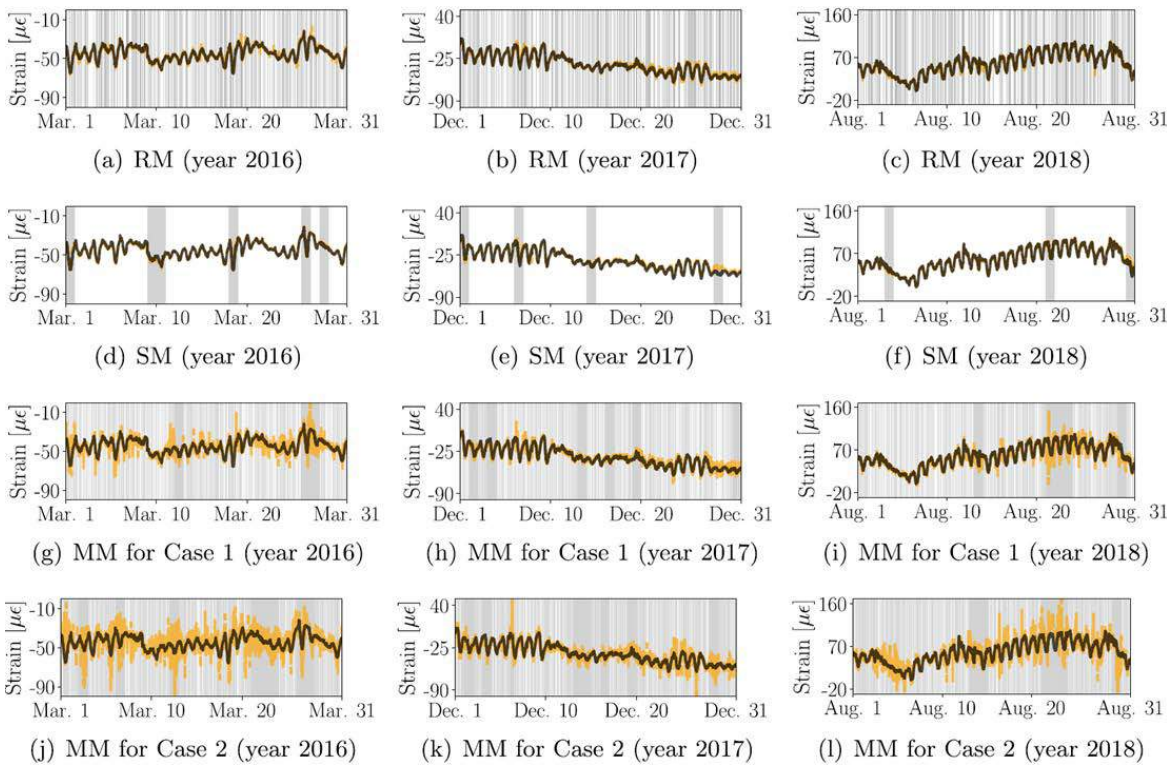


Fig. 10. The zoomed view of the predicted strain time series in Fig. 8. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The black lines and the orange dashed lines depict the one-month field measurement and the forecasted time histories, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

iments on four missing scenarios, including 10% random missing, 10% structured missing, 20% structured missing and the mixed missing with 30% random missing and 20% structured missing. Instead of directly applying RPCA method, we follow the idea of reshaping the dimension of time-series matrix [25] which helps to greatly improve the recovery accuracy for SHM data, e.g., 800×1296. The model parameters in our method are identical to those used in subSection 3.4.

For the random missing scenario, both RPCA and our method achieve high imputation accuracy, 99.2% and 99.7%, respectively, showing excellent performance. However, for the case of structured missing, RPCA fails to capture the dynamic patterns of the imperfect time series data with the accuracy of 76.7% ($\eta = 10\%$) and 60.4% ($\eta = 20\%$), while our method still retains satisfactory accuracy of 98.6% ($\eta = 10\%$) and 95.0% ($\eta = 20\%$). For the mixed missing scenario, both methods exhibit better imputation performance compared with the purely structured missing scenarios. Meanwhile, the accuracy of our method is 99.6%, which is higher than that of RPCA (83.0%). Generally, RPCA is also a matrix decomposition method, which shares similar characteristics with our method where imputation performance is closely related to the approximated rank. The mixed missing scenario surprisingly helps to form a good condition for optimal rank selection of matrix decomposition methods. The imputation results under four missing scenarios are presented in Fig. 15, where a typical month is selected for illustration. In general, our method outperforms the RPCA approach, especially for the imperfect time series with structured missing.

3.7. Data imputation with high-order tensor

To validate the feasibility of the proposed high-order tensor learning, we reshape the sensor data to a third-order tensor $\mathbf{Y}$ with the size of $20 \times 1{,}232 \times 144$, which represents "sensors $\times$ days $\times$ time stamps per day". We set 10% missing rate for both the random and structured missing scenarios. For the model parameters, the rank is set as 8 and the time lags is assumed to be $\{1, 2, 3, 4, 5, 6, 7\}$. The proposed Bayesian tensor decomposition method could achieve 99.9% and 98.6% recovery accuracy for the two missing scenarios, which demonstrates the satisfactory viability of the high-order tensor learning. The overall imputation results are shown in Fig. 16. Nevertheless, it is noteworthy that the above tensor format is inapplicable to response forecasting since the reshaped data format breaks the long-term temporal continuity. Due to the insufficient number of sensors considered in this example, it is less proper to construct the tensor in the form of "transverse sensors $\times$ longitudinal sensors $\times$ time stamps". Such a particular tensor form is more applicable to the practice where measurements from a grid of sensors are available. This will be tested in our future study.

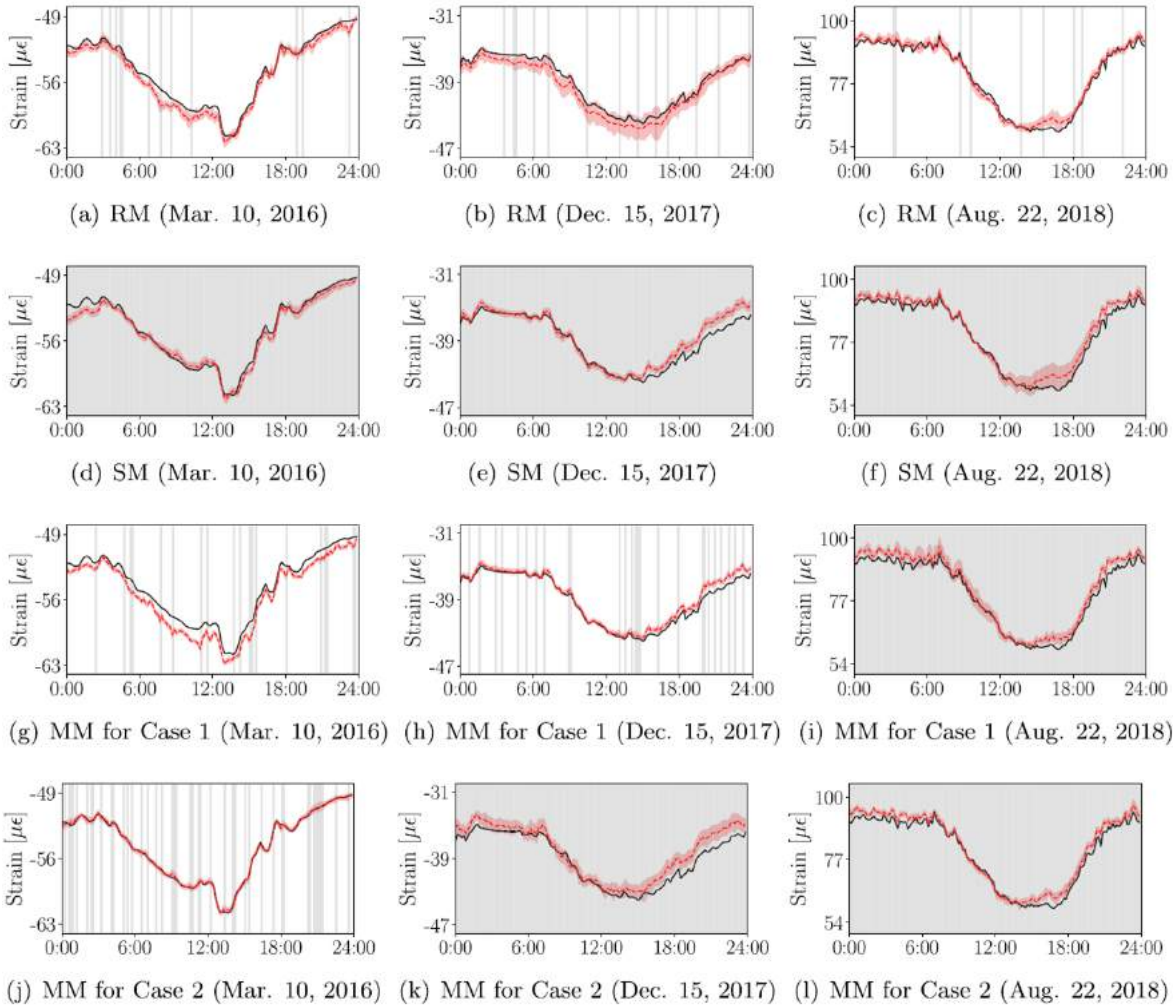


Fig. 11. Uncertainty quantification of imputation for four data missing cases. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The black lines and the red dashed lines depict the one-day field measurement and the imputed means respectively, and the red band is the area between plus/minus three standard deviations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions and discussions

This paper presents an incremental Bayesian matrix/tensor learning method for spatiotemporal data imputation and response forecasting for SHM applications, with the incorporation of the AR process which contributes to the temporal feature modeling in an incremental learning scheme. With the existence of temperature data, the tensor model can easily gain a low-rank structure and utilize the correlation between strain and temperature for robust prediction of the strain response. In our validation experiments, we both consider the ideal random missing scenario and a more realistic missing condition – structured missing. Based on the learned latent features, the accurate estimation and forecasting results show the satisfactory performance of the proposed approach for processing incomplete SHM recordings, with uncertainty quantification capability. In addition, the extreme cases illustrate that acceptable imputation and forecasting accuracy can retain for the missing rate up to $\eta = 80\%$ in random missing and up to $\eta = 40\%$ in structured missing. Furthermore, the investigation into rank selection has revealed that a lower rank helps achieve better prediction performance for structured missing, while a higher rank is preferred for random missing. Lastly, we extend the Bayesian matrix factorization to high-order tensor learning for SHM data imputation, showing satisfactory performance.

There are three highlights of the proposed method. The first and the most notable significance is that we model the temporal dependency via the latent features instead of using incomplete data directly, which offers a robust and flexible modeling scheme for multivariate time series data. Secondly, it is unnecessary to know which of the entries in the tensor data are incomplete beforehand. Thirdly, the fully Bayesian method can avoid overfitting and relax parameter tuning. However, it poses high computational complexity due to the use of MCMC sampling. Notwithstanding the most time consuming process

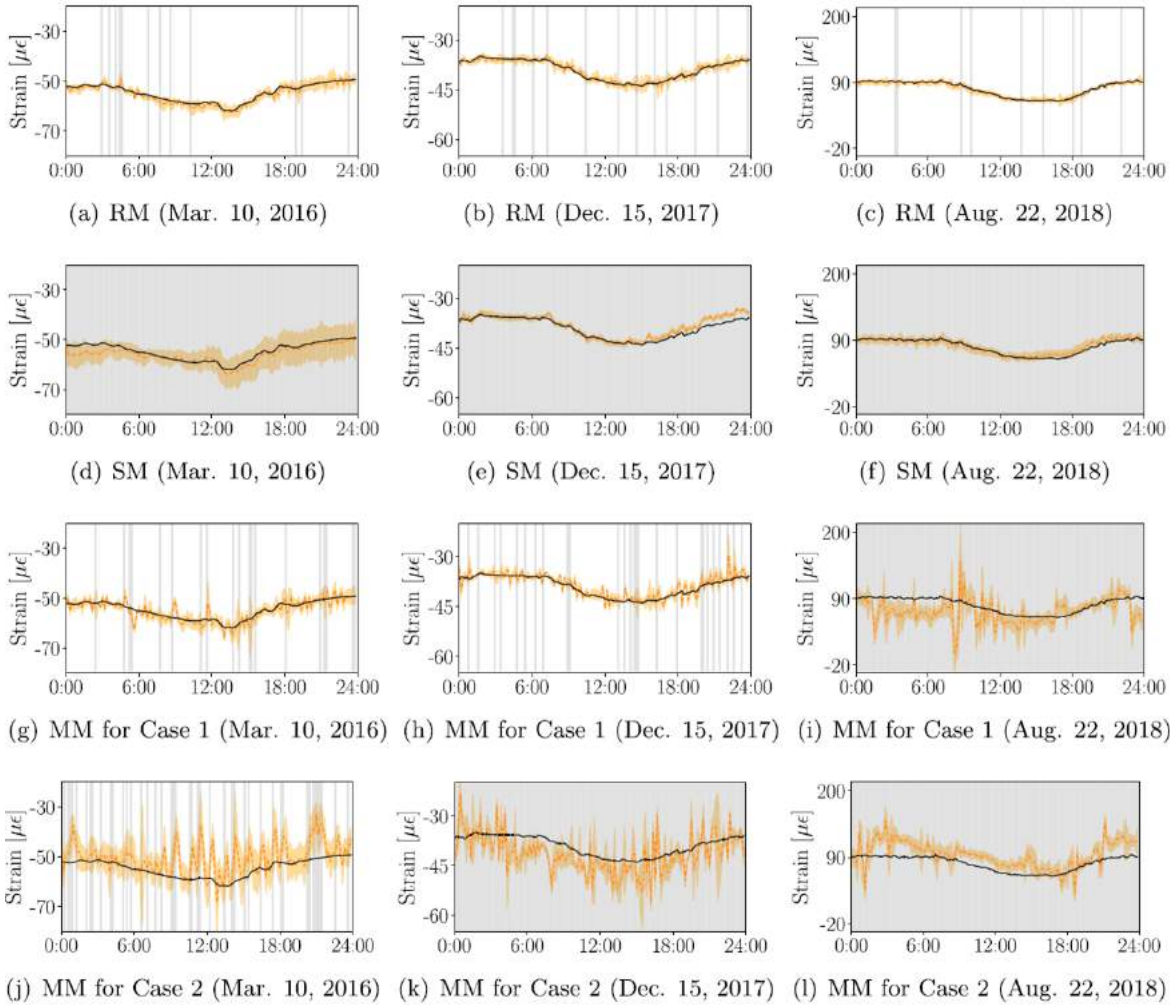


Fig. 12. Uncertainty quantification of forecasting for four data missing cases. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The black lines and the orange dashed lines depict the one-day field measurement and the predicted means respectively, and the yellow band is the area between plus/minus three standard deviations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

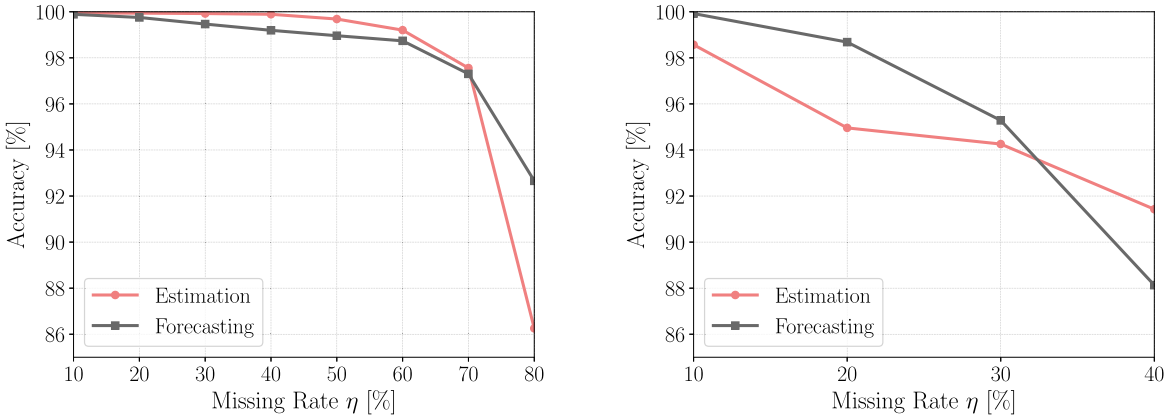
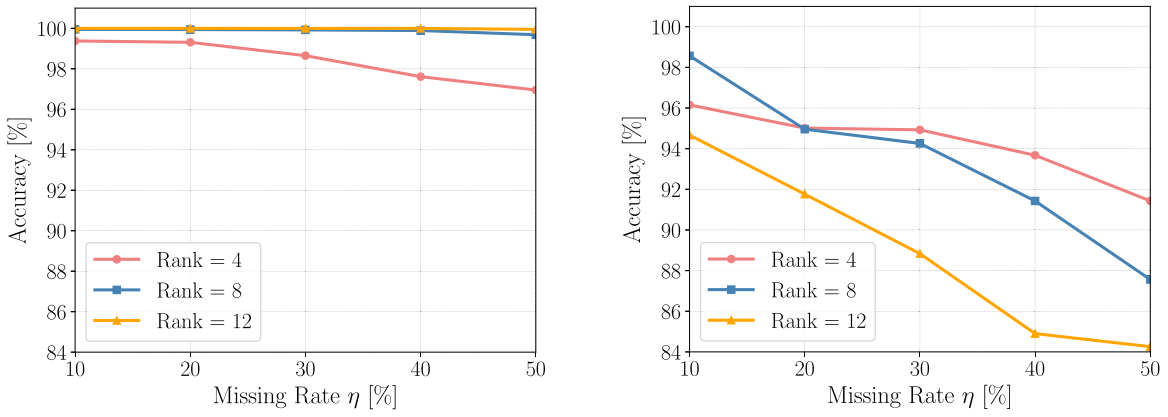
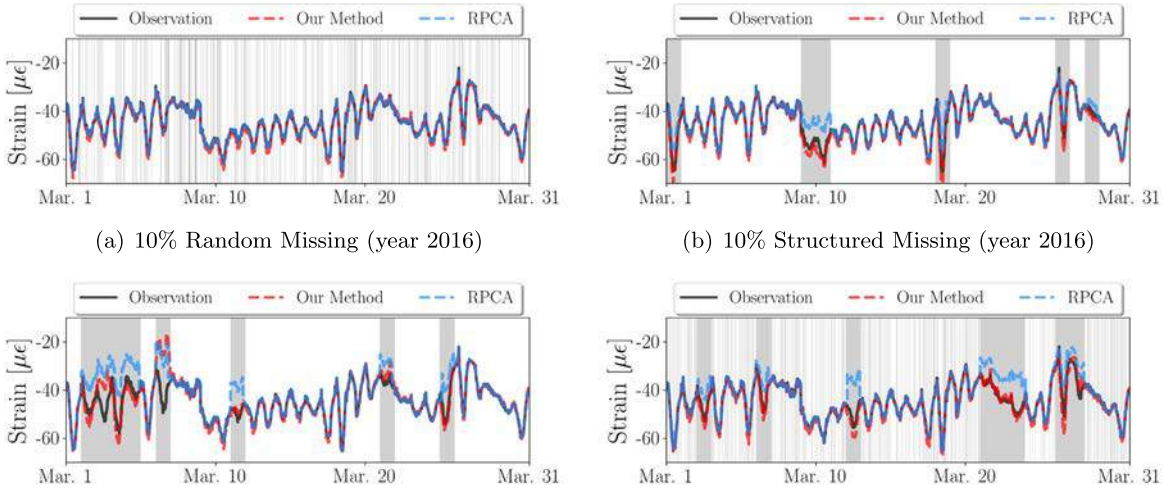


Fig. 13. The accuracy of imputation and forecasting with respect to different data missing rates.



(a) Random Missing (b) Structured Missing

Fig. 14. The performance imputation with respect to different tensor ranks.



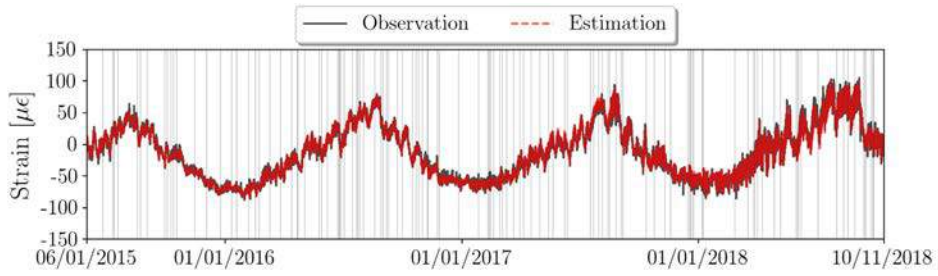
(a) 10% Random Missing (year 2016) (b) 10% Structured Missing (year 2016)

(c) 20% Structured Missing (year 2016) (d) 30% Random + 20% Structured Missing (year 2016)

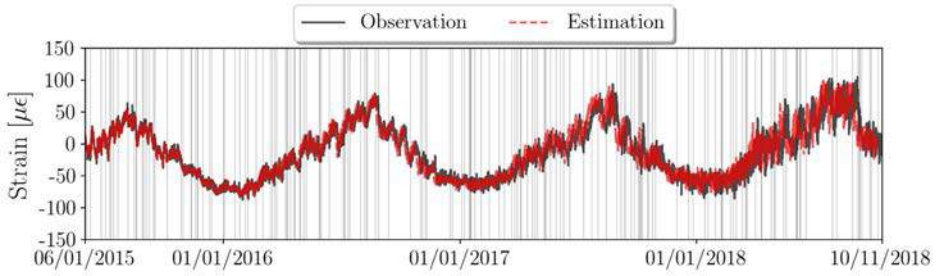
Fig. 15. Comparison between the proposed Bayesian matrix learning approach and RPCA for four missing scenarios. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The black lines denote one-month observations. The red dashed lines and blue dashed lines represent the imputation results by our method and RPCA respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

remains in the imputation process, our proposed incremental Bayesian tensor learning algorithm can drastically reduce the computational time and make it efficient for data imputation and response forecasting for continuous SHM with streaming yet missing data.

The present study demonstrates that tensor learning has potential to become a promising area in SHM applications. Some future research directions and outlook are proposed herein. Firstly, as long as we have enough sensor locations and monitoring zones (e.g., distributed sensing), the higher order tensor decomposition for imputation and forecasting should be explored more thanks to its possibility of outperforming the second-order tensor factorization [82]. Secondly, the proposed approach can be extended to tackle issues of SHM data anomaly detection and de-noising on account of the power of tensor representation. Thirdly, the variational Bayesian inference (VBI) [83], as a faster and more scalable inference alternative, should gain more attention for Bayesian tensor learning, especially when dealing with large datasets. Last but not least, the spatial feature can be described in a more realistic way by considering graph kernels [53,84], which will be worthy to investigate.



(a) Random missing scenario



(b) Structured missing scenario

Fig. 16. The imputation result for two missing cases of Sensor $S_2 - 4$ using a third-order tensor. Note that the shading areas represent the time periods where data missing occurs, while the white box areas denote that the strain time series are successfully recorded. The dataset ranges from June 1, 2015 to October 11, 2018 including 41 months.

CRediT authorship contribution statement

Pu Ren: Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft. **Xinyu Chen:** Methodology. **Lijun Sun:** Conceptualization, Methodology. **Hao Sun:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the Department of Bridge and Structural Engineering, China Merchants Chongqing Communications Technology Research and Design Institute Co. Ltd., for sharing the datasets which were used to validate the proposed methodology. In addition, the authors greatly acknowledge the open source codes [85] of Bayesian temporal matrix factorization (BTMF), which were leveraged for numerical analyses in this study.

References

[1] Y. Bao, H. Li, X. Sun, Y. Yu, J. Ou, Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring, Structural Health Monitoring 12 (1) (2013) 78–95.
[2] Y. Bao, Y. Yu, H. Li, X. Mao, W. Jiao, Z. Zou, J. Ou, Compressive sensing-based lost data recovery of fast-moving wireless sensing for structural health monitoring, Structural Control and Health Monitoring 22 (3) (2015) 433–448.
[3] Y. Huang, J.L. Beck, S. Wu, H. Li, Robust bayesian compressive sensing for signals in structural health monitoring, Computer-Aided Civil and Infrastructure Engineering 29 (3) (2014) 160–179.
[4] Y. Bao, Z. Tang, H. Li, Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach, Structural Health Monitoring 19 (1) (2020) 293–304.
[5] H. Cao, Y. Tian, J. Lei, X. Tan, D. Gao, F. Kopsaftopoulos, F. Chang, Deformation data recovery based on compressed sensing in bridge structural health monitoring, Structural Health Monitoring 2017 (shm).
[6] K.-V. Yuen, H.-Q. Mu, A novel probabilistic method for robust parametric identification and outlier detection, Probabilistic Engineering Mechanics 30 (2012) 48–59.

[7] K.-V. Yuen, G.A. Ortiz, Outlier detection and robust regression for correlated data, Computer Methods in Applied Mechanics and Engineering 313 (2017) 632–646.
[8] H. Sun, D. Feng, Y. Liu, M.Q. Feng, Statistical regularization for identification of structural parameters and external loadings using state space models, Computer-Aided Civil and Infrastructure Engineering 30 (11) (2015) 843–858.
[9] H. Sun, R. Betti, A hybrid optimization algorithm with bayesian inference for probabilistic model updating, Computer-Aided Civil and Infrastructure Engineering 30 (8) (2015) 602–619.
[10] H. Sun, O. Büyüköztürk, Bayesian model updating using incomplete modal data without mode matching, in: Health Monitoring of Structural and Biological Systems 2016, vol. 9805, International Society for Optics and Photonics, 2016, p. 98050D.
[11] I. Behmanesh, S. Yousefianmoghadam, A. Nozari, B. Moaveni, A. Stavridis, Effects of prediction error bias on model calibration and response prediction of a 10-story building, in: Model Validation and Uncertainty Quantification, vol. 3, Springer, 2016, pp. 279–291.
[12] H. Sun, A. Mordret, G.A. Prieto, M.N. Toksöz, O. Büyüköztürk, Bayesian characterization of buildings using seismic interferometry on ambient vibrations, Mechanical Systems and Signal Processing 85 (2017) 468–486.
[13] M. Song, I. Behmanesh, B. Moaveni, C. Papadimitriou, Modeling error estimation and response prediction of a 10-story building model through a hierarchical bayesian model updating framework, Frontiers in Built Environment 5 (2019) 7.
[14] M. Song, I. Behmanesh, B. Moaveni, C. Papadimitriou, Hierarchical bayesian calibration and response prediction of a 10-story building model, in: Model Validation and Uncertainty Quantification, vol. 3, Springer, 2019, pp. 153–165.
[15] M. Uzun, H. Sun, D. Smit, O. Büyüköztürk, Structural damage detection using bayesian inference and seismic interferometry, Structural Control and Health Monitoring 26 (11) (2019) e2445.
[16] Z. Chen, R. Zhang, J. Zheng, H. Sun, Sparse bayesian learning for structural damage identification, Mechanical Systems and Signal Processing 140 (2020) 106689.
[17] T. Yin, Q.-H. Jiang, K.-V. Yuen, Vibration-based damage detection for structural connections using incomplete modal data by bayesian approach and model reduction technique, Engineering Structures 132 (2017) 260–277.
[18] L.D. Avendaño-Valencia, E.N. Chatzi, K.Y. Koo, J.M. Brownjohn, Gaussian process time-series models for structures under operational variability, Frontiers in Built Environment 3 (2017) 69.
[19] L.D. Avendaño-Valencia, K. Tatsis, E.N. Chatzi, Gaussian process vector ar surrogates for identification of structures under varying operational conditions, in: Proceedings of the 8th Conference on Computational Stochastic Mechanics, 2018, pp. 1–12.
[20] L.D. Avendaño-Valencia, E.N. Chatzi, Modelling long-term vibration monitoring data with gaussian process time-series models, IFAC-PapersOnLine 52 (28) (2019) 26–31.
[21] F. Kopsaftopoulos, S. Fassois, A functional model based statistical time series method for vibration based damage detection, localization, and magnitude estimation, Mechanical Systems and Signal Processing 39 (1–2) (2013) 143–161.
[22] A. Amer, F. Kopsaftopoulos, Probabilistic damage quantification via the integration of non-parametric time-series and gaussian process regression models, Structural Health Monitoring (2019).
[23] H.-P. Wan, Y.-Q. Ni, Bayesian multi-task learning methodology for reconstruction of structural health monitoring data, Structural Health Monitoring 18 (4) (2019) 1282–1309.
[24] Z. Chen, Y. Bao, H. Li, B.F. Spencer Jr, A novel distribution regression approach for data loss compensation in structural health monitoring, Structural Health Monitoring 17 (6) (2018) 1473–1490.
[25] Y. Yang, S. Nagarajaiah, Harnessing data structure for recovery of randomly missing structural vibration responses time history: Sparse representation versus low-rank structure, Mechanical Systems and Signal Processing 74 (2016) 165–182.
[26] Z. Chen, H. Li, Y. Bao, Analyzing and modeling inter-sensor relationships for strain monitoring data and missing data imputation: a copula and functional data-analytic approach, Structural Health Monitoring 18 (4) (2019) 1168–1188.
[27] S.-C. Kuok, K.-V. Yuen, Model-free data reconstruction of structural response and excitation via sequential broad learning, Mechanical Systems and Signal Processing 141 (2020) 106738.
[28] X. Fan, Bridge extreme stress prediction based on bayesian dynamic linear models and non-uniform sampling, Structural Health Monitoring 16 (3) (2017) 253–261.
[29] J.-A. Goulet, Bayesian dynamic linear models for structural health monitoring, Structural Control and Health Monitoring 24 (12) (2017) e2035.
[30] H. Wang, Y.-M. Zhang, J.-X. Mao, H.-P. Wan, T.-Y. Tao, Q.-X. Zhu, Modeling and forecasting of temperature-induced strain of a long-span bridge using an improved bayesian dynamic linear model, Engineering Structures 192 (2019) 220–232.
[31] G. Park, D.J. Inman, Structural health monitoring using piezoelectric impedance measurements, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365 (1851) (2007) 373–392.
[32] L. Bornn, C.R. Farrar, G. Park, K. Farinholt, Structural health monitoring with autoregressive support vector machines, Journal of Vibration and Acoustics 131 (2).
[33] N.M. Okasha, D.M. Frangopol, D. Saydam, L.W. Salvino, Reliability analysis and damage detection in high-speed naval craft based on structural health monitoring data, Structural Health Monitoring 10 (4) (2011) 361–379.
[34] H.-P. Wan, Y.-Q. Ni, Bayesian modeling approach for forecast of structural stress response using structural health monitoring data, Journal of Structural Engineering 144 (9) (2018) 04018130.
[35] R.-T. Wu, M.R. Jahanshahi, Deep convolutional neural network for structural dynamic response estimation and system identification, Journal of Engineering Mechanics 145 (1) (2018) 04018125.
[36] G. Fan, J. Li, H. Hao, Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks, Structural Health Monitoring 1475921720916881 (2020).
[37] R. Zhang, Y. Liu, H. Sun, Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling, Engineering Structures 215 (2020) 110704.
[38] R. Zhang, Z. Chen, S. Chen, J. Zheng, O. Büyüköztürk, H. Sun, Deep long short-term memory networks for nonlinear structural seismic response prediction, Computers & Structures 220 (2019) 55–68.
[39] R. Zhang, Y. Liu, H. Sun, Physics-informed multi-lstm networks for metamodeling of nonlinear structures, Computer Methods in Applied Mechanics and Engineering 369 (2020) 113226.
[40] C. Mylonas, I. Abdallah, E. Chatzi, Deep unsupervised learning for condition monitoring and prediction of high dimensional data with application on windfarm scada data, in: Model Validation and Uncertainty Quantification, vol. 3, Springer, 2020, pp. 189–196.
[41] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.
[42] A. Anandkumar, R. Ge, D. Hsu, S.M. Kakade, M. Telgarsky, Tensor decompositions for learning latent variable models, Journal of Machine Learning Research 15 (2014) 2773–2832.
[43] M. Janzamin, R. Ge, J. Kossaifi, A. Anandkumar, Spectral learning on matrices and tensors, arXiv preprint arXiv:2004.07984.
[44] Q. Zhao, L. Zhang, A. Cichocki, Bayesian cp factorization of incomplete tensors with automatic rank determination, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (9) (2015) 1751–1763.
[45] L. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Compression of hyperspectral remote sensing images by tensor approach, Neurocomputing 147 (2015) 358–363.
[46] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, S. Yan, Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5249–5257.
[47] B. Du, M. Zhang, L. Zhang, R. Hu, D. Tao, Pltd: Patch-based low-rank tensor decomposition for hyperspectral images, IEEE Transactions on Multimedia 19 (1) (2016) 67–79.

[48] X. Zhang, G. Wen, W. Dai, A tensor decomposition-based anomaly detection algorithm for hyperspectral image, IEEE Transactions on Geoscience and Remote Sensing 54 (10) (2016) 5801–5820.
[49] Q. Shi, Y.-M. Cheung, Q. Zhao, H. Lu, Feature extraction for incomplete data via low-rank tensor decomposition with feature regularization, IEEE Transactions on Neural Networks and Learning Systems 30 (6) (2018) 1803–1817.
[50] A. Karatzoglou, X. Amatriain, L. Baltrunas, N. Oliver, Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering, in: Proceedings of the Fourth ACM Conference on Recommender Systems, 2010, pp. 79–86.
[51] N. Ifada, R. Nayak, Tensor-based item recommendation using probabilistic ranking in social tagging systems, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 805–810.
[52] A. Seko, H. Hayashi, H. Kashima, I. Tanaka, Matrix-and tensor-based recommender systems for the discovery of currently unknown inorganic compounds, Physical Review Materials 2 (1) (2018) 013805.
[53] H.-F. Yu, N. Rao, I.S. Dhillon, Temporal regularized matrix factorization for high-dimensional time series prediction, in: Advances in Neural Information Processing Systems, 2016, pp. 847–855.
[54] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, F. Li, A tensor-based method for missing traffic data completion, Transportation Research Part C: Emerging Technologies 28 (2013) 15–27.
[55] M.T. Asif, N. Mitrovic, J. Dauwels, P. Jaillet, Matrix and tensor based methods for missing data estimation in large traffic networks, IEEE Transactions on Intelligent Transportation Systems 17 (7) (2016) 1816–1825.
[56] H. Tan, Y. Wu, B. Shen, P.J. Jin, B. Ran, Short-term traffic prediction based on dynamic tensor completion, IEEE Transactions on Intelligent Transportation Systems 17 (8) (2016) 2123–2133.
[57] K. Takeuchi, H. Kashima, N. Ueda, Autoregressive tensor factorization for spatio-temporal predictions, in: 2017 IEEE International Conference on Data Mining (ICDM), IEEE, 2017, pp. 1105–1110.
[58] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, Y. Liu, Latent space model for road networks to predict time-varying traffic, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1525–1534.
[59] X. Chen, Z. He, J. Wang, Spatial-temporal traffic speed patterns discovery and incomplete data recovery via svd-combined tensor decomposition, Transportation Research Part C: Emerging Technologies 86 (2018) 59–77.
[60] X. Chen, Z. He, L. Sun, A bayesian tensor decomposition approach for spatiotemporal traffic data imputation, Transportation Research Part C: Emerging Technologies 98 (2019) 73–84.
[61] X. Chen, L. Sun, Bayesian temporal factorization for multidimensional time series prediction, arXiv preprint arXiv:1910.06366.
[62] X. Chen, J. Yang, L. Sun, A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation, Transportation Research Part C: Emerging Technologies 117 (2020) 102673.
[63] X. Chen, L. Sun, Low-rank autoregressive tensor completion for multivariate time series forecasting, arXiv preprint arXiv:2006.10436.
[64] Y. Xu, B. Chen, C. Ng, K. Wong, W. Chan, Monitoring temperature effect on a long suspension bridge, Structural Control and Health Monitoring 17 (6) (2010) 632–653.
[65] Y.-F. Duan, Y. Li, Y.-Q. Xiang, Strain-temperature correlation analysis of a tied arch bridge using monitoring data, in: 2011 International Conference on Multimedia Technology, IEEE, 2011, pp. 6025–6028.
[66] Q. Xia, Y. Cheng, J. Zhang, F. Zhu, In-service condition assessment of a long-span suspension bridge using temperature-induced strain data, Journal of Bridge Engineering 22 (3) (2017) 04016124.
[67] Y. Zhu, Y.-Q. Ni, A. Jesus, J. Liu, I. Laory, Thermal strain extraction methodologies for bridge structural condition assessment, Smart Materials and Structures 27 (10) (2018) 105051.
[68] P. Rai, Y. Wang, S. Guo, G. Chen, D. Dunson, L. Carin, Scalable bayesian low-rank decomposition of incomplete multiway tensors, in: International Conference on Machine Learning, 2014, pp. 1800–1808.
[69] K. Fujimoto, A. Satoh, S. Fukunaga, System identification based on variational bayes method and the invariance under coordinate transformations, in: 2011 50th IEEE Conference on Decision and Control and European Control Conference, IEEE, 2011, pp. 3882–3888.
[70] B. Li, A. Der Kiureghian, S.-K. Au, A gibbs sampling algorithm for structural modal identification under seismic excitation, Earthquake Engineering & Structural Dynamics 47 (14) (2018) 2735–2755.
[71] R.E. Kalman, A new approach to linear filtering and prediction problems.
[72] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, Springer Science & Business Media, 2006.
[73] G. Petris, S. Petrone, P. Campagnoli, Dynamic linear models, in: Dynamic Linear Models with R, Springer, 2009, pp. 31–84.
[74] R. Salakhutdinov, A. Mnih, Bayesian probabilistic matrix factorization using markov chain monte carlo, in: Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 880–887.
[75] C. Andrieu, A. Doucet, R. Holenstein, Particle markov chain monte carlo methods, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (3) (2010) 269–342.
[76] W.R. Gilks, P. Wild, Adaptive rejection sampling for gibbs sampling, Journal of the Royal Statistical Society: Series C (Applied Statistics) 41 (2) (1992) 337–348.
[77] H. Fan, Y. Chen, Y. Guo, H. Zhang, G. Kuang, Hyperspectral image restoration using low-rank tensor recovery, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (10) (2017) 4589–4604.
[78] X. Chen, Z. Han, Y. Wang, Q. Zhao, D. Meng, L. Lin, Y. Tang, A general model for robust tensor factorization with unknown noise, arXiv preprint arXiv:1705.06755.
[79] P.P. Liang, Z. Liu, Y.-H.H. Tsai, Q. Zhao, R. Salakhutdinov, L.-P. Morency, Learning representations from imperfect time series data via tensor rank regularization, arXiv preprint arXiv:1907.01011.
[80] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, S. Zhong, Weighted low-rank tensor recovery for hyperspectral image restoration, IEEE Transactions on Cybernetics.
[81] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, Journal of the ACM (JACM) 58 (3) (2011) 1–37
[82] B. Ran, H. Tan, Y. Wu, P.J. Jin, Tensor based missing traffic data completion with spatial–temporal correlation, Physica A: Statistical Mechanics and its Applications 446 (2016) 54–63.
[83] J. Paisley, D.M. Blei, M.I. Jordan, Variational bayesian inference with stochastic search, in: Proceedings of the 29th International Conference on Machine Learning, 2012, pp. 1363–1370.
[84] M.T. Bahadori, Q.R. Yu, Y. Liu, Fast multivariate spatio-temporal analysis via low rank tensor learning, in: Advances in Neural Information Processing Systems, 2014, pp. 3491–3499.
[85] https://github.com/xinychen/transdim.