

Discovering Dynamic Patterns from Spatiotemporal Data with Time-Varying Low-Rank Autoregression

Xinyu Chen[§], *Student Member, IEEE*, Chengyuan Zhang[§], *Student Member, IEEE*, Xiaoxu Chen, Nicolas Saunier, and Lijun Sun, *Senior Member, IEEE*

Abstract—The problem of discovering interpretable dynamic patterns from spatiotemporal data is studied in this paper. For that purpose, we develop a time-varying reduced-rank vector autoregression (VAR) model whose coefficient matrices are parameterized by low-rank tensor factorization. Benefiting from the tensor factorization structure, the proposed model can simultaneously achieve model compression and pattern discovery. In particular, the proposed model allows one to characterize nonstationarity and time-varying system behaviors underlying spatiotemporal data. To evaluate the proposed model, extensive experiments are conducted on various spatiotemporal datasets representing different nonlinear dynamical systems, including fluid dynamics, sea surface temperature, USA surface temperature, and NYC taxi trips. Experimental results demonstrate the effectiveness of modeling spatiotemporal data and characterizing spatial/temporal patterns with the proposed model. In the spatial context, the spatial patterns can be automatically extracted and intuitively characterized by the spatial modes. In the temporal context, the complex time-varying system behaviors can be revealed by the temporal modes in the proposed model. Thus, our model lays an insightful foundation for understanding complex spatiotemporal data in real-world dynamical systems. The adapted datasets and Python implementation are publicly available at <https://github.com/xinyuchen/vars>.

Index Terms—Spatiotemporal data, time-varying system, vector autoregression, tensor factorization, pattern discovery.

1 INTRODUCTION

DYNAMIC mechanisms that drive nonlinear systems are universally complex and obscure. Straightforwardly, one can investigate the behavior of a system by discovering the patterns from its observations. In practice, when we take observations from a real-world complex system, spatiotemporal data are one of the most widely encountered form relating to space and time and showing the characteristics of time series. With the remarkable development of sensing technologies, tremendous amounts of spatiotemporal data are accessible for analysis, such as satellite data [1], station-level weather observation data [2], and trajectories of human mobility [3], to name but a few. In the literature, various types of spatiotemporal data were characterized by time series models and data-driven models. Leveraging time series models not only allows one to analyze spatiotemporal data but also makes it possible to discover inherent spatial and temporal patterns from the data over space and time.

- Xinyu Chen and Nicolas Saunier are with the Civil, Geological and Mining Engineering Department, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada. E-mail: chenxy346@gmail.com (Xinyu Chen), nicolas.saunier@polymtl.ca (Nicolas Saunier).
- Chengyuan Zhang, Xiaoxu Chen, and Lijun Sun are with the Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada. E-mail: enzozcy@gmail.com (Chengyuan Zhang), xiaoxu.chen@mail.mcgill.ca (Xiaoxu Chen), lijun.sun@mcgill.ca (Lijun Sun).
- [§]Xinyu Chen and Chengyuan Zhang contributed equally to this work.

(Corresponding author: Lijun Sun)

As a simple yet efficient and classical statistical model, vector autoregression (VAR) explicitly finds the linear relationship among a sequence of time series changing over time, which can also successfully describe the dynamic behavior of time series [4], [5]. Formally, if the given time series $s_1, \dots, s_T \in \mathbb{R}^N$ (i.e., a collection of observations with N variables at consecutive times) are stationary, then the d th-order VAR takes a linear formulation as $s_t = \sum_{k=1}^d A_k s_{t-k} + \epsilon_t, \forall t \in \{d+1, \dots, T\}$, in which the coefficient matrices $A_1, \dots, A_d \in \mathbb{R}^{N \times N}$ are expected to capture the temporal correlations of the multivariate time series, and ϵ_t is the residual at time t . A large body of VAR and its variants such as reduced-rank VAR [6], [7], [8], dynamic mode decomposition (DMD) [9], [10], [11], high-dimensional VAR [12], and time-varying VAR with tensor factorization [13] have been developed for analyzing real-world time series data. Essentially, the modeling intuition of these data-driven approaches is that the coefficient matrix in VAR is rank-deficient, or at least the dominant patterns underlying the coefficient matrix can be revealed by a low-rank structure.

DMD models—the classical variants of VAR—were initially developed for discovering spatiotemporal coherent patterns from fluid flow mechanism, and they are available for interpreting the system behavior and underlying data patterns, giving credit to its low-rank structure (i.e., a reduced-rank linear system of VAR) for dimensionality reduction [9], [10], [11]. However, stationary data are often required when analyzing multivariate time series via

VAR and its variants, which is against the fact that real-world data are usually nonstationary. Therefore, one great challenge of modeling time series with VAR is identifying the time-varying system behaviors in the analysis, which is often associated with the nonstationarity issue. Although the nonstationarity and time-varying system behaviors are pretty clear to verify, the problem of discovering underlying data patterns from time-varying systems is challenging and still demands further exploration. To characterize the underlying time-varying system behaviors hidden behind time series, a classical framework was proposed by utilizing the dynamic linear models (e.g., time-varying vector autoregression) whose latent variables vary over time [14], [15], [16].

Typically, time-varying VAR model takes a sequence of VAR processes at different times, and it is capable of handling the time-varying system behaviors. Nevertheless, the time-varying coefficients in the model give rise to the over-parameterization issue, which implies that the number of parameters (e.g., coefficient matrices in time-varying VAR) inevitably exceeds the number of observations. The over-parameterized coefficients can be regularized by a certain smoothing term from a machine learning perspective, but it does not work in the high-dimensional settings. Another efficient approach to address the over-parameterization issue is using low-rank models such as low-rank matrix/tensor factorization for pursuing perfect model compression. Notably, low-rank models have shown unprecedented compression power on both matrix and tensor data or parameters in several contemporary studies [17], [18], [19], [20], [21].

Complementary to the above methods, recent work provides an online DMD [22] and a time-varying autoregression [13] for modeling time-varying systems. The online DMD model allows one to compute DMD in real time and find the approximation of a system's dynamics for time-varying streaming data. However, as for DMD models, they are sensitive and vulnerable to the noises in the time series data, posing both methodological and practical challenges [23]. Therefore, it is hard to use DMD to discover data patterns and system behaviors from time-varying and noisy time series. In [13], the time-varying autoregression model has a sequence of windowed linear systems consisting of VAR processes, highly motivated by the need for time-varying system identification. It allows one to capture pattern changes of a dynamical system through tensor factorization. Typically, tensor factorization in the time-varying autoregression is capable of compressing the over-parameterized coefficients and identifying the time-varying system behaviors.

In this work, we revisit the classical spatiotemporal data analysis problem and introduce a novel method to discover spatial and temporal patterns from time-varying systems in a data-driven fashion. The cornerstone of this work is the time-varying VAR on nonstationary time series. Formally, we can formulate the first-order VAR with time-varying coefficients as $s_t = A_t s_{t-1} + \epsilon_t, \forall t$, in which the system matrices $\{A_t\}$ are changing over time. In this situation, the system matrices involve $\mathcal{O}(N^2(T - 1))$ parameters, which exceed $\mathcal{O}(NT)$ observations. To address the over-parameterization issue, we utilize a low-rank tensor fac-

torization structure—Tucker decomposition [19], [24]—to achieve the model compression. In the meanwhile, tensor factorization allows one to discover spatiotemporal patterns from spatiotemporal data.

The main contributions of this work are twofold:

- We propose a fully time-varying reduced-rank VAR model which allows one to discover interpretable dynamic modes from spatiotemporal data. In our model, we draw strong connections between time-varying VAR and tensor factorization (i.e., Tucker decomposition) that address the over-parameterization issue and reveal spatial and temporal patterns in the latent spaces.
- We demonstrate the model capability of the time-varying reduced-rank VAR for discovering interpretable patterns from extensive spatiotemporal datasets, including fluid dynamics, sea surface temperature, USA surface temperature, and NYC taxi trips. The evaluation results show that the latent variables in the tensor factorization of our model can reveal both spatial and temporal patterns. Time-varying system behaviors underlying these spatiotemporal data can be clearly identified by our model.

Focusing on the time-varying autoregression with low-rank structures, our model differs from the existing model proposed by [13] on the technical side: i) the existing model takes a sequence of windowed (first-order) VAR processes, while our model is parameterized with fully time-varying coefficients in the higher-order VAR (see Definition 1 and Remark 1 for more details); ii) the existing model utilizes the CANDECOMP/PARAFAC (CP) decomposition (i.e., a special case of Tucker decomposition [19]) to circumvent the over-parameterized coefficients in which CP decomposition does not show any interactions between spatial modes and temporal modes, in contrast, our model applies the Tucker decomposition which allows us to better characterize spatiotemporal patterns. Without loss of generality, the proposed model lays a foundation for modeling real-world spatiotemporal data with the time-varying system behaviors and provides insights into dynamical system modeling.

The remainder of this work is organized as follows. Section 3 introduces a time-varying reduced-rank VAR model and presents an alternating minimization algorithm for solving the involved optimization problem. Section 4 shows extensive experiments on some real-world spatiotemporal datasets. We discuss this research and give a summary in Section 5. For more supplementary material, Appendix A presents spatial/temporal modes achieved by DMD on the fluid dynamics.

2 RELATED WORK

2.1 Low-Rank Autoregression

In the past decades, substantial studies on time series modeling have sought solutions to the over-parameterized autoregression, mainly aiming at model compression on the coefficients. Essentially, the idea of low-rank regression is taking advantage of dimensionality reduction techniques. To solve the over-parameterization issue in the high-dimensional VAR, there are some low-rank matrix and

tensor tools ranging from the linear structure to multilinear structure, e.g., matrix factorization [25], Tucker decomposition [12], and tensor train decomposition [26], to compress the coefficient matrices in VAR. At an early stage, a general reduced-rank setup for VAR models was proposed by Ahn and Reinsel [7], [27]. Very recently, some representative models such as the multivariate autoregressive index model [28] and Bayesian compressed VAR model [25] integrated matrix factorization into the framework. Another idea is to use a low-rank tensor structure to characterize the higher-order VAR (i.e., $d > 1$), e.g., Tucker decomposition formatted VAR model in [12]. Bahadori *et al.* proposed a unified low-rank tensor learning framework for multivariate spatiotemporal analysis [29]. In these high-dimensional VAR models with the higher order, tensor factorization can perform better than matrix factorization in terms of model compression on the coefficients. However, tensor factorization would inevitably involve high complexity and nonconvex optimization. In the field of dynamical systems, DMD and higher-order DMD models can also be categorized as reduced-rank VAR models. They take the form of VAR and can handle high-dimensional time series because truncated singular value decomposition (SVD) and eigenvalue decomposition are adopted for preserving the most important dynamic modes of time series data [9], [10], [11], [30]. Therefore, matrix/tensor factorization in the low-rank autoregression not only addresses the over-parameterization issue, but also provides insights into pattern discovery.

2.2 Time-Varying Autoregression

In real-world applications, time series data are usually nonstationary. Recall that the stationary time series possess mean, average, and autocorrelation that are independent of time [4], but nonstationary time series violate that principle. Therefore, the nonstationarity characteristic poses great methodological and practical challenges for developing well-suited frameworks on such kind of data. Instead of stationarizing time series, the existing studies presented some autoregression models for characterizing time-varying system behaviors. In literature, a classical dynamic linear framework (e.g., time-varying VAR) utilizes time-varying latent variables to characterize the time-varying system behaviors of time series [14], [15], [16]. To reinforce such framework, many counterparts of time-varying autoregression such as time-varying structural VAR [15], [16], online DMD [22], and time-varying autoregression with low-rank tensors [13] have been introduced to time series analysis. Among these works, Harris *et al.* [13] introduced a time-varying VAR and addressed the over-parameterization issue through low-rank tensor factorization—tensor factorization allows one to achieve both model compression and pattern discovery. However, to the best of our knowledge, existing studies have not provided any systematical analysis for showing how to discover dynamic patterns from spatiotemporal data via time-varying autoregression.

3 METHODOLOGY

This section introduces a fully time-varying reduced-rank VAR model for the multivariate time series. In particular,

the model utilizes low-rank tensor factorization to compress the over-parameterized coefficients. Meanwhile, tensor factorization in the model allows us to automatically discover interpretable dynamic modes (corresponding to spatiotemporal patterns) from spatiotemporal data.

3.1 Model Description

In this work, we propose a time-varying VAR model with interpretability regarding the underlying system's dynamic behaviors. This model provides a flexible framework built on the past efforts such as DMD [9], [11] and time-varying autoregression with low-rank tensors [13]. For any spatiotemporal data in the form of multivariate time series, our model considers a time-varying linear system as follows.

Definition 1 (Fully Time-Varying VAR). Given any multivariate time series data $\mathbf{S} \in \mathbb{R}^{N \times T}$ with columns $\mathbf{s}_1, \dots, \mathbf{s}_T \in \mathbb{R}^N$, if a d th-order VAR takes time-varying coefficients at time $t \in \{d+1, \dots, T\}$, then the optimization problem of the fully time-varying VAR can be defined as follows,

$$\min_{\{\mathbf{A}_t\}} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{A}_t \mathbf{z}_t\|_2^2, \quad (1)$$

where $\mathbf{z}_t \triangleq (\mathbf{s}_{t-1}^\top, \dots, \mathbf{s}_{t-d}^\top)^\top \in \mathbb{R}^{dN}$ is an augmented vector and $\mathbf{y}_t \triangleq \mathbf{s}_t$. The data pair $\{\mathbf{y}_t, \mathbf{z}_t\}$ are the inputs for learning the coefficient matrices $\{\mathbf{A}_t \in \mathbb{R}^{N \times (dN)}\}$. The notation $\|\cdot\|_2$ denotes the ℓ_2 -norm of vector. Notably, these coefficient matrices are capable of expressing the time series with a sequence of parameters.

Remark 1. In contrast to the fully time-varying VAR as mentioned in Definition 1, Harris *et al.* [13] presented a windowed time-varying VAR with the first-order form (i.e., $d = 1$), which is given by

$$\min_{\{\mathbf{A}_t\}} \frac{1}{2} \sum_{t=1}^M \|\mathbf{Y}_t - \mathbf{A}_t \mathbf{Z}_t\|_F^2, \quad (2)$$

where $\{\mathbf{A}_t \in \mathbb{R}^{N \times N}\}$ are the coefficient matrices. The time steps $T = MK, M \in \mathbb{N}^+, K \in \mathbb{N}^+$ is comprised of M windows and each window is of length K . At any window t , the data pair $\{\mathbf{Y}_t, \mathbf{Z}_t\}$ are defined as augmented matrices:

$$\mathbf{Y}_t = \begin{bmatrix} | & & | \\ \mathbf{s}_{(t-1)K+1} & \cdots & \mathbf{s}_{tK-1} \\ | & & | \end{bmatrix} \in \mathbb{R}^{N \times (K-1)},$$

$$\mathbf{Z}_t = \begin{bmatrix} | & & | \\ \mathbf{s}_{(t-1)K+2} & \cdots & \mathbf{s}_{tK} \\ | & & | \end{bmatrix} \in \mathbb{R}^{N \times (K-1)}.$$

If $K = 1$, then the windowed time-varying VAR is reduced to our fully time-varying VAR with $d = 1$, showing the flexibility of the fully time-varying VAR over the windowed time-varying VAR.

In Definition 1, gathering $\{\mathbf{A}_t\}$ and stacking these matrices along an additional dimension, there exists a third-order tensor $\mathbf{A} \in \mathbb{R}^{N \times (dN) \times (T-d)}$ for representing these coefficient matrices, in which \mathbf{A}_t is the t th frontal slice of \mathbf{A} . One great concern of the model is that it possesses $\mathcal{O}(dN^2(T-d))$ parameters, which would vastly exceed $\mathcal{O}(NT)$ observations in most cases, regardless of the order

d in the model. Therefore, one can demonstrate that, due to time-varying coefficients, the autoregression model involves the *over-parameterization* issue. To address this issue, low-rank matrix/tensor factorization usually plays as a fundamental tool for compressing the over-parameterized coefficients, e.g., matrix factorization in reduced-rank regression/autoregression [6], [7], [8], [27] and tensor factorization in high-dimensional VAR [12]. In this work, we assume that the tensor \mathcal{A} has a reduced rank, i.e., the rank of \mathcal{A} is deficient, and claim the existence of a multilinear rank- (R, R, R) Tucker decomposition such that

$$\mathcal{A} = \mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{V} \times_3 \mathbf{X}, \quad (3)$$

where $\mathcal{G} \in \mathbb{R}^{R \times R \times R}$ is the core tensor, and $\mathbf{W} \in \mathbb{R}^{N \times R}, \mathbf{V} \in \mathbb{R}^{(dN) \times R}, \mathbf{X} \in \mathbb{R}^{(T-d) \times R}$ are factor matrices. The notation \times_k denotes the modal product between a tensor and a matrix along the mode k [19]. If the autoregressive process is not time-varying, i.e., in the case of static coefficients, then the model would be reduced to the precedent reduced-rank VAR with matrix factorization [7], [8], [27]. According to the property of Tucker decomposition and Kronecker product [19], [24], we can build connections between Eqs. (1) and (3) through

$$\mathcal{A}_t = \mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{V} \times_3 \mathbf{x}_t^\top \equiv \mathbf{W}\mathcal{G}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top, \quad (4)$$

where $\mathbf{x}_t \in \mathbb{R}^R$ is the t th row of \mathbf{X} , and $\mathbf{G} \triangleq \mathcal{G}_{(1)} \in \mathbb{R}^{R \times R^2}$ is the mode-1 unfolding/maticization of the core tensor \mathcal{G} . The symbol \otimes denotes the Kronecker product.

In this case, tensor factorization is used to compress the coefficients in the time-varying VAR. If one integrates the above-mentioned tensor factorization into time-varying VAR processes, then the optimization problem of the resultant time-varying reduced-rank VAR can be formulated as follows,

$$\min_{\mathbf{W}, \mathcal{G}, \mathbf{V}, \mathbf{X}} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{W}\mathcal{G}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2. \quad (5)$$

As can be seen, we draw connections between Tucker decomposition and time-varying VAR and provide an expression for decomposing the coefficients into a sequence of matrices. The model not only addresses the over-parameterization issue in the time-varying VAR, but also provides physically interpretable modes for real-world time series data. As demonstrated by [13], one significant benefit of tensor factorization in the model is revealing the underlying data patterns in the parameters. For instance, on spatiotemporal data, \mathbf{W} and \mathbf{V} can be interpreted as spatial modes, while \mathbf{X} can be interpreted as temporal modes. Time-varying $\{\mathbf{x}_t\}$ can explore the system's behavior of time series data. Since our model takes into account time-varying coefficients, we can offer a better understanding of dynamics of spatiotemporal data with nonstationarity and time-varying system behaviors. In contrast to our model, the time-varying autoregression with low-rank tensors builds VAR for the time series in each window and follows CP tensor factorization for model compression and outlier detection [13], and which seems to be less parsimonious than our model in terms of interpretable structures.

Lemma 1. *The optimization problem in Eq. (5) is equivalent to the following optimization problem in the form of matrix:*

$$\min_{\mathbf{W}, \mathcal{G}, \mathbf{V}, \mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{W}\mathcal{G}(\mathbf{X} \otimes \mathbf{V})^\top \tilde{\mathbf{Z}}\|_F^2, \quad (6)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of matrix, and $\mathbf{Y}, \tilde{\mathbf{Z}}$ are defined as follows,

$$\mathbf{Y} \triangleq \begin{bmatrix} | & | & & | \\ \mathbf{y}_{d+1} & \mathbf{y}_{d+2} & \cdots & \mathbf{y}_T \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{N \times (T-d)},$$

and

$$\tilde{\mathbf{Z}} \triangleq \begin{bmatrix} \mathbf{z}_{d+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{d+2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_T \end{bmatrix} \in \mathbb{R}^{(dN(T-d)) \times (T-d)},$$

respectively.

Remark 2. This matrix-form optimization problem provides insights into understanding the problem of the time-varying reduced-rank VAR. Putting the orthogonal factor matrices with the matrix-form optimization problem together, our model can be written as a higher-order singular value decomposition (HOSVD, [19], [24]) on the coefficient tensor. In this case, we give an unfolded form of HOSVD as follows,

$$\begin{aligned} \min_{\mathbf{W}, \mathcal{G}, \mathbf{V}, \mathbf{X}} \quad & \frac{1}{2} \|\mathcal{A}_{(1)} - \mathbf{W}\mathcal{G}(\mathbf{X} \otimes \mathbf{V})^\top\|_F^2 \\ \text{s.t. } & \begin{cases} \mathbf{W}^\top \mathbf{W} = \mathbf{I}_R, \\ \mathbf{V}^\top \mathbf{V} = \mathbf{I}_R, \\ \mathbf{X}^\top \mathbf{X} = \mathbf{I}_R, \end{cases} \end{aligned} \quad (7)$$

where $\mathcal{A}_{(1)} \triangleq \mathbf{Y}\tilde{\mathbf{Z}}^\dagger \in \mathbb{R}^{N \times (dN(T-d))}$ is the mode-1 unfolding of \mathcal{A} . Herein, \cdot^\dagger denotes the Moore-Penrose pseudoinverse. In the constraint, \mathbf{I}_R is the R -by- R identity matrix.

As can be seen, our model shows the capability of HOSVD for pattern discovery, but in high-dimensional settings, computing with the large and sparse matrix $\tilde{\mathbf{Z}}$ is computationally costly. In what follows, we let f be the objective function of the optimization problem in Eq. (5), and still use the vector-form formula to develop an alternating minimization algorithm for the time-varying reduced-rank VAR.

3.2 Model Estimation

Recall that alternating minimization (e.g., alternating least squares) is a simple yet efficient method for addressing a manifold of nonconvex optimization problems in low-rank matrix/tensor factorization [19], [24]. To solve the involved optimization problem in our model, we apply an alternating minimization algorithm in which we have a sequence of least squares subproblems with respect to $\{\mathbf{W}, \mathcal{G}, \mathbf{V}, \mathbf{X}\}$, respectively. Regarding the subproblem with respect to each unknown variable, we fix the remaining variables as known variables and intend to write down the least squares solutions or obtain the numerically approximated solutions. In addition, as reported in [13], the estimation based on alternating minimization has shown to be effective in such optimization problems.

3.2.1 Updating the Variable \mathbf{W}

With respect to the variable \mathbf{W} , the current estimation task is deriving closed-form solution from

$$\mathbf{W} := \arg \min_{\mathbf{W}} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2, \quad (8)$$

while $\{\mathbf{G}, \mathbf{V}, \mathbf{X}\}$ are fixed as known variables. Since this optimization problem is convex, we can first write down the partial derivative of f with respect to \mathbf{W} as follows,

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{W}} &= - \sum_{t=d+1}^T (\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t) \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \mathbf{G}^\top \\ &= - \sum_{t=d+1}^T \mathbf{y}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \mathbf{G}^\top \\ &\quad + \sum_{t=d+1}^T \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \mathbf{G}^\top. \end{aligned} \quad (9)$$

Then let $\frac{\partial f}{\partial \mathbf{W}} = \mathbf{0}$, there exists a least squares solution to the variable \mathbf{W} :

$$\begin{aligned} \mathbf{W} &= \left(\sum_{t=d+1}^T \mathbf{y}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \mathbf{G}^\top \right) \\ &\quad \cdot \left(\sum_{t=d+1}^T \mathbf{G}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \mathbf{G}^\top \right)^{-1}. \end{aligned} \quad (10)$$

3.2.2 Updating the Variable \mathbf{G}

With respect to the variable \mathbf{G} , the estimation goal is finding the optimal solution from

$$\mathbf{G} := \arg \min_{\mathbf{G}} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2, \quad (11)$$

while $\{\mathbf{W}, \mathbf{V}, \mathbf{X}\}$ are fixed as known variables. To find the closed-form solution from this optimization problem, the partial derivative of f with respect to \mathbf{G} is given by

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{G}} &= - \sum_{t=d+1}^T \mathbf{W}^\top (\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t) \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \\ &= - \sum_{t=d+1}^T \mathbf{W}^\top \mathbf{y}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \\ &\quad + \sum_{t=d+1}^T \mathbf{W}^\top \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}). \end{aligned} \quad (12)$$

Let $\frac{\partial f}{\partial \mathbf{G}} = \mathbf{0}$, then there exists a least squares solution to the variable \mathbf{G} :

$$\begin{aligned} \mathbf{G} &= \mathbf{W}^\dagger \left(\sum_{t=d+1}^T \mathbf{y}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \right) \\ &\quad \cdot \left(\sum_{t=d+1}^T (\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t \mathbf{z}_t^\top (\mathbf{x}_t^\top \otimes \mathbf{V}) \right)^{-1}. \end{aligned} \quad (13)$$

3.2.3 Updating the Variable \mathbf{V}

With respect to the variable \mathbf{V} , the optimization problem is given by

$$\mathbf{V} := \arg \min_{\mathbf{V}} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2, \quad (14)$$

where $\{\mathbf{W}, \mathbf{G}, \mathbf{X}\}$ are fixed as known variables. To get the closed-form solution, one is required to reformulate the objective function because the involved Kronecker product complicates the formula. According to the property of Kronecker product (see Lemma 2), we can rewrite the objective function as follows,

$$\begin{aligned} f &= \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{WG}(\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2 \\ &= \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{WG}((\mathbf{x}_t \mathbf{z}_t^\top) \otimes \mathbf{I}_R) \text{vec}(\mathbf{V}^\top)\|_2^2, \end{aligned} \quad (15)$$

where $\text{vec}(\cdot)$ denotes the operator of vectorization.

Lemma 2. For any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{Y} \in \mathbb{R}^{p \times q}$, $\mathbf{z} \in \mathbb{R}^p$, there exists

$$\begin{aligned} (\mathbf{x}^\top \otimes \mathbf{Y})^\top \mathbf{z} &= (\mathbf{x} \otimes \mathbf{Y}^\top) \mathbf{z} \\ &= \text{vec}(\mathbf{Y}^\top \mathbf{z} \mathbf{x}^\top) \\ &= \text{vec}(\mathbf{I}_q \mathbf{Y}^\top (\mathbf{z} \mathbf{x}^\top)) \\ &= ((\mathbf{x} \mathbf{z}^\top) \otimes \mathbf{I}_q) \text{vec}(\mathbf{Y}^\top). \end{aligned} \quad (16)$$

Remark 3. This lemma stems from one fundamental property of Kronecker product, which is given by

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X}), \quad (17)$$

for any $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{X} \in \mathbb{R}^{m \times n}$, and $\mathbf{B} \in \mathbb{R}^{n \times n}$ commensurate from multiplication in that order.

Therefore, the partial derivative of f with respect to the vectorized variable $\text{vec}(\mathbf{V}^\top)$ is given by

$$\begin{aligned} \frac{\partial f}{\partial \text{vec}(\mathbf{V}^\top)} &= - \sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t \\ &\quad + \sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \\ &\quad \cdot \mathbf{WG}((\mathbf{x}_t \mathbf{z}_t^\top) \otimes \mathbf{I}_R) \text{vec}(\mathbf{V}^\top). \end{aligned} \quad (18)$$

Let $\frac{\partial f}{\partial \text{vec}(\mathbf{V}^\top)} = \mathbf{0}$, then there exists a system of linear equations:

$$\begin{aligned} &\sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \mathbf{WG}((\mathbf{x}_t \mathbf{z}_t^\top) \otimes \mathbf{I}_R) \text{vec}(\mathbf{V}^\top) \\ &= \sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t. \end{aligned} \quad (19)$$

This system of linear equations has a least squares solution. However, the potential inverse of the large matrix would cost $\mathcal{O}((dNR)^3)$, involving high computational complexity. To convert this large-scale and sparse problem into an easy-to-solve problem, we utilize Lemma 3 to establish an equivalent generalized Sylvester equation with respect to the variable \mathbf{V} .

Lemma 3. For any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{Y} \in \mathbb{R}^{q \times n}$, $\mathbf{z} \in \mathbb{R}^p$, there exists

$$((\mathbf{z}\mathbf{x}^\top) \otimes \mathbf{I}_q) \text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{Y}\mathbf{x}\mathbf{z}^\top). \quad (20)$$

According to the property of Kronecker product as mentioned in Lemma 3, Eq. (19) is equivalent to the following generalized Sylvester equation:

$$\begin{aligned} & \sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \mathbf{W} \mathbf{G} (\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t \\ &= \sum_{t=d+1}^T ((\mathbf{z}_t \mathbf{x}_t^\top) \otimes \mathbf{I}_R) \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t. \end{aligned} \quad (21)$$

As can be seen, it yields a generalized Sylvester equation that estimates the solution to \mathbf{V}^\top :

$$\sum_{t=d+1}^T \mathbf{P}_t \mathbf{x}_t \mathbf{z}_t^\top = \sum_{t=d+1}^T \mathbf{Q}_t \mathbf{x}_t \mathbf{z}_t^\top, \quad (22)$$

where we define two auxiliary variables $\mathbf{P}_t, \mathbf{Q}_t \in \mathbb{R}^{R \times R}$ such that

$$\begin{cases} \text{vec}(\mathbf{P}_t) \triangleq \mathbf{G}^\top \mathbf{W}^\top \mathbf{W} \mathbf{G} (\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t, \\ \text{vec}(\mathbf{Q}_t) \triangleq \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t. \end{cases}$$

With respect to the variable \mathbf{V} , the first impulse is to take a transpose operation on both left-hand and right-hand sides of Eq. (22). Then the resultant equation is given by

$$\sum_{t=d+1}^T \mathbf{z}_t \mathbf{x}_t^\top \mathbf{P}_t^\top = \sum_{t=d+1}^T \mathbf{z}_t \mathbf{x}_t^\top \mathbf{Q}_t^\top. \quad (23)$$

In this case, we can use the conjugate gradient, a classical and efficient numerical approximation algorithm [24], to solve the generalized Sylvester equation. To solve Eq. (23) through conjugate gradient, we need to define an operator on the left-hand side of the equation as follows,

$$\mathcal{L}_v(\mathbf{V}) \triangleq \text{vec} \left(\sum_{t=d+1}^T \mathbf{z}_t \mathbf{x}_t^\top \mathbf{P}_t^\top \right) \in \mathbb{R}^{dNR}. \quad (24)$$

Algorithm 1 summarizes the estimation procedure for approximating the solution to the variable \mathbf{V} in Eq. (23). The conjugate gradient method allows one to search for the approximated solution to a system of linear equations with a relatively small number of iterations (e.g., 5 or 10) [24], showing convergence guarantee for solving subproblems of matrix/tensor factorization [31]. Admittedly, the conjugate gradient methods with a small number of iterations cannot match the least squares solution. Nevertheless, the numerical approximated solution via conjugate gradient is rather accurate [24].

3.2.4 Updating the Variable $\{\mathbf{x}_t\}$

According to the property of Kronecker product, we can rewrite the objective function of the optimization problem in Eq. (5) as follows,

$$\begin{aligned} f &= \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{W} \mathbf{G} (\mathbf{x}_t^\top \otimes \mathbf{V})^\top \mathbf{z}_t\|_2^2 \\ &= \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{y}_t - \mathbf{W} \mathbf{G} (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t)) \mathbf{x}_t\|_2^2. \end{aligned} \quad (25)$$

Algorithm 1 Conjugate gradient for estimating \mathbf{V}

Input: Data pair $\{\mathbf{y}_t, \mathbf{z}_t\}$, known variables $\{\mathbf{W}, \mathbf{G}, \mathbf{X}\}$, initialized \mathbf{V} , and the maximum iteration \tilde{L} (e.g., the default value as $\tilde{L} = 5$).

Output: Estimated \mathbf{V} .

- 1: Let $\mathcal{R}_v = \mathbf{0}$.
 - 2: **for** $t = d + 1$ to T **do**
 - 3: Compute \mathbf{Q}_t with $\text{vec}(\mathbf{Q}_t) \triangleq \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t$.
 - 4: Take $\mathcal{R}_v += \mathbf{z}_t \mathbf{x}_t^\top \mathbf{Q}_t^\top$.
 - 5: **end for**
 - 6: Initialize \mathbf{v}_0 by the vectorized \mathbf{V} .
 - 7: Compute residual vector $\mathbf{r}_0 = \text{vec}(\mathcal{R}_v) - \mathcal{L}_v(\mathbf{V})$, and $\mathbf{q}_0 = \mathbf{r}_0$.
 - 8: **for** $\ell = 0$ to $\tilde{L} - 1$ **do**
 - 9: Convert vector \mathbf{q}_ℓ into matrix \mathbf{Q}_ℓ .
 - 10: Compute $\alpha_\ell = \frac{\mathbf{r}_\ell^\top \mathbf{r}_\ell}{\mathbf{q}_\ell^\top \mathcal{L}_v(\mathbf{Q}_\ell)}$.
 - 11: Update $\mathbf{v}_{\ell+1} = \mathbf{v}_\ell + \alpha_\ell \mathbf{q}_\ell$.
 - 12: Update $\mathbf{r}_{\ell+1} = \mathbf{r}_\ell - \alpha_\ell \mathcal{L}_v(\mathbf{Q}_\ell)$.
 - 13: Compute $\beta_\ell = \frac{\mathbf{r}_{\ell+1}^\top \mathbf{r}_{\ell+1}}{\mathbf{r}_\ell^\top \mathbf{r}_\ell}$.
 - 14: Update $\mathbf{q}_{\ell+1} = \mathbf{r}_{\ell+1} + \beta_\ell \mathbf{q}_\ell$.
 - 15: **end for**
 - 16: Convert vector $\mathbf{v}_{\tilde{L}}$ into matrix \mathbf{V} .
-

Thus, the optimization problem with respect to $\mathbf{x}_t, \forall t$ now becomes

$$\mathbf{x}_t := \arg \min_{\mathbf{x}_t} \frac{1}{2} \|\mathbf{y}_t - \mathbf{W} \mathbf{G} (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t)) \mathbf{x}_t\|_2^2, \quad (26)$$

while $\{\mathbf{W}, \mathbf{G}, \mathbf{V}\}$ are fixed as known variables.

Then, we can obtain the partial derivative of f with respect to the variable \mathbf{x}_t (i.e., the t th row of the variable \mathbf{X}), which is given by

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{x}_t} &= (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t))^\top \mathbf{G}^\top \mathbf{W}^\top (\mathbf{W} \mathbf{G} (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t)) \mathbf{x}_t - \mathbf{y}_t) \\ &= (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t))^\top \mathbf{G}^\top \mathbf{W}^\top \mathbf{W} \mathbf{G} (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t)) \mathbf{x}_t \\ &\quad - (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t))^\top \mathbf{G}^\top \mathbf{W}^\top \mathbf{y}_t, \end{aligned} \quad (27)$$

In this case, letting $\frac{\partial f}{\partial \mathbf{x}_t} = \mathbf{0}$ leads to a least squares solution to the variable \mathbf{x}_t :

$$\mathbf{x}_t = (\mathbf{W} \mathbf{G} (\mathbf{I}_R \otimes (\mathbf{V}^\top \mathbf{z}_t)))^\dagger \mathbf{y}_t. \quad (28)$$

3.3 Solution Algorithm

As mentioned above, the variables $\{\mathbf{W}, \mathbf{G}, \mathbf{X}\}$ can be computed by the closed-form least squares solutions, while the solution to the variable \mathbf{V} can be numerically approximated by the conjugate gradient method in an efficient manner. Starting from the initialized variables by using SVD, then we update these variables in an iterative routine. In the iterative process, the basic idea of alternating minimization is that we fix the remaining variables when updating one of these variables. Since each subproblem is convex and there is a unique solution to each of the coordinate-wise minimization problems, the convergence of our algorithm through the block coordinate minimization can be therefore verified [13].

Algorithm 2 Time-varying reduced-rank VAR

Input: Data pair $\{\mathbf{y}_t, \mathbf{z}_t\}$ constructed by the time series data $S \in \mathbb{R}^{N \times T}$, d as the order of VAR, R as the low rank ($R \leq \min\{N, T - d\}$), and L as the maximum iteration.
Output: $\{\mathbf{W}, \mathbf{G}, \mathbf{V}, \mathbf{X}\}$.

- 1: Initialize factor matrices $\mathbf{W}, \mathbf{V}, \mathbf{X}$ by the R left singular vectors of $\mathbf{Y}, \mathbf{Z}, \mathbf{S}^\top$, respectively.
- 2: **for** $\ell = 0$ to $L - 1$ **do**
- 3: Update \mathbf{G} by Eq. (13).
- 4: Update \mathbf{W} by Eq. (10).
- 5: Compute \mathbf{V} from Eq. (23) with conjugate gradient (see Algorithm 1).
- 6: **for** $t = d + 1$ to T **do**
- 7: Update \mathbf{x}_t by Eq. (28).
- 8: **end for**
- 9: **end for**

4 EXPERIMENTS

In this section, we evaluate our model on an artificial dataset for fluid dynamics and some real-world datasets, including sea surface temperature, USA surface temperature, and NYC taxi trips. We demonstrate the performance of our model on characterizing the underlying data patterns and system's behavior, especially showing the modeling capability and interpretability on time-varying systems. All the experiment results are replicable at <https://github.com/xinychen/vars>.

4.1 Fluid Dynamics

Investigating fluid dynamic systems is of great interest for uncovering large-scale spatiotemporal coherent structures because dominant patterns exist in the flow field. The data-driven models, such as proper orthogonal decomposition (POD) [32] and DMD [10], [11], [23], have become an important paradigm. To analyze the underlying spatiotemporal patterns of fluid dynamics, we apply our data-driven model to the cylinder wake dataset in which the flow shows a supercritical Hopf bifurcation. The dataset is collected from the fluid flow passing a circular cylinder with laminar vortex shedding at Reynolds number $Re = 100$, which is larger than the critical Reynolds number, using direct numerical simulations of the Navier-Stokes equations.¹ This is a representative three-dimensional flow dataset in fluid dynamics, consisting of matrix-variate time series of vorticity field snapshots for the wake behind a cylinder. The dataset is of size $199 \times 449 \times 150$, representing 199-by-449 vorticity fields with 150 time snapshots (see some examples in Fig. 1a).

We first manually build an artificial dataset based on fluid dynamics observations to test the proposed model. First, we can reshape the data as a high-dimensional multivariate time series matrix of size 89351×150 . Then, to manually generate a multiresolution system in which the fluid flow takes dynamic-varying system behaviors, we concatenate two parts of data with different frequencies—putting i) the first 50 snapshots (original frequency) together with ii) the uniformly sampled 50 snapshots from the last 100

snapshots (double frequency). As a consequence, the newly-built fluid flow dataset for evaluation has 100 snapshots in total but with a frequency transition in its system behaviors, i.e., possessing different frequencies in two phases. Multiresolution fluid dynamic data come with their own challenges such as multiple frequencies that the standard DMD model cannot work effectively (see Appendix A for further information). Regarding these challenges, finding a spatiotemporal coherent structure from multiresolution data is a challenging task, but of great significance. As demonstrated by [33], uncovering such multiscale system can recognize and separate multiscale spatiotemporal features.

In our model, rank R is a key parameter, which indicates the number of dominant spatial/temporal modes. In practice, a lower rank may only help reveal a few low-frequency dominant modes, but a higher rank would bring some complicated and less interpretable modes, usually referring to high-frequency information [34]. This is consistent with the law that nature typically follows—noise is usually dominant at high frequencies and the system signal is more dominant at lower frequencies. On this dataset, we set the rank of our model as $R = 7$. Fig. 1b shows the spatial modes of the fluid flow revealed by our model. It demonstrates that the spatial mode 1 corresponds to a background mode that is not changing over time because it is consistent with the mean vorticity. The other dominant spatial modes essentially show the waves of fluid flow, which look similar to both DMD and POD modes on the same analysis [11]. With the increase of rank, the spatial modes can be more detailed. In this case, observing the harmonic frequencies of temporal modes (see Fig. 1c), the corresponding spatial modes 4 and 5 are more complicated than the spatial modes 2 and 3, while the spatial modes 6 and 7 are more complicated than the spatial modes 4 and 5. Despite recognizing similar spatial modes as POD, our model can also discover the temporal modes that reveal how the system behaviors evolve. Fig. 1c shows the temporal modes of the fluid flow in \mathbf{X} . As can be seen, the frequency of all temporal modes in \mathbf{X} is changing at the time $t = 50$, and all temporal modes can identify the time series with different frequencies of oscillation. The dynamics of fluid flow essentially consist of the phases of two frequencies. Thus, we can emphasize the model's ability for identifying the time-evolving patterns from multiresolution fluid flow. The temporal mode 1 is the most dominant pattern of the fluid flow, corresponding to the spatial mode 1. With higher rank, the frequency of harmonic cycles increases, implying that the importance of the latter modes is secondary and the last spatial/temporal modes represent high-frequency information. Therefore, we can tell that our model can discover both spatial and temporal modes from the spatiotemporal data with time-varying system behaviors.

4.2 Sea Surface Temperature (SST)

The oceans play a very important role in the global climate system. Exploiting the SST data allows one to monitor the climate change and understand the dynamical processes of energy exchange at the sea surface [33], [35]. Here, we consider the SST dataset that covers weekly means of

1. <http://dmdbook.com/>

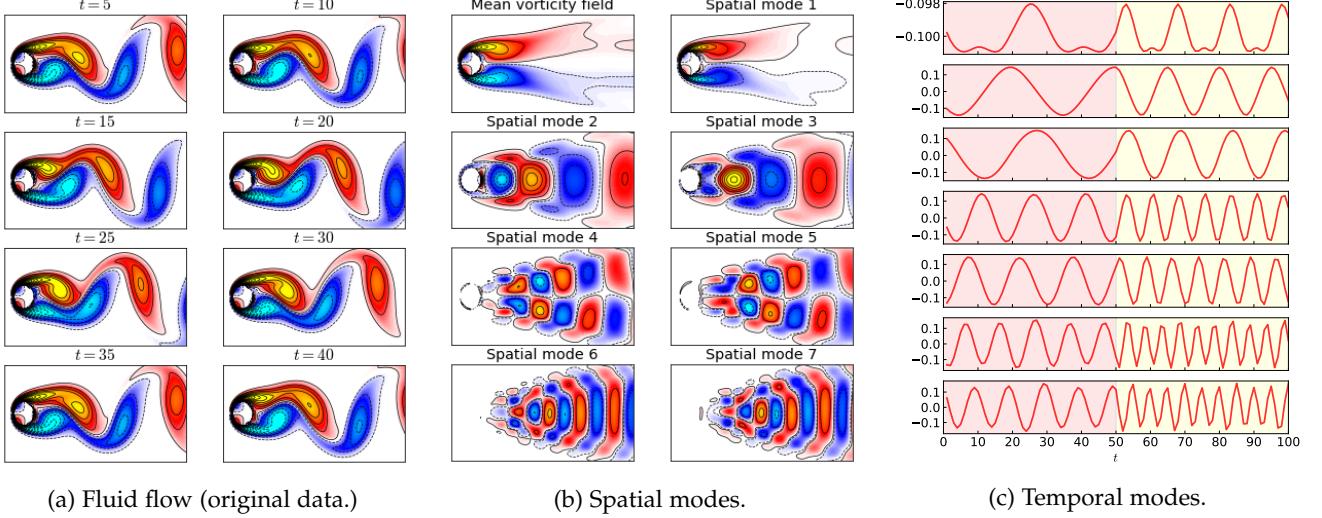


Fig. 1: Fluid flow and spatial/temporal modes to demonstrate the model. (a) Heatmaps (snapshots) of the fluid flow at times $t = 5, 10, \dots, 40$. It shows that the snapshots at times $t = 5$ and $t = 35$ are even same, and the snapshots at times $t = 10$ and $t = 40$ are also even same, allowing one to infer the seasonality as 30 for the first 50 snapshots. (b) Mean vorticity field and spatial modes of the fluid flow. Spatial modes are plotted by the columns of \mathbf{W} in which seven panels correspond to the rank $R = 7$. Note that the colorbars of all modes are on the same scale. (c) Temporal modes of the fluid flow in \mathbf{X} . Seven panels correspond to the rank $R = 7$.

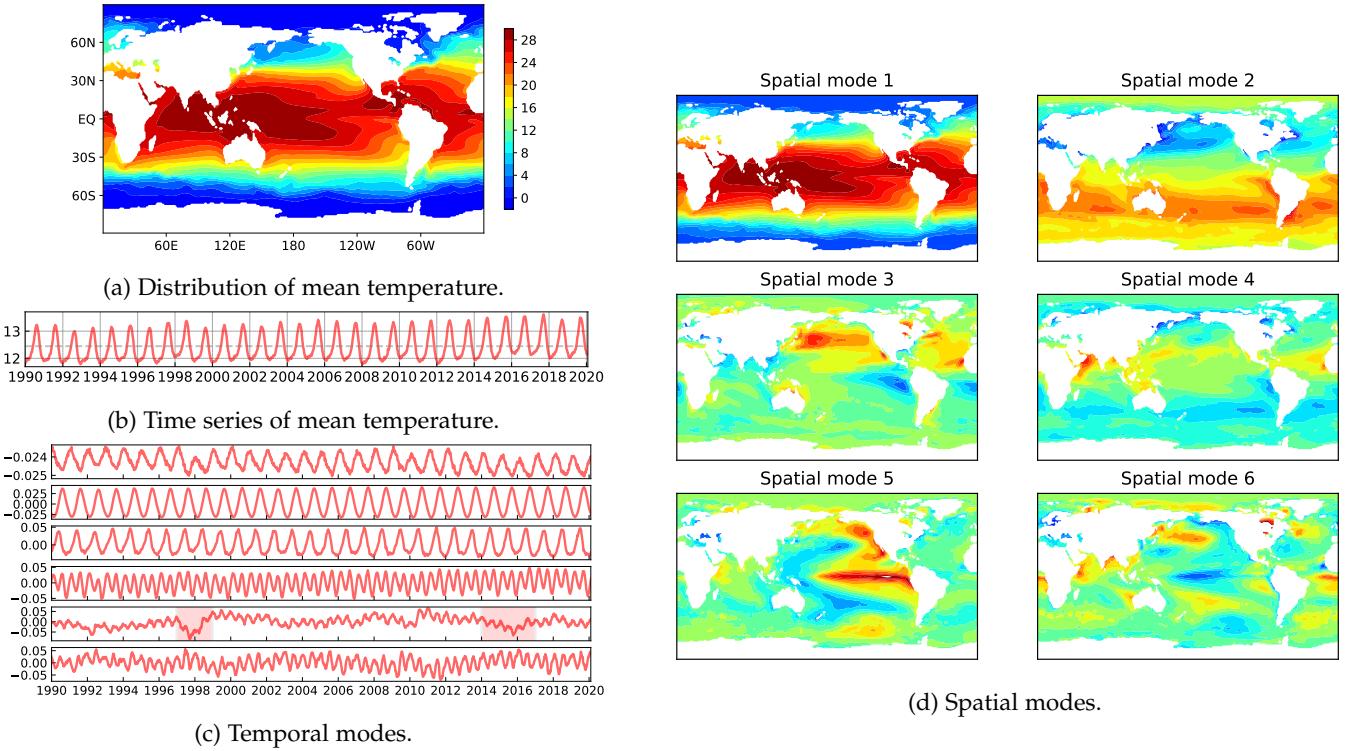


Fig. 2: Model application and results on SST data from 1990 to 2019. (a) Geographical distribution of long-term mean values of the SST over the 30-year period. (b) The time series of mean values of the SST over the 30-year period. The whole average temperature is 12.45°C . (c) Temporal modes of the SST achieved by our model. From the top panel to the bottom panel, we have the temporal modes $1, \dots, 6$, respectively. (d) Geographical distribution of spatial modes of the SST achieved by our model. Spatial modes are plotted by the columns of \mathbf{W} in which six panels correspond to the rank $R = 6$.

temperature on the spatial resolution of (1 degree latitude, 1 degree longitude)-grid, and there are 180×360 global

grids (i.e., 64,800 grids) in total.² The dataset spans a 30-year period from 1990 to 2019, and the time dimension is of

2. <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html>

length 1,565 (weeks). Therefore, the data can be represented as a matrix of size 64800×1565 , which seems to be high-dimensional. Fig. 2a exhibits the long-term mean values of the SST dataset over the 30-year period. It is not hard to see the most basic patterns of geographical distribution of SST. Fig. 2b shows the yearly cycle (≈ 52 weeks) of time series of the mean temperature.

Using the proposed model with rank $R = 6$, we plot the temporal modes and spatial modes of the SST data in Fig. 2c and 2d, respectively. Our model can identify both quasi-periodic and non-periodic behaviors and patterns from complicated and noisy signals. In contrast, DMD models are empirically demonstrated to be sensitive to the data noise in SST [11]. The temporal modes 1, 2, 3, and 4 take a yearly cycle/rhythm, showing dominant patterns of SST. Notably, the spatial modes revealed by our data-driven model is consistent with the previous literature [33]. Specifically, the spatial mode 1, as a background mode, is consistent with the long-term mean temperature. The other spatial modes explain the deviations from the long-term monthly means respectively. It is worth noting that some special oceanic events can also be revealed in these spatial and temporal modes. For instance, the spatial/temporal modes 3 and 4 convey the two Southern Annular Modes in Southern Ocean as well as the positive and negative phases of the North Atlantic Oscillation. In addition, the spatial/temporal modes 5 and 6 demonstrate the phenomenon of El Niño Southern Oscillation (with both of El Niño and La Niña), as well as the Pacific Decadal Oscillation. Combining with the temporal mode 5 in Fig. 2c, we can directly identify two strongest El Niño events on record, i.e., happened on 1997–1998 and 2014–2016, corresponding to the spatial mode 5 as highlighted in Fig. 2d. These results can also be generated through the multiresolution DMD model [33], but the multiresolution DMD are incapable of discovering the complex SST systems.

4.3 USA Surface Temperature Data

Daymet project provides long-term and continuous estimates of daily weather parameters such as maximum and minimum daily temperature for North America.³ There are 5,380 stations over the United States Mainland. In this work, we use the daily maximum temperature data in the United States Mainland from 2010 to 2021 (i.e., 12 years or 4,380 days in total) for evaluation. The data can be represented as a matrix of size 5380×4380 . In particular, we apply the nonstationary temporal matrix factorization model [36] to impute 2.67% missing values in the original data and conduct the following evaluation on the recovered data. Fig. 3a shows the long-term mean values of the temperature dataset over the 12-year period, demonstrating the most basic patterns of geographical distribution of the temperature data. Fig. 3b demonstrates the mean temperature changing over time, showing strong yearly rhythms.

Fig. 4a shows the geographical distribution of the spatial modes of the temperature data revealed by \mathbf{W} in our model, while Fig. 4b and 4c visualize the temporal modes of the temperature data revealed by \mathbf{X} in our model. The spatial mode 1 demonstrates the most dominant mode which is

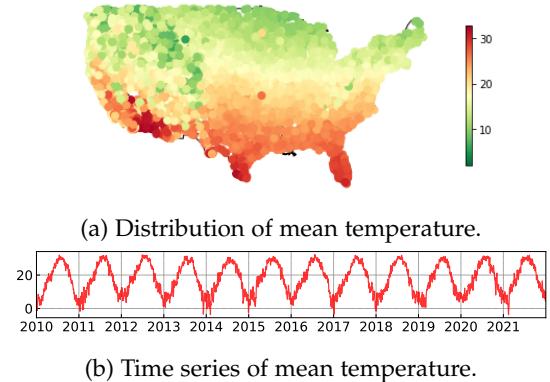


Fig. 3: Mean temperature of the maximum daily temperature data in the United States Mainland from 2010 to 2021. (a) Geographical distribution of the long-term mean temperature over the 12-year period. (b) The time series of mean temperature over the 12-year period.

consistent with the mean temperature as shown in Fig. 3a. In the meanwhile, the temporal mode 1 shows strong yearly cycles/rhythms underlying the daily temperature. The temporal mode 2 shows the similar time series curve as the temporal mode 1, but their values are quite different. The spatial mode 2 highlights the relatively hot areas and relatively cold areas. The temporal modes 3 and 4 are rather complicated, but the corresponding spatial modes 3 and 4 are intuitive for understanding the geographical distribution of these patterns. The spatial mode 3 roughly highlights the eastern areas and the western areas, identifying different characteristics. The spatial mode 4 identifies three areas in which the eastern and western areas follow the similar feature and the central areas take another feature.

Recall that rank is a key parameter of our model, which also refers to the number of spatial/temporal modes. Thus, the prescribed rank determines how many dominant modes we discover. Fig. 5a, 5b, and 5c show the geographical distribution of spatial modes of the temperature data achieved by our model with ranks $R = 6, 7, 8$, respectively. It is clear to see that the first four spatial modes achieved by our model with ranks $R = 6, 7, 8$ are same as the spatial modes in Fig. 4a. As shown in Fig. 5a, 5b, and 5c, the spatial modes achieved by our model with a relatively large rank (e.g., $R = 8$) can cover the spatial modes achieved by our model with a relatively small rank (e.g., $R = 6, 7$). Furthermore, as the rank increases, the spatial modes achieved by our model tend to reveal higher-frequency information and more complicated patterns than the model with a relatively small rank. Thus, these findings demonstrate that our model is capable of effectively discovering dominant spatial and temporal patterns from the real-world data.

In the above experiments, we use the model with order $d = 1$. In practice, we can also set a relatively large order for the time-varying VAR. Recall that we have two variables associated with the spatial dimension, i.e., $\mathbf{W} \in \mathbb{R}^{N \times R}$ and $\mathbf{V} \in \mathbb{R}^{(dN) \times R}$, in which we assume that the spatial modes revealed by \mathbf{W} do not change with the increase of the order d . Fig. 6a and 6b show the spatial modes achieved by our model with rank $R = 4$ and order $d = 2, 3$. These results

3. <https://daac.ornl.gov/DAYMET>

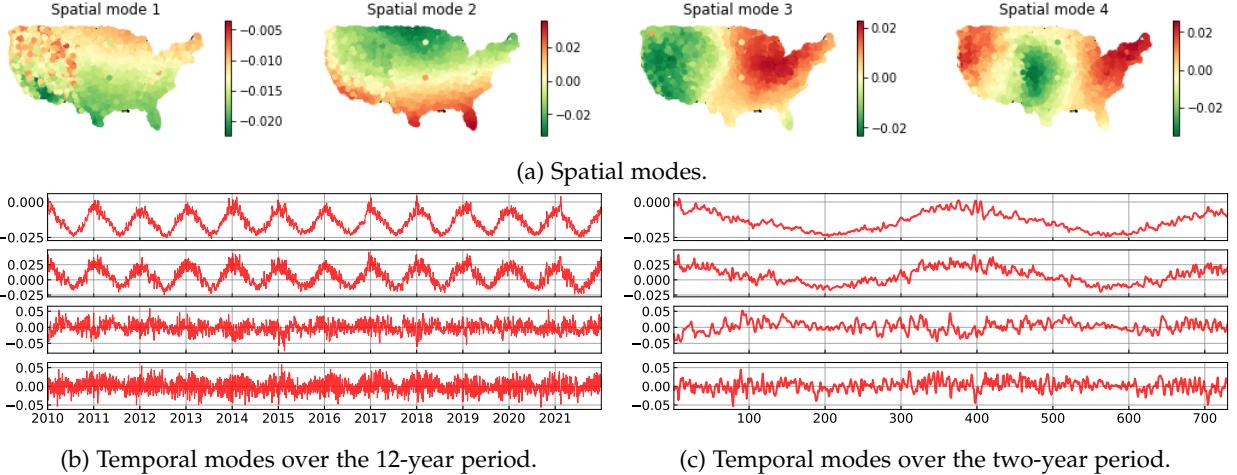


Fig. 4: Model application and results on the daily temperature data in the United States Mainland. (a) Geographical distribution of four spatial modes of the temperature data achieved by our model. (b) Four temporal modes (over the 12-year period) of the temperature data achieved by our model. From the top panel to the bottom panel, we have the temporal modes 1-4, respectively. (c) Four temporal modes (during the two years (730 days) from 2010 to 2011) of the temperature data achieved by our model.

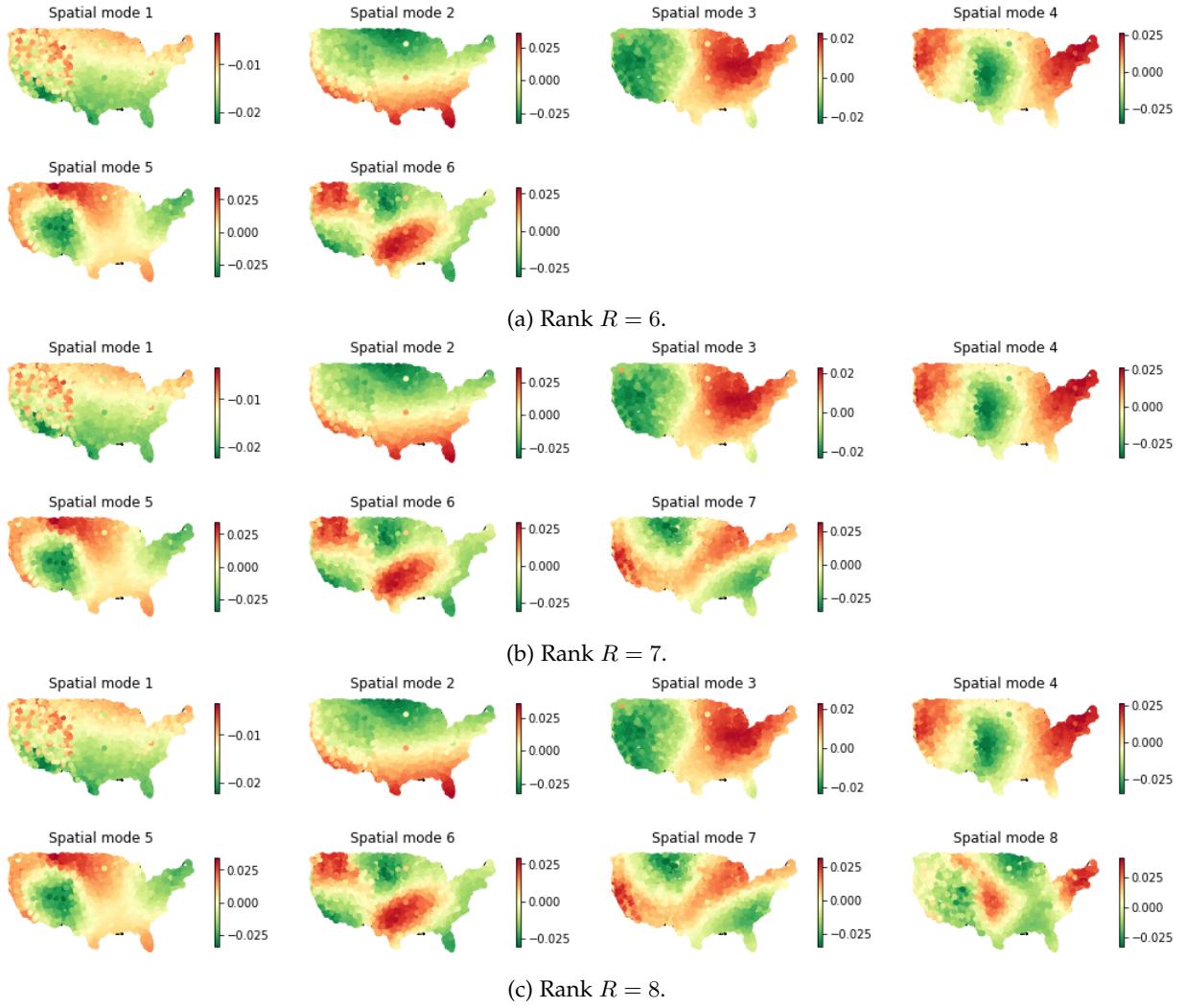


Fig. 5: Geographical distribution of spatial modes of the temperature data achieved by our model.

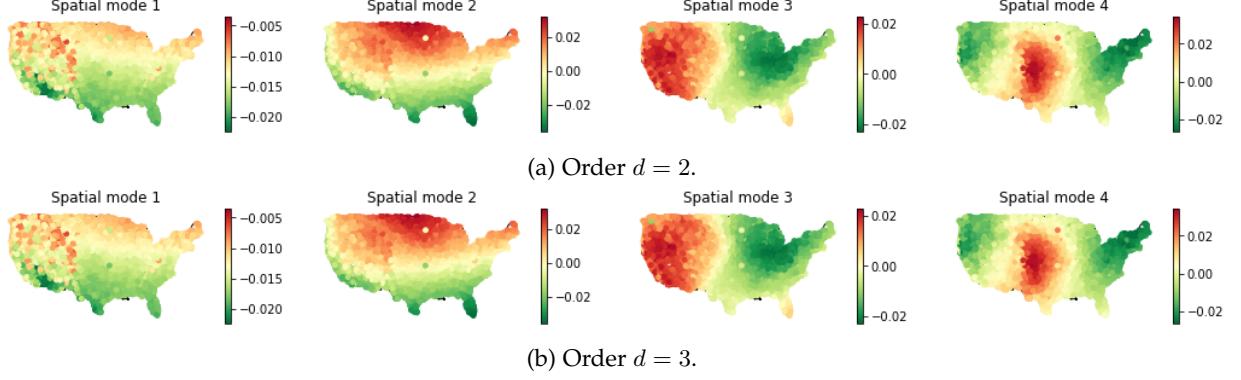


Fig. 6: Geographical distribution of four spatial modes of the temperature data achieved by our model with rank $R = 4$.

demonstrate that our model discovers similar spatial modes with different orders.

4.4 NYC Taxi Trips

Human mobility data are always associated with strong quasi-seasonality (e.g., weekly rhythms) and trends (e.g., decline of trips due to unplanned events like COVID-19), it is challenging to characterize these nonlinear and time-varying system behaviors. In this work, we consider to use an NYC (yellow) taxi trip dataset.⁴ We use 69 zones in Manhattan as pickup/dropoff zones and aggregate daily taxi trip volume of the data from 2012 to 2021. Therefore, the daily trip volume tensor is of size $69 \times 69 \times 3653$. In the following analysis, we aggregate the trip volume tensor as a pickup trip volume matrix and a dropoff trip volume matrix, both of size 69×3653 . The left panel of Fig. 8a and 9a shows the total pickup trips and total dropoff trips, respectively. As can be seen, most pickup/dropoff trips are created in the central urban areas of Manhattan.

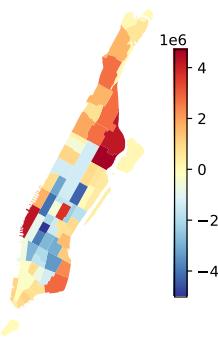


Fig. 7: Total dropoff trips minus total pickup trips in the 69 zones of Manhattan.

Fig. 7 visualizes the long-term changes of human mobility via the dropoff trips minus the pickup trips. The dropoff trips are greater than the pickup trips in edges of urban areas (see the red zones such as the Upper East Side of Manhattan). In contrast, the pickup trips are greater than the dropoff trips in the central urban areas (see the blue

zones). Therefore, it demonstrates the difference between the pickup trips and the dropoff trips in the spatial context.

As shown in Fig. 8a and 9a, the spatial and temporal modes can explain the long-term trip patterns of NYC taxi trips. The spatial mode 1 essentially demonstrates the long-term patterns, which is consistent with the total trips (see the left panel of Fig. 8a and 9a). Other spatial modes can be used to reveal the trip patterns of some specific zones. In terms of pickup trips, Fig. 8b shows both the temporal mode 2 and two trip time series curves of the highlighted zones (i.e., zones 161 and 237) in the spatial mode 2. The temporal mode 2 is informative before the COVID-19 pandemic and shows consistent patterns with the daily trips of zones 161 and 237.⁵ Fig. 8c shows both the temporal mode 4 and three trip time series curves of the highlighted zones (i.e., zones 107, 231, and 234) in the spatial mode 4. The temporal mode 4 reveals the patterns of trips before and during the COVID-19 pandemic, but it shows a great change in March 2020. Fig. 8d shows both the temporal mode 5 and two trip time series curves of the highlighted zones (i.e., zones 170 and 186). The temporal mode 5 is consistent with the daily trips of zones 170 and 186, and it also reveals the change of trips in March 2020.

In terms of the dropoff trips, Figure 9b shows both the temporal mode 2 and three trip time series curves of the highlighted zones (i.e., zones 161, 236, and 237). The temporal mode 2 reveals the (weekly) seasonal patterns of trips before the crisis of COVID-19. It is consistent with the daily trips of these zones, showing strong quasi-seasonality before the COVID-19 pandemic. It is clear to see the difference of trip patterns before and during the COVID-19 pandemic. Figure 9c shows both the temporal mode 4 and three trip time series curves of the highlighted zones (i.e., zones 79, 234, and 249). The temporal mode 4 demonstrates the trip patterns that are consistent with daily trips of these zones. Figure 9d shows both the temporal mode 5 and the trip time series curve of highlighted zone 246. The temporal mode 5 is consistent with the daily trips of zone 246 before the COVID-19 pandemic, and the pattern changes since March 2020. Remarkable traffic/trip reduction has been reported due to the travel restriction during the COVID-19 pandemic.

5. The unique zone identifiers are generated by the NYC Taxi and Limousine Commission (TLC), and the taxi zone map is available at https://www1.nyc.gov/assets/tlc/images/content/pages/about/taxi_zone_map_manhattan.jpg.

4. <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

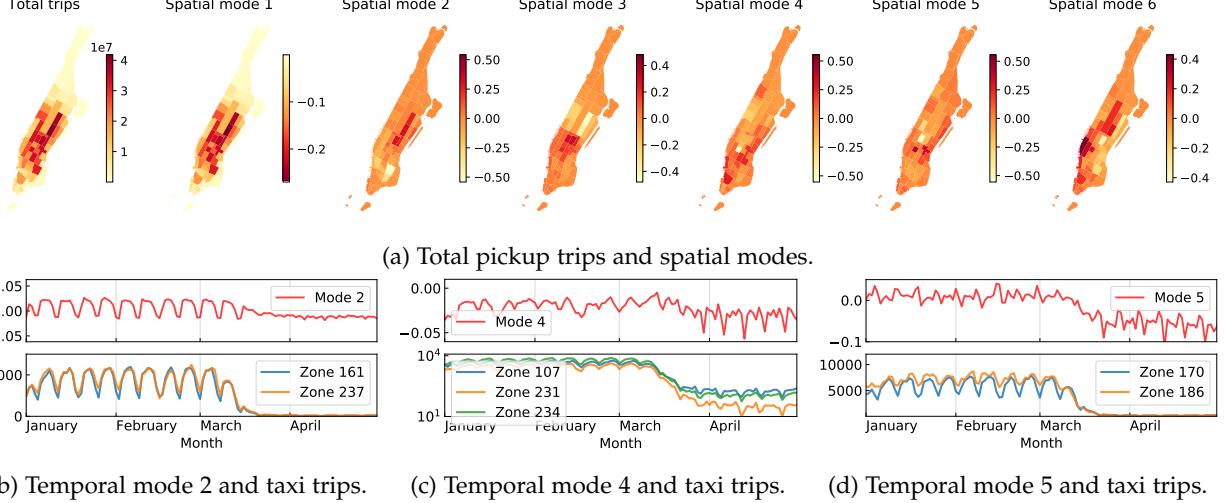


Fig. 8: NYC taxi pickup trips and their spatial and temporal modes achieved by our model. We zoom in the temporal modes in the first four months of 2020. These modes reveal the total traffic reduction due to the COVID-19 pandemic since March 2020. (a) Total trips and spatial modes revealed by \mathbf{W} . (b-d) refer to temporal mode 2, 4, 5, respectively; note that the bottom panels of these temporal modes uncover the trip time series of certain taxi zones.

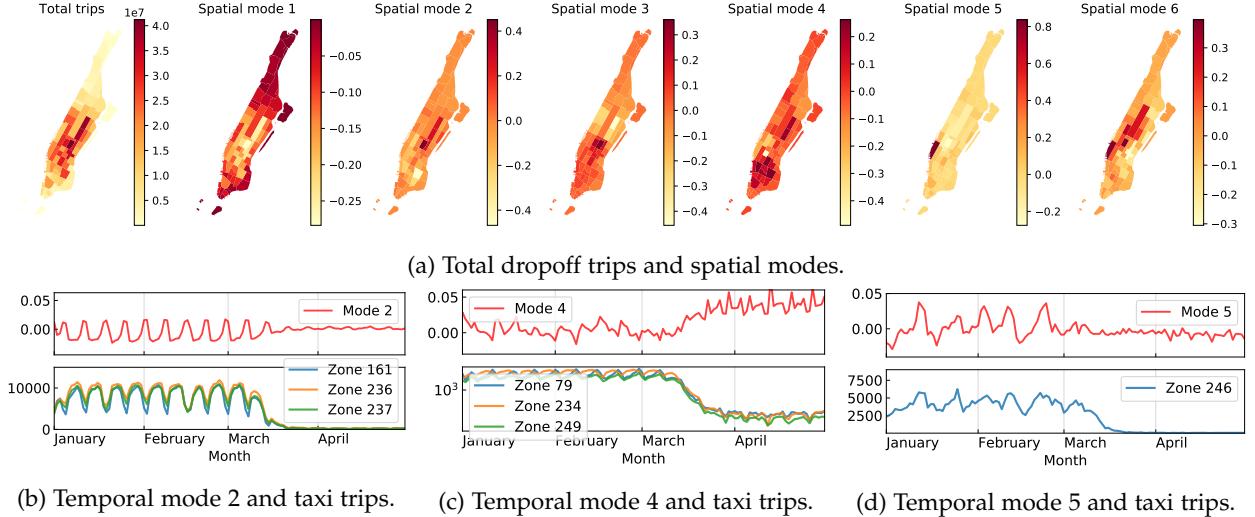


Fig. 9: NYC taxi dropoff trips and their spatial and temporal modes achieved by our model. We zoom in the temporal modes in the first four months of 2020. These modes reveal the total traffic reduction due to the COVID-19 pandemic since March 2020. (a) Total trips and spatial modes revealed by \mathbf{W} . (b-d) refer to temporal mode 2, 4, 5, respectively; note that the bottom panels of these temporal modes uncover the trip time series of certain taxi zones.

5 CONCLUSION AND DISCUSSION

Spatiotemporal data are increasingly available due to the remarkable development of sensing technologies and advanced information systems. These data provide unprecedented opportunities for discovering complex dynamic mechanisms and data patterns from nonlinear systems, and it is essential to understand them through data-driven approaches. Introducing reduced-rank VAR models such as DMD into spatiotemporal data has been demonstrated to be efficient for discovering spatiotemporal patterns [10], [11]; however, these models are sensitive to practical data noises [23] and incapable of characterizing the time-varying system behaviors of data. Therefore, it still demands data-driven frameworks that are well-suited to spatiotemporal data.

To this end, this work presents a time-varying reduced-rank VAR model for discovering interpretable modes from time series, providing insights into modeling real-world time-varying spatiotemporal systems. The model is built on the time-varying VAR and allows one to compress the over-parameterized coefficients by low-rank tensor factorization. Experiments demonstrated that the model can reveal meaningful spatial and temporal patterns underlying the time series through interpretable modes. To evaluate the performance, we test our model on several real-world spatiotemporal datasets, including fluid dynamics, SST, USA surface temperature, and NYC taxi trips. Our model can discover the underlying spatial patterns through the spatial modes and identify the time-varying system behaviors through the

temporal modes.

Last but not least, the limitations of this work are from several aspects, overcoming which would be of interest to further improve the rudimentary method proposed. For instance, it is essential to develop efficient time-varying autoregression models for spatiotemporal data in the presence of high-dimensional time series. Also, when the volume of data is too large to fit in memory all at once and arrives incrementally, adapting the time-varying autoregression models into an online learning setting has great potential to improve the efficiency [37], [38]. Additionally, forecasting is a significant task for time-series modeling [39], thus, it is essential to develop time-varying autoregression models for spatiotemporal data forecasting.

APPENDIX A DMD ON FLUID DYNAMICS

To highlight the advantages of the proposed model for discovering interpretable modes from spatiotemporal data, we consider DMD [10], [11] as an important baseline for comparison. Fig. 10 and 11 show the spatial modes and the temporal modes achieved by the DMD model, respectively. In contrast to the temporal modes achieved by our model (see Fig. 1c), the temporal modes achieved by DMD as shown in Fig. 11 is hard to reveal the time-varying system behaviors, failing for the changing dynamics in this case. Comparing to DMD, multiresolution DMD proposed by [33] still requires a well designed scheme, and it requires fixed frequency in each window. Going back to our model, it is fully time-varying, making it more flexible than the DMD models.

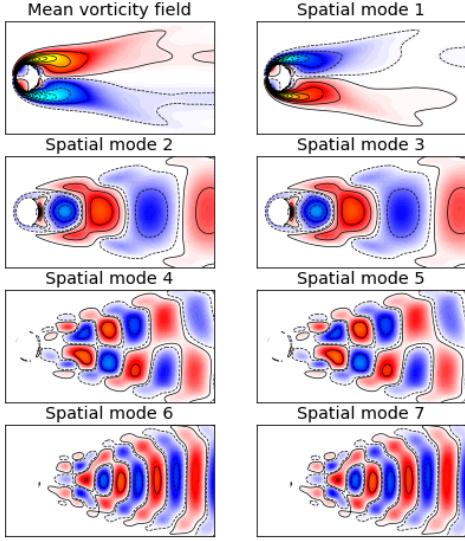


Fig. 10: Mean vorticity field and spatial modes of the fluid flow achieved by DMD with the rank $R = 7$. Note that the colorbars of all modes are on the same scale.

ACKNOWLEDGMENTS

Xinyu Chen and Chengyuan Zhang would like to thank the Institute for Data Valorisation (IVADO), Fonds de recherche

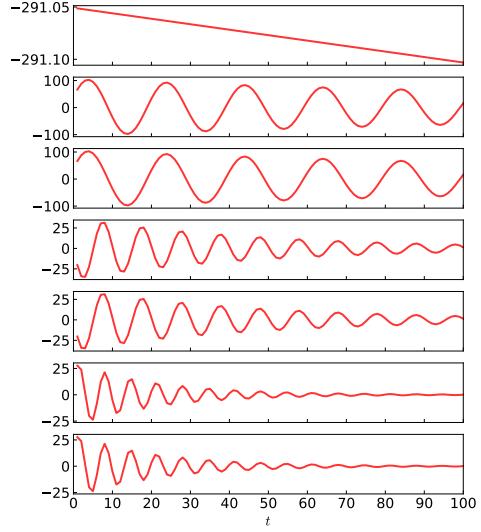


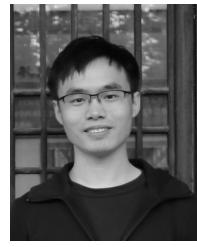
Fig. 11: Temporal modes of the fluid flow achieved by DMD. Seven panels correspond to the rank $R = 7$.

du Québec – Nature et technologies (FRQNT), and the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) for providing the PhD Excellence Scholarship to support this study. We also thank Dr. Wenshuo Wang for providing insightful suggestions for improving this work.

REFERENCES

- [1] R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, "An improved in situ and satellite sst analysis for climate," *Journal of climate*, vol. 15, no. 13, pp. 1609–1625, 2002.
- [2] M. Thornton, Y. Wei, P. Thornton, R. Shrestha, S. Kao, and B. Wilson, "Daymet: Station-level inputs and cross-validation result for north america, version 4," 2020.
- [3] Y. Zheng, "Trajectory data mining: an overview," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, pp. 1–41, 2015.
- [4] J. D. Hamilton, *Time series analysis*. Princeton university press, 1994.
- [5] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [6] A. J. Izenman, "Reduced-rank regression for the multivariate linear model," *Journal of multivariate analysis*, vol. 5, no. 2, pp. 248–264, 1975.
- [7] S. K. Ahn and G. C. Reinsel, "Nested reduced-rank autoregressive models for multiple time series," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 849–856, 1988.
- [8] R. Velu and G. C. Reinsel, *Multivariate reduced-rank regression: theory and applications*. Springer Science & Business Media, 1998, vol. 136.
- [9] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of fluid mechanics*, vol. 656, pp. 5–28, 2010.
- [10] J. H. Tu, "Dynamic mode decomposition: Theory and applications," Ph.D. dissertation, Princeton University, 2013.
- [11] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.
- [12] D. Wang, Y. Zheng, H. Lian, and G. Li, "High-dimensional vector autoregressive time series modeling via tensor decomposition," *Journal of the American Statistical Association*, pp. 1–19, 2021.
- [13] K. D. Harris, A. Aravkin, R. Rao, and B. W. Brunton, "Time-varying autoregression with low-rank tensors," *SIAM Journal on Applied Dynamical Systems*, vol. 20, no. 4, pp. 2335–2358, 2021.
- [14] M. West and J. Harrison, *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 1997.

- [15] G. E. Primiceri, "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, vol. 72, no. 3, pp. 821–852, 2005.
- [16] M. Del Negro and G. E. Primiceri, "Time varying structural vector autoregressions and monetary policy: a corrigendum," *The review of economic studies*, vol. 82, no. 4, pp. 1342–1345, 2015.
- [17] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [18] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.
- [19] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [20] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, "Tensor decompositions for learning latent variable models," *Journal of machine learning research*, vol. 15, pp. 2773–2832, 2014.
- [21] J. Wright and Y. Ma, *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications*. Cambridge University Press, 2022.
- [22] H. Zhang, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, "Online dynamic mode decomposition for time-varying systems," *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 3, pp. 1586–1609, 2019.
- [23] Z. Wu, S. L. Brunton, and S. Revzen, "Challenges in dynamic mode decomposition," *Journal of the Royal Society Interface*, vol. 18, no. 185, p. 20210686, 2021.
- [24] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.
- [25] G. Koop, D. Korobilis, and D. Pettenuzzo, "Bayesian compressed vector autoregressions," *Journal of Econometrics*, vol. 210, no. 1, pp. 135–154, 2019.
- [26] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [27] G. C. Reinsel and S. K. Ahn, "Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting," *Journal of time series analysis*, vol. 13, no. 4, pp. 353–375, 1992.
- [28] A. Carriero, G. Kapetanios, and M. Marcellino, "Structural analysis with multivariate autoregressive index models," *Journal of Econometrics*, vol. 192, no. 2, pp. 332–348, 2016.
- [29] M. T. Bahadori, Q. R. Yu, and Y. Liu, "Fast multivariate spatio-temporal analysis via low rank tensor learning," *Advances in neural information processing systems*, vol. 27, 2014.
- [30] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [31] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5239–5269, 2019.
- [32] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annual review of fluid mechanics*, vol. 25, no. 1, pp. 539–575, 1993.
- [33] J. N. Kutz, X. Fu, and S. L. Brunton, "Multiresolution dynamic mode decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 2, pp. 713–735, 2016.
- [34] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*. SIAM, 2006.
- [35] C. Deser, M. A. Alexander, S.-P. Xie, A. S. Phillips *et al.*, "Sea surface temperature variability: Patterns and mechanisms," *Annu. Rev. Mar. Sci.*, vol. 2, no. 1, pp. 115–143, 2010.
- [36] X. Chen, C. Zhang, X.-L. Zhao, N. Saunier, and L. Sun, "Nonstationary temporal matrix factorization for multivariate time series forecasting," *arXiv preprint arXiv:2203.10651*, 2022.
- [37] E. Gujral, R. Pasricha, and E. E. Papalexakis, "Sambaten: Sampling-based batch incremental tensor decomposition," in *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 2018, pp. 387–395.
- [38] E. Gujral, R. Pasricha, T. Yang, and E. E. Papalexakis, "Octen: Online compression-based tensor decomposition," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2019, pp. 455–459.
- [39] A. Afshar, I. Perros, H. Park, C. Defilippi, X. Yan, W. Stewart, J. Ho, and J. Sun, "Taste: temporal and static tensor factorization for phenotyping electronic health records," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 193–203.



Xinyu Chen

(Student Member, IEEE) received the B.S. degree in Traffic Engineering from Guangzhou University, Guangzhou, China, in 2016, and M.S. degree in Transportation Information Engineering & Control from Sun Yat-Sen University, Guangzhou, China, in 2019. He is currently a Ph.D. student with the Civil, Geological and Mining Engineering Department, Polytechnique Montreal, Montreal, QC, Canada. His current research centers on machine learning, spatiotemporal data modeling, and intelligent transportation systems.



Chengyuan Zhang

(Student Member, IEEE) is a Ph.D. student with the Department of Civil Engineering, McGill University. He received his B.S. degree in Vehicle Engineering from Chongqing University in 2019. He has worked as a visiting researcher with the Department of Mechanical Engineering, UC Berkeley in 2019–2020, and the Robotics Institute, Carnegie Mellon University in 2023. His research interests are Bayesian learning, macro/micro human driving behavior analysis, and spatiotemporal pattern recognition

in intelligent transportation systems.



Xiaoxu Chen

is a Ph.D. student with the Department of Civil Engineering, McGill University, Montreal, QC, Canada. He received the B.S. degree in transportation engineering from Harbin Institute of Technology in 2017 and the M.S. degree in transportation engineering from Tongji University, Shanghai, China, in 2020. His research currently focuses on Bayesian statistics and spatiotemporal transportation modeling.



Nicolas Saunier

received an engineering degree and a Doctorate (Ph.D.) in computer science from Telecom ParisTech, Paris, France, respectively in 2001 and 2005. He is currently a Full Professor with the Civil, Geological and Mining Engineering Department at Polytechnique Montreal, Montreal, QC, Canada. His research interests include intelligent transportation, road safety, and data science for transportation.



Lijun Sun

(Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in civil engineering (transportation) from the National University of Singapore in 2015. He is currently an Associate Professor and William Dawson Scholar in the Department of Civil Engineering, McGill University, Montreal, QC, Canada. His research centers on intelligent transportation systems, traffic control and management, spatiotemporal modeling,

Bayesian statistics, and agent-based simulation.