Statistica Sinica Preprint No: SS-2025-0419	
Title	Local Interaction Autoregressive Model for High Dimension Time Series Data
Manuscript ID	SS-2025-0419
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0419
Complete List of Authors	Jingyang Li and Yang Chen
Corresponding Authors	Jingyang Li
E-mails	jjyyli@fudan.edu.cn
Notice: Accepted author version.	

Local Interaction Autoregressive Model for High-Dimensional Time Series Data

Jingyang Li[1], Yang Chen[2]

[1] *Fudan University;* [2] *University of Michigan, Ann Arbor*

Abstract: High-dimensional matrix and tensor time series often exhibit local dependency, where each entry interacts mainly with a small neighborhood. Accounting for local interactions in a prediction model can greatly reduce the dimensionality of the parameter space, leading to more efficient inference and more accurate predictions. We propose a Local Interaction Autoregressive (LIAR) framework for high-dimensional matrix and tensor time series forecasting problems, and study Separable LIAR, a matrix variant with shared row and column components. We derive a scalable parameter estimation algorithm via parallel least squares with a BIC-type neighborhood selector. Theoretically, we show consistency of neighborhood selection and derive error bounds for kernel and auto-covariance estimation. Numerical simulations show that the BIC selector recovers the true neighborhood with high success rates, the LIAR achieves small estimation errors, and the forecasts outperform matrix time-series baselines. In a Total Electron Content (TEC) application, the proposed method identifies localized spatio-temporal propagation patterns and achieves improved prediction compared with non-local time series prediction models.

Key words and phrases: Autoregressive model; Matrix time series; Tensor time series; Local interaction; Neighborhood selection; Total Electron Content.

1. Introduction

Multivariate time series analysis is a well-established field with a rich literature of research (see e.g., Lütkepohl (2005); Tsay (2013, 2023)). Recently, there has been a growing interest in high-dimensional time series modeling. Existing research in this area can be broadly categorized into several complementary directions. The vector autoregression (VAR) framework has been extensively studied with sparsity or banded assumptions to accommodate high dimensionality while retaining interpretability (Davis et al., 2016; Guo et al., 2016; Gao et al., 2019). Building on this, matrix autoregression (MAR) models leverage the row–column structure of matrix-valued series for more efficient parameterization and interpretation (Chen et al., 2021; Hsu et al., 2021; Sun et al., 2025). Further extensions to tensor autoregressive settings capture higher-order interactions and preserve multi-way dependencies in array-valued data (Li and Xiao, 2021). Meanwhile, dynamic factor models offer an alternative route to dimension reduction by representing high-dimensional processes through a few latent dynamic factors (Lam and Yao, 2012; Wang et al., 2019; Chen et al., 2022; Chen and Fan, 2023; Han et al., 2024).

In particular, matrix- and tensor-valued time series have become increasingly common in economics, geophysics, and environmental science nowadays (Hsu et al., 2021; Chen et al., 2021; Sun et al., 2025) with the availability of high-resolution data collected at grid points over an extended amount of time. For instance, to analyze

the dynamics of multiple economic factors across various countries simultaneously, data can be organized into a matrix form (Chen et al., 2021). On the other hand, geophysical data are typically higher-dimensional. For instance, the global total electron content (TEC) distribution quantifies the electron density within the Earth's ionosphere along the vertical path between a radio transmitter and a ground-based receiver. Accurately predicting global TEC is essential for anticipating the effects of space weather on positioning, navigation, and timing (PNT) services. Each TEC map takes the form of a 181×361 matrix, organized on a spatial-temporal grid with a resolution of 1 degree in both latitude and longitude (see e.g., Sun et al., 2022, 2023). As shown in Figure 1, consecutive TEC maps evolve coherently over time, forming a high-dimensional matrix-valued time series.

While we can always resort to the well-studied vector time series analysis through vectorization, this would lead to a problem of a significantly larger size and introduce many redundant parameters. Recent studies, such as the factor autoregressive model (Chen et al., 2022; Chen and Fan, 2023), have addressed this issue, but the resulting parameters often lack clear interpretability. Additionally, the vectorization destroys the spatial information that is crucial in geophysics and environmental science. To address these challenges, several methods based on autoregressive (AR) structures specifically designed for matrix/tensor time series have been developed. For instance, Chen et al. (2021) proposed the following matrix autoregressive (MAR)

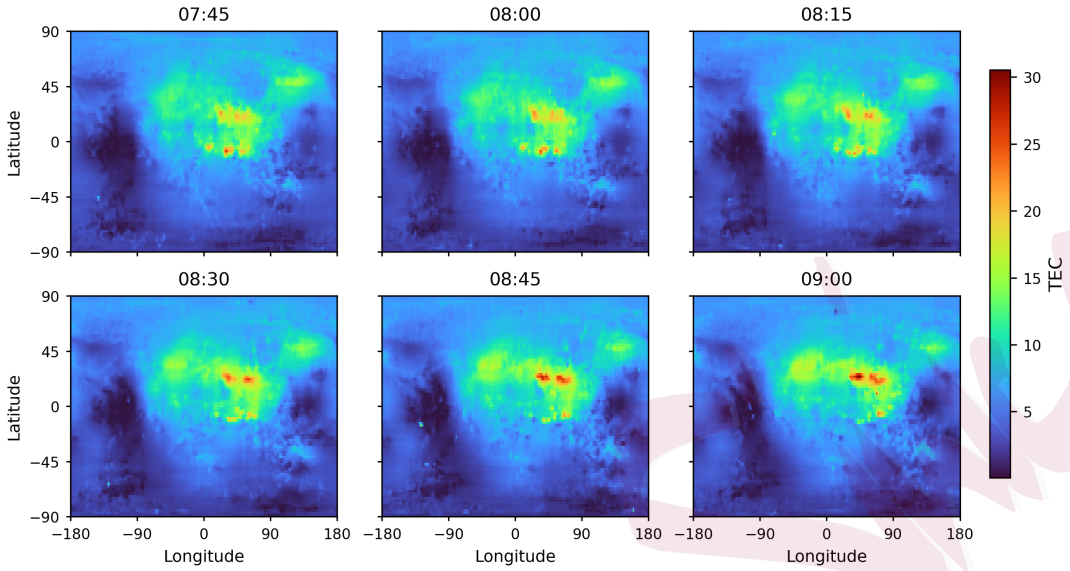


Figure 1: Consecutive global total electron content (TEC) maps from June 14, 2019, shown at 15-minute intervals.

model with a bilinear form: $\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1}\boldsymbol{B}^\top + \boldsymbol{E}_t$, where $\boldsymbol{X}_t$ is the $M \times N$ matrix. Bayesian method has also been proposed for variable selection in high-dimensional matrix autoregressive models (Celani et al., 2024). Subsequently, Sun et al. (2025) incorporated vector-valued time series data into the matrix autoregression model. However, tensor time series remains a less explored area. Li and Xiao (2021) proposed a multi-linear form for tensor time series as a generalization of the matrix case. Despite these advancements, none of these approaches fully leverage the spatial dependency in the AR coefficient matrices. In contrast, Hsu et al. (2021); Jiang et al. (2024) took spatial dependency into account by assuming $\boldsymbol{A}$ and $\boldsymbol{B}$ to be banded

matrices, which significantly decreases the number of parameters. Related rank-R matrix autoregressive structures for spatio-temporal data were further considered by Hsu et al. (2024).

Although MAR (Chen et al., 2021) reduces the number of parameters compared with vectorization, and the MAR-ST (spatio-temporal) model (Hsu et al., 2021) and banded MAR (Jiang et al., 2024) further exploits spatial structures for matrix time series, computational cost remains a significant challenge for high-dimensional data. This high cost is largely because their algorithms are inherently iterative, a requirement that stems directly from the model assumptions being bilinear in the parameters. Additionally, their model's underlying constraints can be overly stringent. Specifically, the MAR model can be understood as performing MN low-rank matrix regressions, necessitating that the coefficient matrices have a rank of 1 and share identical column subspaces across rows and identical row subspaces across columns. While this structure is manageable for the MAR model due to its adequate number of parameters, it becomes excessively restrictive when banded structures are simultaneously imposed.

In this paper, we propose a general framework for matrix- and tensor-variate autoregressive modeling that incorporates spatial structure. Our framework is flexible and covers MAR (Chen et al., 2021), MAR-ST (Hsu et al., 2021), banded MAR (Jiang et al., 2024), and tensor AR (Li and Xiao, 2021) as special cases. To incorporate the

spatial structure, we specifically focus on the Local Interaction Autoregressive (LIAR) model, which applies to both matrix and tensor data. Our model enjoys a locally adaptive dependency property in that it allows different entries to depend on different sizes of neighborhoods. In practice, the size of the neighborhood is unknown. We propose a Bayesian information criterion for the selection of the neighborhood. We show that this criterion leads to consistent neighborhood selection under various asymptotic regimes. When the neighborhoods are given, we propose parallel least squares (LS) estimators for the coefficients. We also provide asymptotic properties of the parallel LS estimator. To further reduce the number of parameters, we also consider a separable structure on the local coefficients, following related low-rank and rank-(R) autoregressive coefficient structures in the MAR and tensor autoregressive literature (Hsu et al., 2024; Li and Xiao, 2021). We refer to this variant as Separable-LIAR (SP-LIAR).

1.1 Our Contributions

Our contributions are summarized as follows. First, we introduce a unified framework for matrix and tensor time series that accommodates high dimensionality and multi-way structure in a coherent setup. It subsumes much of the existing literature as special cases. Specifically, we propose a new model that relaxes the structural constraints of MAR-ST and banded MAR while preserving the same order of param-

eters. Second, we design algorithms that are both simple to implement and suitable for parallel execution. Our algorithms provide closed-form estimators, eliminating the need for iterative processes. This greatly reduces the computational complexity (see Table 1 for the comparison with existing literature). Third, we establish asymptotic guarantees for our estimators without imposing structural assumptions on the noise, and identifiability poses no obstacle for LIAR or SP-LIAR because inference is conducted at the equivalence-class level rather than on individual components. Moreover, our neighborhood-selection criterion is consistent under mild conditions across multiple regimes, including jointly growing sample size, ambient dimension, and neighborhood size. Finally, on synthetic data, our criterion selects the true neighborhood with high success rates, kernel and auto-covariance/auto-correlation estimates are accurate, and compared with matrix time-series baselines our method achieves the lowest prediction RMSE with much shorter runtime. On large-scale total electron content (TEC) data, our approach attains competitive predictive accuracy with markedly faster runtime than existing methods, and the learned neighborhoods exhibit interpretable spatial patterns. Additional tensor-based TEC analysis and experimental results are provided in the Supplementary Material.

Model	Algorithm	Parallel?	Parameters	Computation
MAR (Chen et al., 2021)	Projection Method	No	$O(M^2)$	$O(M^4T + M^6)$
	Iterative LS	No	$O(M^2)$	$O(M^3T \cdot N_{\mathsf{iter}})$
MAR-ST (Hsu et al., 2021)	Iterative LS	No	$O(M)$	$O\big((M^3 + M^2T) \cdot N_{\mathsf{iter}}\big)$
LIAR (This paper)	Parallel LS	Yes	$O(M^2)$	$O(M^2T)$
SP-LIAR (This paper)	Parallel LS	Yes	$O(M)$	$O(M^2T + M^3)$

Table 1: Comparison between our proposed methods and existing methods for matrix autoregressive model. The matrix data is of size $M \times M$, and there are in total T samples. We assume the size of neighborhood considered in MAR-ST and our model to be constant. The computational cost displayed for MAR-ST is the minimized one when the innovation is also structured.

1.2 Organization of the Paper

The rest of the paper is organized as follows. Section 2 presents the local interaction autoregressive framework, the separable SP-LIAR variant, the parallel least-squares estimator, the neighborhood-selection criterion, and a brief extension to tensor time series. Section 3 studies the asymptotic properties of the proposed estimators and establishes neighborhood-selection consistency. Sections 4 and 5 present numerical experiments and the TEC real-data application, respectively. Further tensor implementation and tensor-based TEC experiments are provided in the Supplementary

Material.

2. Methodology

Bold lower-case (e.g., $\boldsymbol{z}$) denote vectors, bold capitals (e.g., $\boldsymbol{A}, \boldsymbol{B}$) denote matrices. All vectorization operations, denoted by $\mathrm{vec}(\cdot)$, and any implicit index orderings strictly adhere to the column-major convention. We index matrix location $\mathcal{S} = [M] \times [N]$, and write an index $\boldsymbol{i} = (i_1, i_2) \in \mathcal{S}$. Matrix entries are explicitly denoted with square brackets, as $[\boldsymbol{A}]_{\boldsymbol{i}}$. Denote $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm of matrices, and denote $\|\cdot\|_{\ell_p}$ the ℓ_p-norm of vectors or vectorized tensors for $0 \le p \le \infty$. Here $\|\boldsymbol{v}\|_{\ell_\infty}$ denotes the largest magnitude of the entries of $\boldsymbol{v}$. The spectral radius of a matrix $\boldsymbol{A} \in \mathbb{R}^{M \times M}$ (denoted by $\rho(\boldsymbol{A})$) is defined to be the maximum modulus of eigenvalues. The operator norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ (denoted by $\|\boldsymbol{A}\|$) is defined to be the largest singular value. We denote $a \vee b = \max\{a, b\}, a \wedge b = \min\{a, b\}$.

2.1 Local Interaction Autoregressive (LIAR) Model

Let $\{\boldsymbol{X}_t\}_{t \ge 0} \subset \mathbb{R}^{\mathcal{S}}$ be a matrix-valued time series. For each site $\boldsymbol{i} = (i_1, i_2) \in \mathcal{S}$ and lag $p \in [P]$, let $\mathcal{J}_{\boldsymbol{i}} \subset \mathcal{S}$ be a (location-specific) spatial neighborhood, and let $\boldsymbol{M}_{p,\boldsymbol{i}} \in \mathbb{R}^{\mathcal{S}}$ satisfy $\mathsf{supp}(\boldsymbol{M}_{p,\boldsymbol{i}}) = \mathcal{J}_{\boldsymbol{i}}$. The LIAR specification assumes that each entry

of $\boldsymbol{X}_t$ depends only on past values within its neighborhood:

$$\boldsymbol{X}_t = \sum_{\boldsymbol{i}\in\mathcal{S}} \sum_{p\in[P]} \langle \boldsymbol{X}_{t-p}, \boldsymbol{M}_{p,\boldsymbol{i}} \rangle \; \boldsymbol{e}_{i_1} \boldsymbol{e}_{i_2}^\top + \boldsymbol{E}_t, \tag{2.1}$$

where $\boldsymbol{E}_t$ is a mean-zero innovation independent of the past series, and $\boldsymbol{e}_{i_1} \in \mathbb{R}^M$ and $\boldsymbol{e}_{i_2} \in \mathbb{R}^N$ are standard basis vectors. Both the *shape* and *size* of $\mathcal{J}_{\boldsymbol{i}}$ may vary with $\boldsymbol{i}$; i.e., neighborhoods need not be rectangular. Our goal is to recover the neighborhoods $\{\mathcal{J}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$ and estimate the coefficients $\{\boldsymbol{M}_{p,\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S},p\in[P]}$.

In particular, this formulation admits a VAR representation and includes several familiar special cases:

1. *VAR representation.* Denote $\boldsymbol{x}_t = \mathsf{vec}(\boldsymbol{X}_t)$ and $\boldsymbol{\epsilon}_t = \mathsf{vec}(\boldsymbol{E}_t)$. Flattening both sides of (2.1) yields

$$\boldsymbol{x}_t \;=\; \sum_{p\in[P]} \boldsymbol{M}_p^{\mathsf{vec}} \, \boldsymbol{x}_{t-p} \;+\; \boldsymbol{\epsilon}_t,$$

   where $\boldsymbol{M}_p^{\mathsf{vec}}$ has the structured form $\boldsymbol{M}_p^{\mathsf{vec}} \;=\; \sum_{\boldsymbol{i}\in\mathcal{S}} \big(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1}\big)\mathsf{vec}(\boldsymbol{M}_{p,\boldsymbol{i}})^\top$.

2. *Banded VAR (Guo et al., 2016).* If $N = 1$ and $\mathcal{J}_i = \{u : |u - i| \le K\}$, the model reduces to a banded VAR on a length-M vector with bandwidth K.

3. *MAR (Chen et al., 2021).* Assume each local coefficient is rank one with shared components: $\boldsymbol{M}_{p,\boldsymbol{i}} \;=\; \boldsymbol{a}_{p,i_1} \, \boldsymbol{b}_{p,i_2}^\top$. Let $\boldsymbol{e}_{i_1}^\top \boldsymbol{A}_p \;=\; \boldsymbol{a}_{p,i_1}^\top$ and $\boldsymbol{e}_{i_2}^\top \boldsymbol{B}_p \;=\; \boldsymbol{b}_{p,i_2}^\top$ for $\boldsymbol{i} \;=\; (i_1, i_2) \in \mathcal{S}$. And $\mathcal{J}_{\boldsymbol{i}} = \mathcal{S}$ for all $\boldsymbol{i}$, then

$$\boldsymbol{X}_t \;=\; \sum_{p\in[P]} \boldsymbol{A}_p \, \boldsymbol{X}_{t-p} \, \boldsymbol{B}_p^\top \;+\; \boldsymbol{E}_t,$$

   which is the matrix autoregressive (MAR) model.

4. *MAR-ST (Hsu et al., 2021) or banded MAR (Jiang et al., 2024).* Under the same rank-one form $\boldsymbol{M}_{p,\boldsymbol{i}} = \boldsymbol{a}_{p,i_1}\boldsymbol{b}_{p,i_2}^\top$ with location-specific rectangular neighborhoods $\mathcal{J}_{\boldsymbol{i}} = \{(u_1,u_2) : |u_1 - i_1| \le K_{1,i_1},\ |u_2 - i_2| \le K_{2,i_2}\}$, the representation above holds with $\boldsymbol{A}_p$ and $\boldsymbol{B}_p$ banded, yielding the structured MAR.

5. *Separable LIAR (SP-LIAR).* To further reduce the number of parameters, we consider a separable structure on the local coefficients. Related low-rank and rank-R autoregressive coefficient structures have been studied in the MAR literature (Hsu et al., 2024). Specifically, let $\boldsymbol{M}_{p,\boldsymbol{i}} = \sum_{r\in[R]} \boldsymbol{a}_{p,r,i_1}\boldsymbol{b}_{p,r,i_2}^\top$, with $[\boldsymbol{a}_{p,r,i_1}]_{u_1} = 0$ if $|i_1 - u_1| > K_{1,i_1}$, and $[\boldsymbol{b}_{p,r,i_2}]_{u_2} = 0$ if $|i_2 - u_2| > K_{2,i_2}$, and $R \le \min_{ij}\{K_{1,i_1}, K_{2,i_2}\}$. Then the model can be written as

$$\boldsymbol{X}_t = \sum_{p\in[P]}\sum_{r\in[R]} \boldsymbol{A}_{p,r}\boldsymbol{X}_{t-p}\boldsymbol{B}_{p,r}^\top + \boldsymbol{E}_t, \tag{2.2}$$

where $\boldsymbol{e}_{i_1}^\top\boldsymbol{A}_{p,r} = \boldsymbol{a}_{p,r,i_1}^\top$, $\boldsymbol{e}_{i_2}^\top\boldsymbol{B}_{p,r} = \boldsymbol{b}_{p,r,i_2}^\top$ are banded matrices. Here R governs the rank of local interactions: $R = 1$ recovers MAR-ST or banded MAR, and SP-LIAR approaches the unrestricted LIAR as R increases.

In small-scale settings (e.g., macroeconomic panels), global MAR can be adequate (Chen et al., 2021), but in large-scale spatio–temporal systems it becomes computationally and statistically burdensome due to dense, high-dimensional kernels. MAR-ST (Hsu et al., 2021) and banded MAR (Jiang et al., 2024) mitigate this cost by imposing banded structure for matrices. But this rank-1 separable form restriction

limits expressiveness and cannot capture location-specific variation.

By contrast, LIAR introduces local dependence through neighborhoods $\{\mathcal{J}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$ that are allowed to vary in shape and size across locations, which captures spatially varying (location-specific) dependence while dramatically reducing the number of parameters compared with MAR. As an intermediate option, SP-LIAR imposes a low-rank separable structure within each neighborhood, offering a tunable bridge between structured MAR and the fully flexible LIAR: smaller rank yields more parsimonious models, whereas increasing the rank brings SP-LIAR closer to unrestricted LIAR while preserving locality and interpretability.

2.2 Estimating Coefficients

We now present estimation methods for LIAR, and we then extend the procedure to the separable SP-LIAR variant. For clarity of exposition, we will focus on $P = 1$ in the main text, deferring the treatment of the general lag P case to Section S3.2 of the Supplementary Material. We drop the subscript p in (2.1) and consider the following lag-1 model:

$$\boldsymbol{X}_t = \sum_{\boldsymbol{i}\in\mathcal{S}} \langle \boldsymbol{X}_{t-1}, \boldsymbol{M}_{\boldsymbol{i}} \rangle \, \boldsymbol{e}_{i_1} \boldsymbol{e}_{i_2}^\top + \boldsymbol{E}_t. \tag{2.3}$$

For each $\boldsymbol{i} \in \mathcal{S}$, we denote $\boldsymbol{x}_t^{(\boldsymbol{i})} := \boldsymbol{X}_t(\mathcal{J}_{\boldsymbol{i}}), \boldsymbol{m}_{\boldsymbol{i}} = \boldsymbol{M}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}}) \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}|}$. Here $\boldsymbol{M}(\mathcal{J}) \in \mathbb{R}^{|\mathcal{J}|}$ is the vector that collects the entries $\{[\boldsymbol{M}]_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{J}\}$ in column-major order for $\mathcal{J} \subset \mathcal{S}$. Then the scalar regression for entry $\boldsymbol{i}$ reads, for $t = 2, \ldots, T$, $[\boldsymbol{X}_t]_{\boldsymbol{i}} =$

$\langle \boldsymbol{x}_{t-1}^{(i)}, \boldsymbol{m}_i \rangle + [\boldsymbol{E}_t]_i$. Stacking over t gives the standard linear model $\boldsymbol{z}_i = \boldsymbol{Y}_i \boldsymbol{m}_i + \boldsymbol{\nu}_i$, where

$$\begin{aligned} \boldsymbol{z}_i &= \begin{bmatrix} [\boldsymbol{X}_2]_i & \cdots & [\boldsymbol{X}_T]_i \end{bmatrix}^\top, \quad \boldsymbol{\nu}_i = \begin{bmatrix} [\boldsymbol{E}_2]_i & \cdots & [\boldsymbol{E}_T]_i \end{bmatrix}^\top \in \mathbb{R}^{T-1}, \\ \boldsymbol{Y}_i &= \begin{bmatrix} \boldsymbol{x}_1^{(i)} & \cdots & \boldsymbol{x}_{T-1}^{(i)} \end{bmatrix}^\top \in \mathbb{R}^{(T-1)\times|\mathcal{J}_i|}. \end{aligned} \tag{2.4}$$

The least squares estimator is $\widehat{\boldsymbol{m}}_i = \arg\min_{\boldsymbol{m}} \frac{1}{2}\|\boldsymbol{z}_i - \boldsymbol{Y}_i \boldsymbol{m}\|_{\ell_2}^2 = (\boldsymbol{Y}_i^\top \boldsymbol{Y}_i)^{-1} \boldsymbol{Y}_i^\top \boldsymbol{z}_i$. This procedure can be carried out in parallel for different $\boldsymbol{i} \in \mathcal{S}$. We summarize the procedure in Algorithm 1.

Algorithm 1 Parallel Least Squares

Input: Data $\{\boldsymbol{X}_t\}_{t=1}^T$, local neighborhood $\{\mathcal{J}_i\}_{i\in\mathcal{S}}$

ParFor $\boldsymbol{i} \in \mathcal{S}$

Collect the covariate $\boldsymbol{Y}_i$ and response $\boldsymbol{z}_i$ defined in (2.4)

Solve the least-squares problem $\widehat{\boldsymbol{m}}_i = (\boldsymbol{Y}_i^\top \boldsymbol{Y}_i)^{-1} \boldsymbol{Y}_i^\top \boldsymbol{z}_i$

End ParFor

Output: $\{\widehat{\boldsymbol{m}}_i\}_{i\in\mathcal{S}}$

Efficiency and Computational Cost Analysis of Parallel LS. Due to the spatial locality, our parallel LS is highly efficient. Additionally, the estimator is non-iterative with a closed-form solution. Let $J = \max_i |\mathcal{J}_i|$. Then for each sub-problem, we only need to solve a least squares problem, resulting in a computational cost of order $O(J^2T + J^3)$. We note that the computation of coefficients for a single entry is

independent of the matrix data's ambient dimension. The total computational cost is $O\big(MN(J^2T + J^3)\big)$. In most applications, J is small constant, so the computational cost reduces to $O(MNT)$. Moreover, the parallel LS can be implemented concurrently, further decreasing the algorithm's runtime. In contrast, while the spatial structure was utilized in Hsu et al. (2021), their iterative least squares algorithm does not fully exploit it, leading to a significantly higher computational cost (see Table 1).

Coefficient Estimation for SP-LIAR. The estimation for the coefficients for SP-LIAR is more challenging as it admits no closed-form solution due to the separable structure. We proceed in two stages: first obtain per-location LS estimates as in LIAR, then enforce separability via a low-rank projection. In SP-LIAR, $\mathcal{J}_i$ are rectangles by construction, and we denote $\boldsymbol{M}^{[i]} := \boldsymbol{M}_i[\mathcal{J}_i]$ as a sub-matrix of $\boldsymbol{M}_i$. Arrange these matrices into the block matrix

$$\boldsymbol{M}^{\mathsf{blk}} = \begin{bmatrix} \boldsymbol{M}^{[1,1]} & \cdots & \boldsymbol{M}^{[1,N]} \\ \vdots & & \vdots \\ \boldsymbol{M}^{[M,1]} & \cdots & \boldsymbol{M}^{[M,N]} \end{bmatrix}. \tag{2.5}$$

Then under SP-LIAR, $\boldsymbol{M}^{\mathsf{blk}}$ has rank R. This motivates us to consider the best rank R approximation of the following estimator of $\boldsymbol{M}^{\mathsf{blk}}$, where $\widehat{\boldsymbol{m}_i}$ are obtained from

Algorithm 1:

$$\widehat{\boldsymbol{M}}^{\mathsf{blk}} = \begin{bmatrix} \mathsf{mat}(\widehat{\boldsymbol{m}}_{(1,1)}) & \cdots & \mathsf{mat}(\widehat{\boldsymbol{m}}_{(1,N)}) \\ \vdots & & \vdots \\ \mathsf{mat}(\widehat{\boldsymbol{m}}_{(M,1)}) & \cdots & \mathsf{mat}(\widehat{\boldsymbol{m}}_{(M,N)}) \end{bmatrix}.$$

Our algorithm is summarized in Algorithm 2.

Algorithm 2 Parallel Least Square for SP-LIAR

Input: Local neighborhood $\mathcal{J}_i$, data $\{\boldsymbol{X}_t\}_{t=1}^T$, rank R

Run Algorithm 1 for $\{\widehat{\boldsymbol{m}}_i\}_{i\in\mathcal{S}}$

Perform rank R SVD approximation on $\widehat{\boldsymbol{M}}^{\mathsf{blk}}$ and get: $\widehat{\boldsymbol{M}}_R^{\mathsf{blk}} = \mathsf{SVD}_R(\widehat{\boldsymbol{M}}^{\mathsf{blk}})$

Output: $\{\widehat{\boldsymbol{M}}_R^{[i]}\}_{i\in\mathcal{S}}$, where $\widehat{\boldsymbol{M}}_R^{[i]}$ is the (i_1, i_2)-th block of $\widehat{\boldsymbol{M}}_R^{\mathsf{blk}}$

Identifiability Issue. In the case where $R = 1$ (Chen et al., 2021), it is assumed that one of the coefficient matrices has a unit Frobenius norm. They also mentioned the identifiability issue becomes subtler when $R > 1$. In our formulation of SP-LIAR, this concern disappears as the primary quantity of interest is the equivalence class rather than two individual components.

2.3 Neighborhood Selection

We now discuss how to select the neighborhood $\mathcal{J}_i$ for LIAR. Ideally, one would hope to select from all subsets of $\mathcal{S}$. However, due to the *non-nested* model structure,

this is difficult without refining the candidate class. We illustrate this with a simple example.

Example (Non-nested ambiguity). Consider the linear regression problem $y_t = \langle \boldsymbol{X}_t, \boldsymbol{M} \rangle$, $t = 1, \cdots, T$, with $\boldsymbol{M}$ supported on $\mathcal{J} = \{(u, v) : |u - i_0| \le K_1, |v - j_0| \le K_2\}$. Consider an alternative candidate $\widetilde{\mathcal{J}} = \{(u, v) : |u - i_0| \le k_1, |v - j_0| \le k_2\}$ for some $k_1 < K_1, k_2 > K_2$ (so neither rectangle contains the other). There are natural designs $\boldsymbol{X}_t$ under which one can construct $\widehat{\boldsymbol{M}}$ supported on $\widetilde{\mathcal{J}}$ that reproduces exactly the same responses y_t as $\boldsymbol{M}$. A formal statement and proof are provided in Section S3.1 of the Supplementary Material.

Without nesting, empirical criteria cannot reliably separate such candidates. We therefore restrict the search to a nested sequence indexed by k for each location $\boldsymbol{i} = (i_1, i_2)$:

$$\mathcal{J}_{\boldsymbol{i}}[1] \subsetneq \cdots \subsetneq \mathcal{J}_{\boldsymbol{i}}[k_0] \subsetneq \cdots \subsetneq \mathcal{J}_{\boldsymbol{i}}[K_0] \subset \mathcal{S},$$

where $\mathcal{J}_{\boldsymbol{i}}[k_0] = \mathcal{J}_{\boldsymbol{i}}$, and K_0 is a prescribed cap. Write $s_k = |\mathcal{J}_{\boldsymbol{i}}[k]|$ (depends on k, and implicitly on $\boldsymbol{i}$). For any level k, let $\boldsymbol{x}_t[k] := \boldsymbol{X}_t\big(\mathcal{J}_{\boldsymbol{i}}[k]\big) \in \mathbb{R}^{s_k}$. Then we form the design matrix $\boldsymbol{Y}[k] = \begin{bmatrix} \boldsymbol{x}_1[k] & \cdots & \boldsymbol{x}_{T-1}[k] \end{bmatrix}^\top$. Also recall $\boldsymbol{z} = \begin{bmatrix} [\boldsymbol{X}_2]_{\boldsymbol{i}} & \cdots & [\boldsymbol{X}_T]_{\boldsymbol{i}} \end{bmatrix}^\top$. The residual sum of squares is $\mathsf{RSS}_{\boldsymbol{i}}(k) := \|\boldsymbol{z} - \boldsymbol{Y}[k](\boldsymbol{Y}[k]^\top \boldsymbol{Y}[k])^{-1} \boldsymbol{Y}[k]^\top \boldsymbol{z}\|_{\ell_2}^2$. And we use the Bayesian information criterion for the bandwidth selection:

$$\mathsf{BIC}_{\boldsymbol{i}}(k) = \log \mathsf{RSS}_{\boldsymbol{i}}(k) + D_0 \frac{|\mathcal{J}_{\boldsymbol{i}}[k]|}{T} \log(M \vee N \vee T). \tag{2.6}$$

For each $\boldsymbol{i} \in \mathcal{S}$, we select $\widehat{k}_{\boldsymbol{i}}$ by $\widehat{k}_{\boldsymbol{i}} = \arg\min_{k \le K_0} \mathsf{BIC}_{\boldsymbol{i}}(k)$.

Choice of D_0 and K_0. We set $D_0 = \log\log T$ when considering the regime where $T, M, N \to \infty$ while $J = \max_{\boldsymbol{i}\in\mathcal{S}} |\mathcal{J}_{\boldsymbol{i}}|$ remains fixed. The cutoff K_0 can be chosen by examining the curvature of $\mathsf{BIC}_{\boldsymbol{i}}(k)$. When J also diverges, choosing D_0 is more delicate; this high-dimensional regime is addressed in Section 3.

2.4 Extension to Tensor Time Series

The LIAR framework extends naturally to tensor-valued time series. Let $\{\boldsymbol{\mathcal{X}}_t\}_{t\ge 0} \subset \mathbb{R}^{\mathcal{S}_d}$, where $\mathcal{S}_d = [N_1] \times \cdots \times [N_d]$. For each tensor entry $\boldsymbol{i} = (i_1, \ldots, i_d) \in \mathcal{S}_d$, let $\mathcal{J}_{\boldsymbol{i}} \subset \mathcal{S}_d$ denote its local neighborhood, and let $\boldsymbol{\mathcal{M}}_{\boldsymbol{i}}$ be supported on $\mathcal{J}_{\boldsymbol{i}}$. The tensor LIAR model is

$$\boldsymbol{\mathcal{X}}_t = \sum_{\boldsymbol{i}\in\mathcal{S}_d} \langle \boldsymbol{\mathcal{X}}_{t-1}, \boldsymbol{\mathcal{M}}_{\boldsymbol{i}} \rangle \, \boldsymbol{e}_{i_1} \circ \cdots \circ \boldsymbol{e}_{i_d} + \boldsymbol{\mathcal{E}}_t.$$

After vectorization, this model can be written as a structured VAR,

$$\boldsymbol{x}_t = \boldsymbol{M}\boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{M} = \sum_{\boldsymbol{i}\in\mathcal{S}_d} (\boldsymbol{e}_{i_d} \otimes \cdots \otimes \boldsymbol{e}_{i_1}) \mathsf{vec}(\boldsymbol{\mathcal{M}}_{\boldsymbol{i}})^\top,$$

where $\boldsymbol{x}_t = \mathsf{vec}(\boldsymbol{\mathcal{X}}_t)$ and $\boldsymbol{\epsilon}_t = \mathsf{vec}(\boldsymbol{\mathcal{E}}_t)$. Writing $\boldsymbol{x}_t^{(\boldsymbol{i})} = \boldsymbol{\mathcal{X}}_t(\mathcal{J}_{\boldsymbol{i}})$ and $\boldsymbol{m}_{\boldsymbol{i}} = \boldsymbol{\mathcal{M}}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}})$, the scalar regression for entry $\boldsymbol{i}$ is $[\boldsymbol{\mathcal{X}}_t]_{\boldsymbol{i}} = \langle \boldsymbol{x}_{t-1}^{(\boldsymbol{i})}, \boldsymbol{m}_{\boldsymbol{i}} \rangle + [\boldsymbol{\mathcal{E}}_t]_{\boldsymbol{i}}$. Thus, coefficient estimation again reduces to a local least-squares problem. The BIC neighborhood selector and the matrix-case theoretical arguments carry over analogously after replacing

matrix neighborhoods by tensor neighborhoods. The detailed tensor least-squares implementation is given in the Supplementary Material.

3. Theoretical Guarantees

In this section, we study the asymptotic property of the estimators $\{\widehat{\boldsymbol{m}_{\boldsymbol{i}}}\}_{\boldsymbol{i}\in\mathcal{S}}$ from Algorithm 1. We define

$$J_{\square} = \min\Big\{(2k_1+1)(2k_2+1) : \mathcal{J}_{\boldsymbol{i}} \subset \{i_1-k_1,\cdots,i_1+k_1\}\times \{i_2-k_2,\cdots,i_2+k_2\},\ \forall \boldsymbol{i}\Big\}.$$

Here $J_{\square}$ is the size of the minimal rectangular that uniformly contains all supports, directly capturing geometric locality. The asymptotic regimes are: Regime 1, only $T\to\infty$, while $M, N, J_{\square}$ remain fixed; Regime 2, $T, M, N \to \infty$, while $J_{\square}$ remains fixed; and Regime 3, $T, M, N, J_{\square} \to \infty$. We consider the case when $P = 1$, and then (2.3) can be equivalently written as $\boldsymbol{x}_t = \boldsymbol{M}\boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{M} = \sum_{\boldsymbol{i}\in\mathcal{S}}(\boldsymbol{e}_{i_2}\otimes \boldsymbol{e}_{i_1})\mathsf{vec}(\boldsymbol{M}_{\boldsymbol{i}})^\top$, where $\boldsymbol{x}_t = \mathsf{vec}(\boldsymbol{X}_t)$, and $\boldsymbol{\epsilon}_t = \mathsf{vec}(\boldsymbol{E}_t)$.

3.1 Consistency and Asymptotic Normality

When the spectral radius $\rho(\boldsymbol{M}) < 1$, (2.3) is a stationary matrix time series and $\boldsymbol{X}_t$ admits the following expansion in terms of the innovations: $\boldsymbol{x}_t = \sum_{l=0}^{\infty}\boldsymbol{M}^l\boldsymbol{\epsilon}_{t-l}$.

Assumption 1. There exist $q, \widetilde{q} > 0, \delta \in (0,1)$, and $\mu_{2q} > 0$ independent of $T, M, N, J_\square$, such that the following holds under different regimes:

(1) Regime 1 (classical): $\rho(\boldsymbol{M}) < 1$; the innovation process $\{\boldsymbol{E}_t\}$ are i.i.d. with zero mean, and each entry has a finite $2+\widetilde{q}$ moments, $\max_{ij} \|[\boldsymbol{E}_t]_{\boldsymbol{i}}\|_{2+\widetilde{q}} < +\infty$;

(2) Regime 2 (growing dimension): $\|\boldsymbol{M}\| \le \delta$; the innovation process $\{\boldsymbol{E}_t\}$ are i.i.d. with zero mean, and $\max_{ij} \|[\boldsymbol{E}_t]_{\boldsymbol{i}}\|_{2q} \le \mu_{2q}$ for some $q > 2$; and $MN = O(T^\beta)$, where $\beta \in (0, \frac{q-2}{4})$;

(3) Regime 3 (growing dimension and locality): in addition to (2), $T \ge \widetilde{C}_1 J_\square^4 \log(M \vee N \vee T)$, for some $\widetilde{C}_1 > 0$ depending only on δ and μ_{2q}.

These three regimes can be viewed as increasingly high-dimensional extensions of standard autoregressive theory. Regime 1 corresponds to the classical fixed-dimensional time-series setting. Regime 2 allows the ambient dimension MN to grow while the uniform locality size $J_\square$ remains fixed. The restriction $\beta < (q-2)/4$ comes from controlling sample covariance terms uniformly over the growing dimension under bounded $2q$-th moments, and is the same type of dimension-growth condition as that used in high-dimensional banded VAR models Guo et al. (2016). Regime 3 further allows the locality size $J_\square$ to increase. The lower bound $T \gtrsim J_\square^4 \log(M \vee N \vee T)$ ensures that the local sample Gram matrices remain uniformly well-conditioned when the number of local regressors also grows.

The regimes in which M, N, and possibly $J_\square$, diverge are more challenging. Given the fact $\rho(\boldsymbol{M}) \leq \|\boldsymbol{M}\|$, Regime 2 and Regime 3 require stronger conditions on both the coefficient matrix and the moments on the innovation process. Let the auto-correlation $\boldsymbol{\Gamma}_0 := \mathsf{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_t)$. We denote $\boldsymbol{\Gamma}_0(\mathcal{J}_i; \mathcal{J}_i) \in \mathbb{R}^{|\mathcal{J}_i| \times |\mathcal{J}_i|}$ the sub-matrix of $\boldsymbol{\Gamma}_0$ by extracting the elements corresponding to the row (column) indices in $\mathcal{J}_i$, arranged using column-major ordering. The following theorem states the asymptotic property for $\widehat{\boldsymbol{m}}_i$ under either regimes.

Theorem 1. *Suppose Assumption 1 holds. Then we have*

$$\sqrt{T} \cdot (\widehat{\boldsymbol{m}}_i - \boldsymbol{m}_i) \implies N\big(0, \sigma_i^2 \cdot [\boldsymbol{\Gamma}_0(\mathcal{J}_i; \mathcal{J}_i)]^{-1}\big)$$

under any of Regimes 1–3, where $\sigma_i^2 = \mathsf{Var}([\boldsymbol{E}_t]_i)$.

This result does not require extra structure on the full innovation covariance as in (Chen et al., 2021; Sun et al., 2025). Our theorem implies that the coefficients $\widehat{\boldsymbol{m}}_i$ have a convergence rate of $\sqrt{T}$, which matches known results for matrix autoregression (Chen et al., 2021) and matrix autoregression model with auxiliary covariate in (Sun et al., 2025), and continues to hold when M, N, even $J_\square$ grow.

Inference based on Theorem 1. Theorem 1 also provides a direct way to construct confidence intervals for the local autoregressive coefficients. For a fixed loca-

tion $\boldsymbol{i} \in \mathcal{S}$, define the plug-in estimators

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{i}} = \frac{1}{T-1}\boldsymbol{Y}_{\boldsymbol{i}}^{\top}\boldsymbol{Y}_{\boldsymbol{i}}, \qquad \widehat{\sigma}_{\boldsymbol{i}}^2 = \frac{\|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}\widehat{\boldsymbol{m}}_{\boldsymbol{i}}\|_{\ell_2}^2}{T-1-|\mathcal{J}_{\boldsymbol{i}}|}.$$

For any fixed vector $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}|}$, since $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{i}} \xrightarrow{p} \boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{i}})$ under the same regimes as in Theorem 1 and $\widehat{\sigma}_{\boldsymbol{i}}^2$ is a consistent estimator of $\sigma_{\boldsymbol{i}}^2$, Slutsky's theorem gives

$$\frac{\boldsymbol{a}^{\top}(\widehat{\boldsymbol{m}}_{\boldsymbol{i}} - \boldsymbol{m}_{\boldsymbol{i}})}{\widehat{\sigma}_{\boldsymbol{i}}\{\boldsymbol{a}^{\top}(\boldsymbol{Y}_{\boldsymbol{i}}^{\top}\boldsymbol{Y}_{\boldsymbol{i}})^{-1}\boldsymbol{a}\}^{1/2}} \implies N(0,1).$$

Therefore, an asymptotic $1-\alpha$ confidence interval for the linear functional $\boldsymbol{a}^{\top}\boldsymbol{m}_{\boldsymbol{i}}$ is

$$\boldsymbol{a}^{\top}\widehat{\boldsymbol{m}}_{\boldsymbol{i}} \pm z_{1-\alpha/2}\widehat{\sigma}_{\boldsymbol{i}}\{\boldsymbol{a}^{\top}(\boldsymbol{Y}_{\boldsymbol{i}}^{\top}\boldsymbol{Y}_{\boldsymbol{i}})^{-1}\boldsymbol{a}\}^{1/2},$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution. In particular, for the j-th coordinate of $\boldsymbol{m}_{\boldsymbol{i}}$, by taking $\boldsymbol{a} = \boldsymbol{e}_j$, we obtain $\widehat{\boldsymbol{m}}_{\boldsymbol{i},j} \pm z_{1-\alpha/2}\widehat{\sigma}_{\boldsymbol{i}}\left\{[(\boldsymbol{Y}_{\boldsymbol{i}}^{\top}\boldsymbol{Y}_{\boldsymbol{i}})^{-1}]_{j,j}\right\}^{1/2}$. When the neighborhood is selected by the BIC criterion in (2.6), we use the selected neighborhood $\mathcal{J}_{\boldsymbol{i}}[\widehat{k}_{\boldsymbol{i}}]$ in the above construction. Under the selection consistency result in Theorem 5, this plug-in confidence interval has the same first-order asymptotic validity as the oracle interval based on the true neighborhood $\mathcal{J}_{\boldsymbol{i}}$.

Since $\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$ are estimated based on possibly overlapped data, they are dependent to each other. In order to state the joint asymptotic behavior, we adopt column-major stacking operators: $\mathsf{vecstack}\{\boldsymbol{v}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\}$ denotes the vector formed by concatenating $\boldsymbol{v}_{\boldsymbol{i}}$ in column-major order, and $\mathsf{blkstack}\{\boldsymbol{M}_{\boldsymbol{i},\boldsymbol{j}} : \boldsymbol{i}, \boldsymbol{j} \in \mathcal{S}\}$ denotes

the block matrix whose block rows and block columns follow the same column-major order.

Theorem 2. *Suppose Assumption 1 holds. Then we have*

$$\sqrt{T} \cdot \mathsf{vecstack}\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}} - \boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} \implies N(0, \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1})$$

under any of Regimes 1–3*, where the matrices* $\boldsymbol{C}, \boldsymbol{D}$ *are arranged in a column-major order*

$$\boldsymbol{C} = \mathsf{blkstack}\{[\boldsymbol{\Sigma}]_{\boldsymbol{i},\boldsymbol{j}} \cdot \boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{j}}) : \boldsymbol{i}, \boldsymbol{j} \in \mathcal{S}\}, \quad \boldsymbol{D} = \mathsf{blkdiag}\{\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{i}}) : \boldsymbol{i} \in \mathcal{S}\},$$

and $\boldsymbol{\Sigma} = \mathsf{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t)$.

For SP-LIAR, we denote the compact rank R SVD of $\boldsymbol{M}^{\mathsf{blk}}$ (defined in (2.5)) by $\boldsymbol{M}^{\mathsf{blk}}$ be $\boldsymbol{M}^{\mathsf{blk}} = \boldsymbol{U}^{\mathsf{blk}}\boldsymbol{\Sigma}^{\mathsf{blk}}\boldsymbol{V}^{\mathsf{blk}\top}$. Note that the vectorization of $\boldsymbol{M}^{\mathsf{blk}}$ differs from $\mathsf{vecstack}\{\boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\}$ a permutation matrix $\boldsymbol{Q}$, that is $\mathsf{vec}(\boldsymbol{M}^{\mathsf{blk}}) = \boldsymbol{Q} \cdot \mathsf{vecstack}\{\boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\}$. We have the following asymptotic result.

Theorem 3. *Suppose Assumption 1 holds. Then we have*

$$\sqrt{T} \cdot \mathsf{vec}(\widehat{\boldsymbol{M}}_R^{blk} - \boldsymbol{M}^{blk}) \implies N(0, \boldsymbol{P}\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}\boldsymbol{Q}^\top\boldsymbol{P})$$

under any of Regimes 1–3*, where* $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{V}_\perp^{blk}\boldsymbol{V}_\perp^{blk\top} \otimes \boldsymbol{U}_\perp^{blk}\boldsymbol{U}_\perp^{blk\top}$.

When there is separable structure with the coefficient matrices, the estimator $\widehat{\boldsymbol{M}}_R^{\mathsf{blk}}$ has the same convergence rate but the covariance matrix is smaller in the sense $\boldsymbol{P}\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}\boldsymbol{Q}^\top\boldsymbol{P} \leq \boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}\boldsymbol{Q}^\top$ since $\boldsymbol{P}$ is a projection matrix.

3.2 Estimation for Auto-Correlation

In this section, we consider the estimation for the auto-correlation. We consider the following estimator $\widehat{\mathbf{\Gamma}}_0 = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^\top$. And we have the following entry-wise bound for $\|\widehat{\mathbf{\Gamma}}_0 - \mathbf{\Gamma}_0\|_{\ell_\infty}$, where $\|\boldsymbol{A}\|_{\ell_\infty} := \|\mathsf{vec}(\boldsymbol{A})\|_{\ell_\infty}$ for any matrix $\boldsymbol{A}$.

Theorem 4. *Under Assumption 1, we have*

$$\|\widehat{\mathbf{\Gamma}}_0 - \mathbf{\Gamma}_0\|_{\ell_\infty} = O_p\big((T^{-1}\log(M \vee N \vee T))^{1/2} J_\square\big)$$

under any of Regimes 1–3.

The proof of Theorem 4 is given in Section S1.4 of the Supplementary Material. Under Regime 1 or 2, the entry-wise convergence rate is of order $\widetilde{O}(1/\sqrt{T})$, where $\widetilde{O}$ hides the logarithm factor. In either regimes, $\widehat{\mathbf{\Gamma}}_0$ converges in probability to $\mathbf{\Gamma}_0$.

3.3 Consistency of the Neighborhood Selection

We denote $J_\infty = \max\big\{J_\square, \max_{\boldsymbol{i}\in\mathcal{S}} |\mathcal{J}_{\boldsymbol{i}}[K_0]|\big\}$. Also recall $\mathcal{J}_{\boldsymbol{i}}[k_0] = \mathcal{J}_{\boldsymbol{i}}$ is the ground-truth. In order to establish the consistency of $\widehat{k}_{\boldsymbol{i}}$, we need the following assumption:

Assumption 2. There exists $\kappa_1 > 0$ independent of T, M, N, J_∞, such that

(1) $\lambda_{\min}(\mathbf{\Gamma}_0) \geq \kappa_1$;

(2) For each $\boldsymbol{i} \in \mathcal{S}$, there exists $\widetilde{C}_2$ depending only on κ_1, μ_{2q}:

$$\|\boldsymbol{M}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}} \backslash \mathcal{J}_{\boldsymbol{i}}[k_0 - 1])\|_{\mathrm{F}} \geq \left(\widetilde{C}_2 D_0 \frac{|\mathcal{J}_{\boldsymbol{i}}|}{T}\log(M \vee N \vee T)\right)^{1/2};$$

(3) Under Regime 3, in addition to (2) in Assumption 1, $T \geq \widetilde{C}_1 J_\infty^4 \log(M \vee N \vee T)$ for some $\widetilde{C}_1 > 0$ depending only on δ, μ_{2q}.

Here the first condition in Assumption 2 guarantees that the auto-correlation $\mathsf{Cov}(\boldsymbol{x}_t)$ is strictly positive definite. The second condition ensures that the index set $\mathcal{J}_{\boldsymbol{i}}[k_0]$ is asymptotically identifiable, and

$$\left(\widetilde{C}_2 D_0 \frac{|\mathcal{J}_{\boldsymbol{i}}|}{T} \log(M \vee N \vee T)\right)^{1/2}$$

is the minimum order of the non-zero coefficients. And in the third assumption, we need a slightly stronger assumption on the rate at which J_∞ grows.

Together with this assumption, we can guarantee the consistency of the selection of neighborhood. Recall $J = \max_{\boldsymbol{i}} |\mathcal{J}_{\boldsymbol{i}}|$.

Theorem 5. *Suppose Assumption 1, 2 holds. As long as we choose $D_0 \geq \widetilde{C} J^2 J_\square$ for some constant $\widetilde{C}$ depending only on $\kappa_1, \delta, \mu_{2q}$, we have for each $\boldsymbol{i} \in \mathcal{S}$, $\mathbb{P}(\widehat{k}_{\boldsymbol{i}} = k_0) \to 1$ under any of Regimes 1–3.*

Under Regime 1 or 2, when $J_\square$ remains fixed, it suffices to set $D_0 = \log(\log T)$. However, Theorem 5 allows $J_\square$ to go to infinity as well. In this case, we need to set $D_0 = C_J \log(\log T)$, where $C_J \geq J^2 J_\square$. Although J is not known in practice, we can choose C_J to be large and the condition in Theorem 5 is satisfied. However, a too large choice of C_J would lead to a stricter marginal condition in Assumption 2.

4. Numerical Experiments

In this section, we present a series of numerical experiments to demonstrate the effectiveness of our proposed methodology. Throughout, we focus on both model selection and estimation accuracy, and we compare against several existing benchmarks. Unless otherwise noted, we conduct 50 independent trials for each experiment.

Our first experiment evaluates the entrywise BIC in (2.6) for neighborhood-size selection under the lag-one LIAR model (2.3):

$$\boldsymbol{X}_t = \sum_{\boldsymbol{i}} \langle \boldsymbol{X}_{t-1}, \boldsymbol{M}_{\boldsymbol{i}} \rangle \boldsymbol{e}_{i_1} \boldsymbol{e}_{i_2}^\top + \boldsymbol{E}_t,$$

where $\boldsymbol{M}_{\boldsymbol{i}}$ is supported on the region with center at $\boldsymbol{i}$ and neighborhood size $K = 3$ (although our framework allows the neighborhood size to vary across locations, in this simulation we set uniform K for all pixels). For data generation, the nonzero local coefficients were generated within the true neighborhood and then rescaled to ensure stationarity of the resulting time series. To maintain a clear signal at the boundary of the true neighborhood, we used slightly larger coefficients on the outermost ring of the true neighborhood. The innovations had independent mean-zero Gaussian entries with standard deviation 0.12, and the first 600 observations were discarded as burn-in. We vary the size of the matrix (M, N) from $\{(10, 10), (15, 15), (20, 20)\}$, and the sample size $T \in \{600, 800, 1000\}$. For each configuration, the entrywise BIC is applied over the candidate neighborhood sizes $\{0, 1, 2, 3, 4, 5\}$ and success is

recorded when the selected neighborhood size equals the truth $K = 3$.

Figure 2 displays the success rates for each (M, N) and T combination (together with a per-pixel success-rate heatmap). The results show that the BIC selector remains informative in these moderate sample settings. The success rate increases with T but declines as the size (M, N) grows. The per-pixel heatmap further shows slightly lower rates near the boundaries than in the interior. This phenomenon can be explained by the edge effect: boundary pixels have truncated neighborhoods, weakening the signal for identifying K and leading BIC to favor smaller sizes; interior pixels, with full neighborhoods, provide stronger signal.

Our second experiment evaluates estimation accuracy under the lag-one LIAR model with $P = 1$, $(M, N) = (10, 10)$, and true neighborhood size $K = 3$. We vary the sample size $T \in \{500, 1000, 1500, \ldots, 5000\}$. We consider three quantities: (i) kernel estimation error (Frobenius norm), (ii) auto-covariance $\mathbf{\Gamma}_0$ error (Frobenius norm), and (iii) auto-covariance $\mathbf{\Gamma}_0$ error (entrywise ℓ_∞ norm).

Figure 3 displays the log-scale errors versus T and the ratio $\|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}_0\|_{\mathrm{F}}/\|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}_0\|_{\ell_\infty}$. All three errors decrease as T increases. The ratio remains approximately constant (about 23) for larger T, indicating that the error is broadly distributed across entries rather than concentrated on a few. Figure 4 plots the means of the three errors against $1/\sqrt{T}$. The errors scale approximately linearly with $1/\sqrt{T}$, especially for larger T, in line with the theoretical $T^{-1/2}$ rate and supporting consistency under

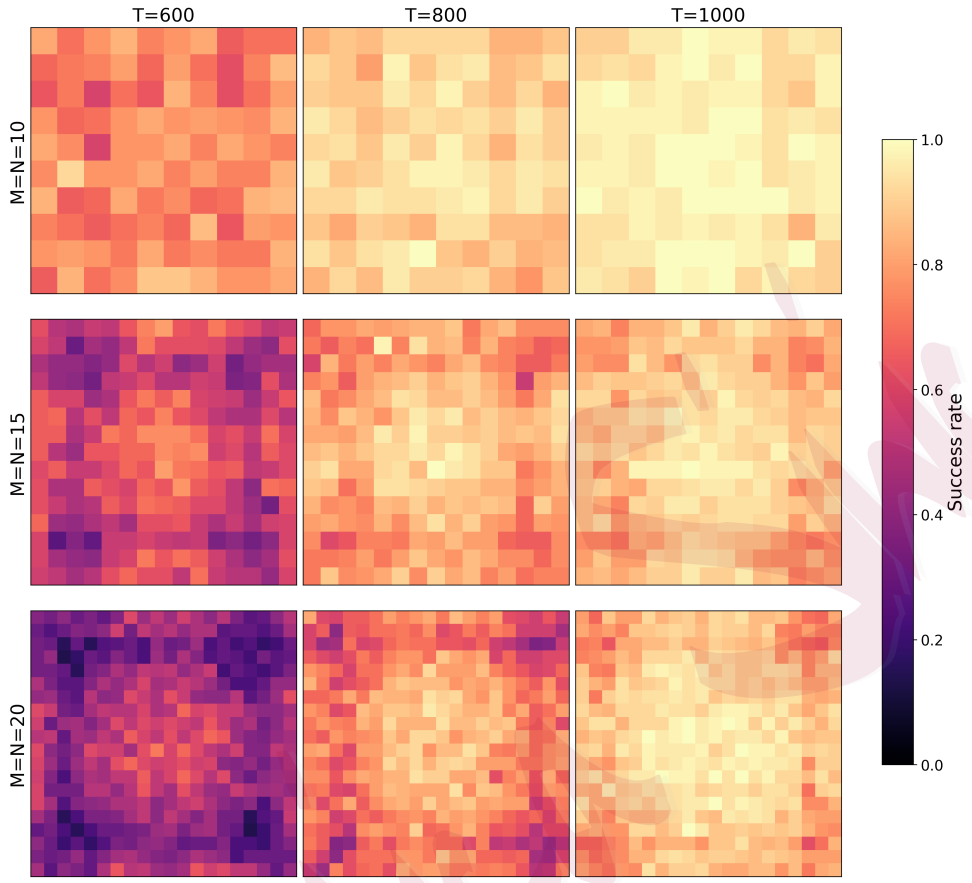


Figure 2: Success rates for $M = N \in \{10, 15, 20\}$ and $T \in \{600, 800, 1000\}$.

large samples.

In the third experiment, we compare the performance of our proposed model LIAR and its separable variant SP-LIAR against three existing methods: (1) MAR (Chen et al., 2021), $\boldsymbol{X}_t = \sum_{p=1}^{P} \boldsymbol{A}_p \boldsymbol{X}_{t-p} \boldsymbol{B}_p^\top + \boldsymbol{E}_t$; (2) MAR-ST (Hsu et al., 2021), $\boldsymbol{X}_t = \sum_{p=1}^{P} \boldsymbol{A}_p \boldsymbol{X}_{t-p} \boldsymbol{B}_p^\top + \boldsymbol{E}_t$, where $\boldsymbol{A}_p, \boldsymbol{B}_p$ are banded matrices; and (3) pixel-wise autoregression (LIAR-P), $[\boldsymbol{X}_t]_{\boldsymbol{i}} = \sum_{p=1}^{P} \beta_{\boldsymbol{i}p} [\boldsymbol{X}_{t-p}]_{\boldsymbol{i}} + [\boldsymbol{E}_t]_{\boldsymbol{i}}$ for each $\boldsymbol{i} \in \mathcal{S}$. The

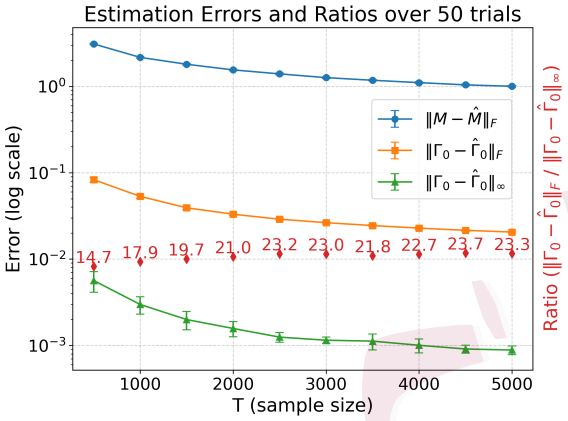


Figure 3: Log scale of estimation errors of the kernels and auto-covariance $\mathbf{\Gamma}_0$ in Frobenius and infinity norms versus T. The ratio of errors for the auto-covariance is also displayed.

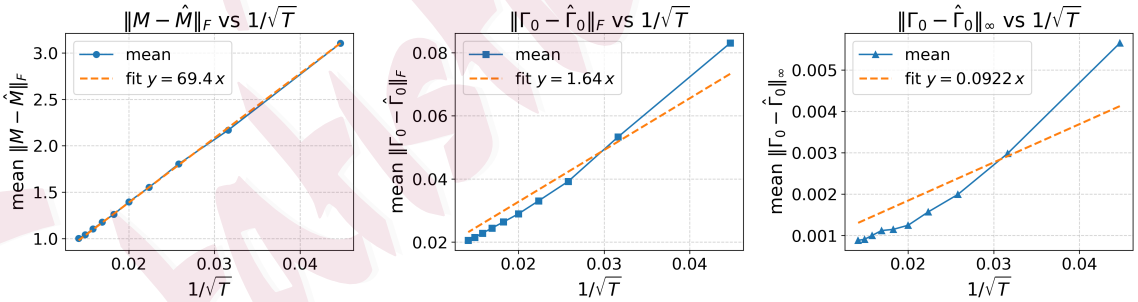


Figure 4: Mean of the estimation errors of the kernels and auto-covariance $\mathbf{\Gamma}_0$ in Frobenius and infinity norms versus $1/\sqrt{T}$.

last model can be seen as a special case of LIAR with $\mathcal{J}_i = \{i\}$. We set $(M, N) = (20, 20)$, $T = 200$, and true neighborhood size $K = 2$. The observations are split chronologically, with the first 90% used for model fitting and the remaining 10% used for testing. Prediction accuracy is measured by one-step-ahead full-matrix RMSE: at each test time point t, the fitted model uses the observed previous lag(s) to form $\widehat{\boldsymbol{X}}_t$, which is compared with the observed $\boldsymbol{X}_t$; the RMSE is then computed over all matrix entries and all test time points. For MAR-ST we fix the bandwidth at $K = 2$. For MAR and MAR-ST we cap the iterations at 50 with tolerance 10^{-7}.

We report prediction RMSE and runtime. As shown in Figure 5, LIAR is substantially faster than MAR and MAR-ST while achieving a markedly smaller prediction RMSE. LIAR-P is faster than LIAR but attains an RMSE that is about 7% higher.

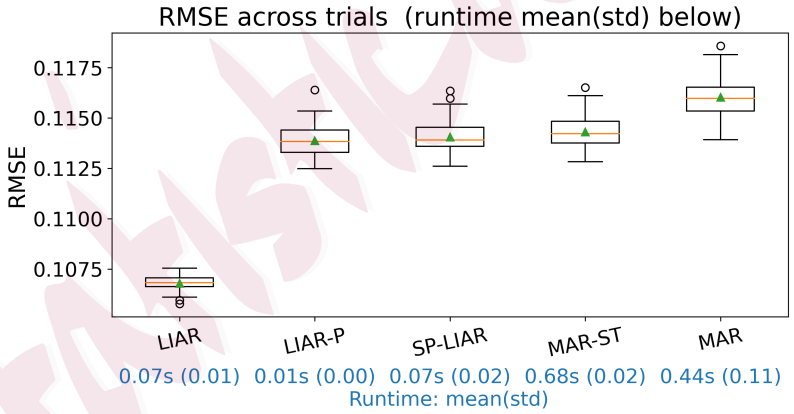


Figure 5: RMSE and runtime comparison of LIAR, SP-LIAR, MAR, MAR-ST, and LIAR-P.

5. Real data: Total Electron Content

In this section, in addition to prediction accuracy and computational cost, we examine what the fitted local interaction structure reveals about TEC dynamics. We use the BIC-selected neighborhoods as data-driven summaries of local spatio-temporal dependence and compare them across different time periods. Additional tensor-based TEC experiments are reported in the Supplementary Material.

5.1 LIAR for Matrix Time Series

We analyze ionospheric total electron content (TEC) fields derived from multi-frequency Global Navigation Satellite System (GNSS) signals, a widely used proxy in space-weather and ionospheric studies. We use the completed TEC product of Sun et al. (2023). The original data are sampled on a $1^\circ \times 1^\circ$ (latitude $\times$ longitude) grid every 5 minutes (181×361 per time point). For preprocessing, we aggregate the data to $2^\circ \times 2^\circ$ and 15-minute resolution by averaging non-overlapping 2×2 spatial blocks and averaging consecutive three-frame windows in time, yielding a 91×181 matrix time series at each time point.

Neighborhood selection. We analyze six representative 10-day periods: {201706, 201709, 201806, 201903, 201906, 201912}. For each month, we use TEC data from the 11th to the 20th (inclusive), yielding $T = 960$ time points at 15-minute resolution.

For each period, we apply the entrywise BIC to select the neighborhood size over square candidates $K \in \{0, 1, 2, 3, 4, 5\}$. Figure 7a summarizes the selections; across all six periods, the most frequently chosen sizes are $K = 1$ or $K = 2$. The concentration on small neighborhoods suggests that short-term TEC evolution is mainly driven by localized spatial propagation. The variation across locations further indicates that the strength and spatial range of local TEC dependence are spatially heterogeneous. Figure 6 illustrates the connection between the raw TEC evolution and the fitted local interaction structure for a representative period on June 14, 2019. The first six panels show consecutive TEC fields at 15-minute intervals, and the last panel shows the corresponding BIC-selected neighborhood size K. The white boundaries separate regions with $K < 2$ from those with $K \geq 2$. The selected neighborhood sizes are broadly related to local TEC behavior: relatively stable regions, such as parts of the polar areas and the dark low-TEC region in the left-central part of the maps, tend to have smaller neighborhoods, whereas regions with more visible temporal and spatial variation tend to have broader neighborhoods. Thus, the selected neighborhood map gives an interpretable summary of the spatially heterogeneous local dependence learned by LIAR.

For cross-period comparison, we coarsen the selected neighborhood sizes into two categories, $K < 2$ and $K \geq 2$, separating the most localized selections from relatively broader local neighborhoods. For each anchor period, we compare it with

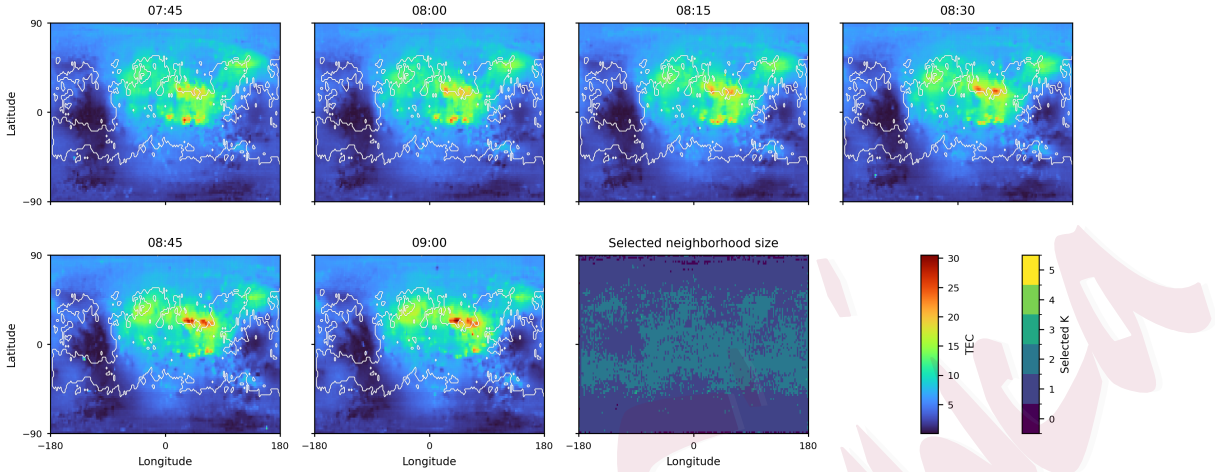
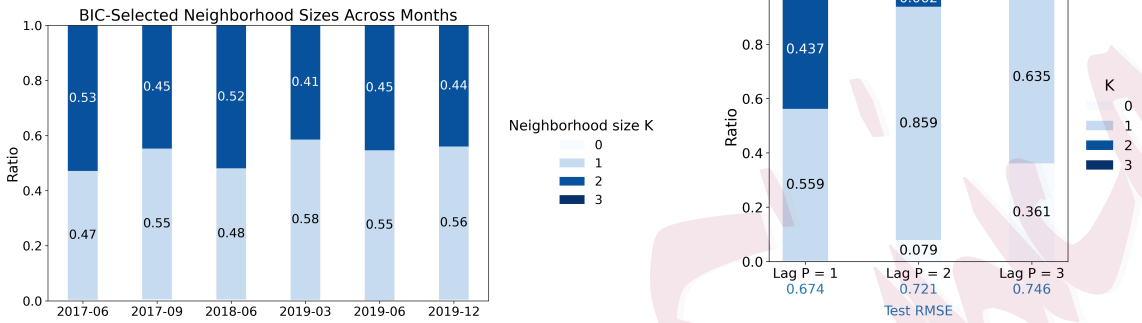


Figure 6: Representative TEC fields over consecutive 15-minute time points and the corresponding BIC-selected neighborhood-size map. For visualization, the boundary overlay is slightly smoothed to highlight the large-scale spatial pattern.

each of the other five periods by computing (i) Cohen's κ between the corresponding two-category maps and (ii) the Pearson correlation between the corresponding raw TEC fields. The results are shown in Figure 8.

In short, Figures 7a and 8 show two points: (i) The share of pixels with $K < 2$ versus $K \geq 2$ is relatively stable across periods (Figure 7a). (ii) There is a clear positive association between the Pearson correlation of the raw TEC fields and Cohen's κ of the neighborhood maps (Figure 8). The positive association suggests that the selected local interaction structure is related to the underlying ionospheric state: when the TEC fields are similar across periods, the *local* predictive dependence learned by

LIAR is also similar. Thus, the fitted LIAR model gives an interpretable summary of how the local dependence structure changes with the TEC field.



(a) Distribution of BIC-selected neighborhood sizes across six time periods

(b) Distribution of BIC-selected neighborhood sizes versus lag P

Figure 7: Distribution of BIC-selected neighborhood sizes

Lag dependence of selected neighborhoods. We focus on June 2019, using 90% of the series for training and 10% for testing. We vary the lag $P \in \{1, 2, 3\}$ and, for each P, apply the entrywise BIC to select the neighborhood size from $K \in \{0, 1, 2, 3, 4, 5\}$. Figure 7b reports the selection frequencies versus P; the bottom panel shows the test RMSE obtained with the BIC-selected neighborhoods.

Figure 7b shows that (i) as P increases, the selected neighborhoods tend to shrink, suggesting that additional temporal lags reduce the need for wider spatial neighborhoods; and (ii) the held-out RMSE does not improve with larger P, which

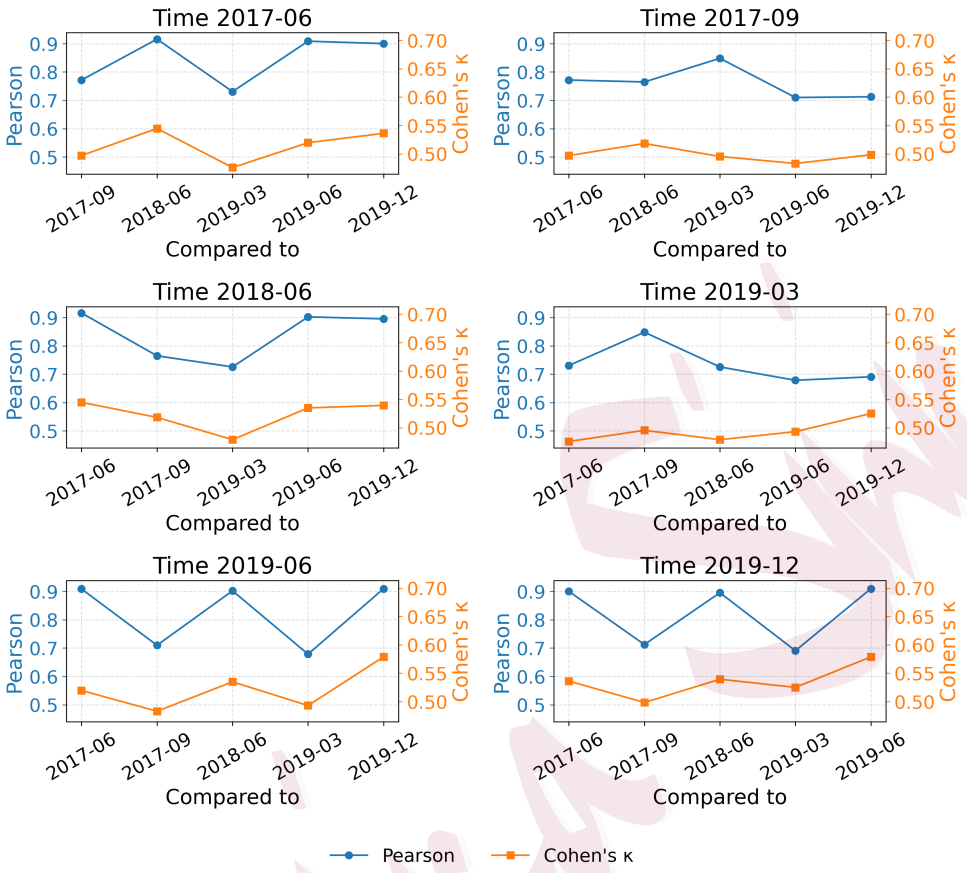


Figure 8: Pairwise similarity across six periods: Pearson correlation of TEC fields and Cohen's κ for the two-category neighborhood maps.

supports the use of $P = 1$ in the reported TEC analysis.

Comparisons with other methods. We compare LIAR and its separable variant SP-LIAR with MAR (Chen et al., 2021), MAR-ST (Hsu et al., 2021), and LIAR-P across the six 10-day periods described above, using the first 90% of each series for

training and the remaining 10% for testing. Test RMSE is computed from one-step-ahead full-field predictions over the test period: at each test time point t, the fitted model uses the observed previous lag to form $\widehat{\boldsymbol{X}}_t$, which is compared with $\boldsymbol{X}_t$; the RMSE is then averaged over all test time points and spatial grid entries. For each period, we fit all methods and report test RMSE in Figure 9; runtimes for LIAR, MAR, and MAR-ST are reported in Table 2.

The pixel-wise baseline LIAR-P is the fastest method, but it also has the largest RMSE. SP-LIAR is closest in structure to MAR-ST; its slightly higher RMSE is expected, since SP-LIAR uses a one-shot SVD projection whereas MAR-ST refines the factors iteratively, a pattern also observed by Chen et al. (2021). The unrestricted LIAR has RMSE close to MAR and slightly below MAR-ST, while taking much less time than both. Thus, in this TEC example, imposing locality keeps most of the predictive information while reducing the computational cost. This agrees with the neighborhood maps above: the selected local neighborhoods reflect meaningful TEC dependence and are useful for prediction, improving one-step-ahead full-field forecasts relative to pixel-wise autoregression.

6. Discussion and Conclusion

We proposed the *local interaction autoregressive* (LIAR) framework, a general principle for modeling high-dimensional matrix and tensor time series via short-range

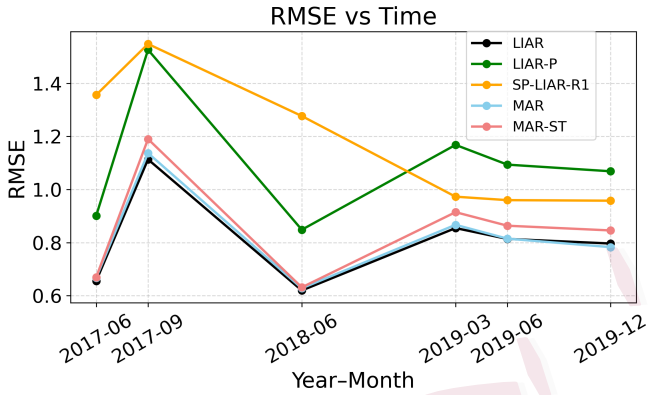


Figure 9: Prediction RMSE over time on TEC data, comparing LIAR, SP-LIAR, MAR, MAR-ST, and LIAR-P

spatio-temporal dependence. In the matrix setting, the proposed SP-LIAR serves as a bridge between traditional MAR/MAR-ST modeling and the LIAR framework. By restricting each entry's evolution to a data-driven neighborhood, LIAR ensures parsimony and interpretability while maintaining predictive accuracy.

A natural extension of the proposed neighborhood selector is to allow the selected neighborhoods to vary smoothly over space. In the current implementation, the BIC-based neighborhood selection is performed separately for each spatial location. This location-wise selection preserves local adaptivity, but it does not explicitly impose spatial coherence on the resulting neighborhood-size map. A simple practical remedy is to post-process the selected neighborhood-size map, for example by applying a local median filter or a local majority rule over a small spatial window. Such a step

Table 2: Runtime (s) by method and period

	2017-06	2017-09	2018-06	2019-03	2019-06	2019-12
LIAR	11.26	8.41	8.90	8.45	4.43	6.84
MAR	467.03	462.89	451.69	438.63	225.81	452.26
MAR-ST	81.30	63.00	88.03	81.54	35.18	69.74

is straightforward to implement and may reduce isolated selections while retaining spatial heterogeneity.

A more integrated approach is to incorporate spatial smoothness directly into the model-selection criterion. Let $\mathcal{E}$ be the set of adjacent spatial pairs. One may consider a spatially regularized BIC criterion

$$\widehat{\boldsymbol{k}} = \arg \min_{\{\boldsymbol{k}=(k_i)_{i\in\mathcal{S}}: k_i \le K_0\}} \left\{ \sum_{i} \mathrm{BIC}_i(k_i) + \lambda \sum_{(i,j)\in\mathcal{E}} |k_i - k_j| \right\},$$

where $\lambda \ge 0$ controls the degree of spatial smoothness. The first term retains the location-wise BIC fit, while the second term discourages abrupt changes of neighborhood sizes between adjacent locations. This formulation can preserve spatially varying local interactions while encouraging smoothly varying neighborhood patterns. Its implementation, however, requires solving a discrete spatial labeling problem and selecting the additional tuning parameter λ. Ordered-label optimization methods or approximate graph-based algorithms may be useful for this purpose; we leave a

detailed computational and theoretical study of this spatially regularized selection procedure to future work.

We developed scalable estimation procedures, including parallel least squares and projection-based methods, a BIC-type neighborhood selector, and statistical theory for kernel and auto-covariance estimation. Extensive simulations show that LIAR achieves lower prediction error and markedly better runtime over competing baselines, including MAR and MAR-ST, across diverse settings. On real-world data, LIAR captures salient local dynamics and provides practical forecasting gains.

7. Acknowledgements

We thank Hu Sun (University of Michigan, Ph.D. 2024) and Professor Han Xiao (Rutgers University) for discussions on the matrix autoregressive model with banded structures. YC acknowledges support from NSF AGS Award 2419187, NASA Federal Award No. 80NSSC23M0192, and No. 80NSSC23M0191.

References

Celani, A., P. Pagnottoni, and G. Jones (2024). Bayesian variable selection for matrix autoregressive models. *Statistics and Computing 34*(2), 91.

Chen, E. Y. and J. Fan (2023). Statistical inference for high-dimensional matrix-variate factor models. *Journal of the American Statistical Association 118*(542), 1038–1055.

Chen, R., H. Xiao, and D. Yang (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics 222*(1), 539–560.

Chen, R., D. Yang, and C.-H. Zhang (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association 117*(537), 94–116.

Davis, R. A., P. Zang, and T. Zheng (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics 25*(4), 1077–1096.

Gao, Z., Y. Ma, H. Wang, and Q. Yao (2019). Banded spatio-temporal autoregressions. *Journal of Econometrics 208*(1), 211–230.

Guo, S., Y. Wang, and Q. Yao (2016). High-dimensional and banded vector autoregressions. *Biometrika 103*(4), 889–903.

Han, Y., D. Yang, C.-H. Zhang, and R. Chen (2024). Cp factor model for dynamic tensors. *Journal of the Royal Statistical Society Series B: Statistical Methodology 86*(5), 1383–1413.

Hsu, N.-J., H.-C. Huang, and R. S. Tsay (2021). Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics 30*(4), 1143–1155.

Hsu, N.-J., H.-C. Huang, R. S. Tsay, and T.-C. Kao (2024). Rank-r matrix autoregressive models for modeling spatio-temporal data. *Statistics and Its Interface 17*(2), 275–290.

Jiang, H., B. Shen, Y. Li, and Z. Gao (2024). Regularized estimation of high-dimensional matrix-variate autoregressive models. *Statistica Sinica*.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 694–726.

Li, Z. and H. Xiao (2021). Multi-linear tensor autoregressive models. *arXiv preprint*

arXiv:2110.00928.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Sun, H., Y. Chen, S. Zou, J. Ren, Y. Chang, Z. Wang, and A. Coster (2023). Complete global total electron content map dataset based on a video imputation algorithm vista. *Scientific Data 10*(1), 236.

Sun, H., Z. Hua, J. Ren, S. Zou, Y. Sun, and Y. Chen (2022). Matrix completion methods for the total electron content video reconstruction. *The Annals of Applied Statistics 16*(3), 1333–1358.

Sun, H., Z. Shang, and Y. Chen (2025). Matrix autoregressive model with vector time series covariates for spatio-temporal data. *Statistica Sinica*.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.

Tsay, R. S. (2023). Matrix-variate time series analysis: A brief review and some new developments. *International Statistical Review*.

Wang, D., X. Liu, and R. Chen (2019). Factor models for matrix-valued high-dimensional time series. *Journal of econometrics 208*(1), 231–248.

Jingyang Li, Fudan University

E-mail: jjyyli@fudan.edu.cn

Yang Chen, University of Michigan, Ann Arbor

E-mail: ychenang@umich.edu