

# Interpretable Time Series Autoregression for Periodicity Quantification

Xinyu Chen, Vassilis Digalakis Jr, Lijun Ding, Dingyi Zhuang, and Jinhua Zhao

**Abstract**—Time series autoregression is a classical statistical model for capturing auto-correlations and identifying temporal patterns such as periodicity and seasonality. In this work, we propose a novel sparse autoregression framework from an interpretable machine learning perspective and the model interpretability for periodicity quantification is reinforced by  $\ell_0$ -norm induced sparsity constraints. On the time-varying time series data, we reformulate the sparse autoregression and convert the involved optimization problem into a mixed-integer optimization (MIO). To accelerate it, we develop a subspace pursuit based decision variable pruning (DVP) strategy to reduce the search space. On the multidimensional time series that involves complicated spatial and temporal dimensions, we propose a spatially- and time-varying sparse autoregression model and resolve the corresponding MIO problem by developing a two-stage optimization scheme. In particular, the proposed scheme makes the model scalable to large problems even with millions of decision variables. Empirically, we conduct extensive experiments to evaluate the proposed models on real-world time series data. First, we demonstrate that the MIO solver can be drastically accelerated through the DVP strategy, while maintaining the same solution quality as a full MIO solver. Applying the time-varying sparse autoregression model to ridesharing trip data, we uncover both daily and weekly periodicities and reveal long-term changes in regularity of human mobility. Second, we demonstrate the spatial patterns of yearly seasonality in climate variable time series such as temperature and precipitation across the past four decades, and our model allows to discover dynamic climate patterns and identify climate phenomena such as El Niño in sea surface temperature. Thus, our framework lays an insightful foundation for understanding periodicity and seasonality in real-world time series.

**Index Terms**—Interpretable machine learning, time series analysis, sparse autoregression, periodicity quantification, mixed-integer optimization, urban transportation systems, human mobility, climate systems

## 1 INTRODUCTION

MANY real-world systems exhibit complex temporal patterns, including periodicity, seasonality, and anomalies. Detecting and understanding these patterns is essential for anticipating system behavior, identifying disruptions, and supporting operational decision-making. In dynamic environments such as urban transportation and climate systems, periodicities can shift due to external factors—policy interventions, demand changes, environmental variability, global events, or extreme climate phenomena—making interpretable, adaptive, and data-driven methods indispensable. A central challenge is to automatically identify dominant periodic components from spatially- and time-varying systems, track their evolution over a long-term time period, and distinguish true structural changes from random variability.

**Urban transportation systems** display strong periodicity driven by commuting patterns, business cycles, and travel demand. Fig. 1 shows the regularity of ridesharing activity in Chicago during the first two weeks (336 hours) starting April 1, 2024. The data exhibits strong periodicity with a weekly cycle  $\Delta t = 7 \times 24 = 168$ .

infrastructure changes, economic conditions, and disruptive events. For example, the COVID-19 pandemic in 2020 led to a collapse in established mobility periodicities due to lockdowns and shifts to remote work. These disruptions raise key questions: *How do periodic structures evolve over time? Can we systematically quantify such changes in a time-varying system?* Addressing these questions is critical for planning, forecasting, and adaptive resource allocation.

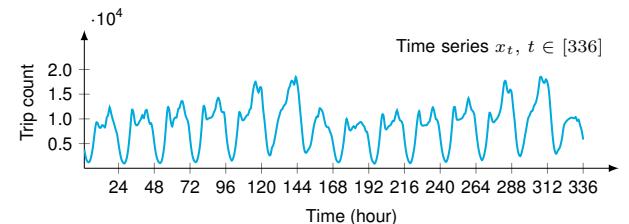


Fig. 1: Hourly time series of aggregated ridesharing trip counts in Chicago during the first two weeks (336 hours) starting April 1, 2024. The data exhibits strong periodicity with a weekly cycle  $\Delta t = 7 \times 24 = 168$ .

- Xinyu Chen, Dingyi Zhuang, and Jinhua Zhao are with the Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: chenxy346@gmail.com; dingyi@mit.edu; jinhua@mit.edu).
- Vassilis Digalakis Jr is with the Questrom School of Business, Boston University, Boston, MA 02115, USA (e-mail: vvdigalakis@gmail.com).
- Lijun Ding is with the Department of Mathematics at the University of California, San Diego, La Jolla, CA 92093, USA (e-mail: l2ding@ucsd.edu).

(Corresponding author: Jinhua Zhao)

**Climate systems** also exhibit periodic and seasonal patterns that shape temperature, ocean circulation, and atmospheric dynamics. Yet these patterns evolve over time due to long-term variability and climate change. Traditional time series decomposition methods [1], [2], [3] assume fixed seasonal structure and often ignore gradual or region-specific shifts. Accurately monitoring such changes is essential for understanding climate dynamics, anticipating ex-

treme events, and supporting policy-making. This calls for *interpretable, data-driven models that can robustly uncover the dominant seasonal components and track their evolution across space and time.*

To address these needs, we develop a unified interpretable machine learning framework for identifying and quantifying time-varying periodicity and structural shifts in real-world time series. Our models build on classical autoregression (AR) [4], [5] but incorporate sparse structure via  $\ell_0$ -norm induced sparsity constraints to promote interpretability. Inspired by recent advances in sparse regression [6] and time series convolution [7], we formulate interpretable AR models capable of isolating dominant periodic patterns over time and space. To the best of our knowledge, this represents the first application of exact sparse AR to the task of periodicity quantification in real-world complex time series. Overall, this work makes the following contributions:

- **Sparse Autoregression (SAR):** We introduce an interpretable framework for identifying dominant auto-correlations from time series by reformulating AR with  $\ell_0$ -norm induced sparsity constraints. The problem is solved exactly via mixed-integer optimization (MIO) techniques, providing more accurate and reliable periodicity quantification than conventional greedy methods.
- **Time-Varying SAR (TV-SAR):** We extend SAR to characterize non-stationary time series, enforcing consistent support sets across time segments to enhance interpretability. To improve scalability, we introduce a decision variable pruning (DVP) strategy that narrows the MIO search space using fast greedy approximations such as subspace pursuit.
- **Spatially- and Time-Varying SAR (STV-SAR):** We propose a scalable model for multidimensional time series that vary over both space and time. A two-stage optimization procedure—global support set selection via MIO, followed by local coefficient estimation via quadratic optimization—makes the method tractable for millions of decision variables.
- **Extensive Real-World Validation:** We demonstrate the effectiveness of our models on large-scale transportation and climate datasets. TV-SAR reveals dynamic changes in daily and weekly mobility patterns related to periodicity in New York City (NYC) during the period of the COVID-19 pandemic. STV-SAR captures evolving spatial patterns of seasonality in North American climate variables and identifies global sea surface temperature dynamics related to El Niño.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces notation and background on AR models. Section 4 presents the core SAR model. Sections 5 and 6 develop the TV-SAR and STV-SAR extensions, respectively. Section 7 reports empirical results. Finally, we conclude this study in Section 8.

## 2 LITERATURE REVIEW

### 2.1 Classical and Time-Varying Autoregression

Time series modeling is a foundational tool in transportation, climate science, econometrics, and other fields [8], [9],

[10]. Classical models such as autoregression (AR), moving average (MA), and their combinations (e.g., ARIMA) are widely used for capturing temporal dependencies and seasonal structures in univariate time series data [4], [5], [11]. Among these, AR models remain popular due to their simplicity and interpretability. More recently, applications have demanded models that adapt to non-stationary and time-varying systems. Time-varying AR extends classical AR by allowing coefficients to change over time [12], often incorporating structural or smoothness constraints to ensure interpretability and stability [13]. AR has also been generalized to multivariate and spatial-temporal settings, such as Vector Autoregressive (VAR) models [14], [15]. These ideas have been successfully applied in finance, neuroscience, and dynamic mobility systems.

### 2.2 Sparse Autoregression and Interpretability

Traditional AR models include all lagged terms up to a fixed order, which can lead to overfitting and obscure the most meaningful temporal dependencies—particularly in high-order settings [16]. Sparse AR addresses this limitation by selecting a subset of informative lags, thereby improving both parsimony and interpretability. LASSO-based methods [17] have been widely used to induce sparsity in AR models, offering scalability and robustness in high-dimensional contexts [18]. Recent work has further explored structured sparsity in time series models. For instance, [19], [20], [21] develop interpretable sparse formulations for nonlinear and dynamical systems, while [6] introduces structured sparsity constraints over graphs to model slowly evolving regression coefficients. A related line of research by [22] proposes a sparsity-controlled VAR framework, allowing users to tune multiple dimensions of sparsity for enhanced interpretability of causal discovery. While this approach offers flexibility and shows improved predictive accuracy over LASSO-based alternatives, it lacks scalability and does not provide exact solutions with guaranteed optimality.

These advances move beyond classical LASSO penalties by incorporating domain-specific sparsity structures. However, existing methods typically do not enforce structured sparsity across time and space simultaneously, nor do they leverage exact combinatorial optimization. Our work bridges this gap by introducing a framework for structured sparse AR with exact support set control over spatiotemporal dimensions.

### 2.3 Exact Sparse Regression via Mixed-Integer Optimization

Sparse regression via  $\ell_0$ -norm regularization—also known as best subset selection—has long been recognized for its statistical optimality and interpretability, but was historically limited by its combinatorial complexity. Early heuristic methods include Orthogonal Matching Pursuit and CoSaMP [23], [24], [25], while convex relaxations such as the LASSO [17] and non-convex penalties such as SCAD and MCP offered tractable alternatives [26]. The solution quality of these methods often negatively impact the interpretability of sparse structures. Recent breakthroughs in optimization have made it practically feasible to solve  $\ell_0$ -norm regularized problems exactly using MIO. Following the seminal

work of [27], several studies have scaled MIO-based sparse regression to large datasets [28], [29], [30]. These methods retain the interpretability of exact sparse models while achieving near-LASSO speed.

Our work builds directly on this line, leveraging the MIO machinery developed in [6] to solve structured sparse regression with controlled support set consistency. However, unlike prior work which focused on static or graph-based regression, we apply these methods to time series with dynamic and spatial variation, developing the first exact sparse AR framework for periodicity quantification in large-scale real-world systems.

### 3 PRELIMINARIES

#### 3.1 Notation

In Table 1, we summarize the basic symbols and notation used throughout the paper. Notably,  $\mathbb{R}$  denotes the set of real numbers, and  $\mathbb{Z}^+$  denotes the set of positive integers.

TABLE 1: Summary of basic notation.

Notation	Description
$x \in \mathbb{R}$	Scalar
$\mathbf{x} \in \mathbb{R}^n$	Vector of length $n$
$\mathbf{X} \in \mathbb{R}^{m \times n}$	Matrix of size $m \times n$
$[i]$	Integer set $\{1, 2, \dots, i\}, i \in \mathbb{Z}^+$
$[i, j]$	Integer set $\{i, i+1, \dots, j\}, i < j$
$\ \cdot\ _0$	$\ell_0$ -norm (number of nonzero entries)
$\ \cdot\ _1$	$\ell_1$ -norm (sum of absolute values)
$\ \cdot\ _2$	$\ell_2$ -norm (Euclidean norm)
$\text{tr}(\cdot)$	Trace of a square matrix
$\text{supp}(\cdot)$	Support set (indices of nonzero entries)
$\mathcal{N}(\cdot)$	Gaussian distribution
$\mathbb{E}[\cdot]$	Expectation
$\cup$	Union of sets
$\cap$	Intersection of sets

#### 3.2 Time Series Autoregression

AR is a widely used technique for modeling temporal dependencies in univariate time series [4], [5]. It expresses each observation as a linear combination of its past values, plus noise. For a univariate time series  $\mathbf{x} = (x_1, x_2, \dots, x_T)^\top \in \mathbb{R}^T$ , the order- $d$  AR model is written as:

$$x_t = \sum_{k=1}^d w_k x_{t-k} + \epsilon_t, \quad \forall t \in [d+1, T], \quad (1)$$

where  $d \in \mathbb{Z}^+$  is the AR order, and  $\mathbf{w} = (w_1, w_2, \dots, w_d)^\top \in \mathbb{R}^d$  is the coefficient vector. As the residual,  $\epsilon_t$  denotes noise, typically modeled by a Gaussian assumption:  $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$  for the variance  $\sigma^2 > 0$ . The coefficient  $w_k$  captures the linear dependence between  $x_t$  and its  $k$ -lagged value  $x_{t-k}$ .

To estimate the coefficients, we minimize the sum of squared residuals such that

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{t=d+1}^T \left( x_t - \sum_{k=1}^d w_k x_{t-k} \right)^2. \quad (2)$$

By defining the  $(T-d)$ -by- $d$  design matrix and length- $(T-d)$  target vector as:

$$\mathbf{A} = \begin{bmatrix} x_d & x_{d-1} & \cdots & x_1 \\ x_{d+1} & x_d & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-1} & x_{T-2} & \cdots & x_{T-d} \end{bmatrix}, \quad \tilde{\mathbf{x}} = \begin{bmatrix} x_{d+1} \\ x_{d+2} \\ \vdots \\ x_T \end{bmatrix}, \quad (3)$$

respectively. Then, Problem (2) becomes

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2.$$

This leads to the standard least squares solution:  $\hat{\mathbf{w}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \tilde{\mathbf{x}}$ , assuming  $\mathbf{A}^\top \mathbf{A}$  is invertible. In more general settings, this solution can also be expressed using the Moore-Penrose pseudoinverse as  $\hat{\mathbf{w}} = \mathbf{A}^\dagger \tilde{\mathbf{x}}$ .

A time series  $\mathbf{x}$  is said to exhibit periodicity with period  $\Delta t \in \mathbb{Z}^+$  if  $x_t \approx x_{t+\Delta t}$  for many  $t$ , typically reflected by high auto-correlation at lag  $\Delta t$ , i.e.,  $\text{Cov}(x_t, x_{t+\Delta t}) \gg 0$ . Seasonality refers to periodic patterns tied to calendar cycles (e.g., daily, weekly, and yearly) and can be modeled as a component  $s_t$  satisfying  $s_t = s_{t+\Delta t}$ . In the AR framework, strong periodicity at lag  $\Delta t$  manifests as a large positive coefficient  $w_{\Delta t}$ , allowing us to infer dominant cycles directly from the estimated sparse coefficient vector  $\mathbf{w}$ .

Fig. 2 illustrates the coefficient vector  $\mathbf{w}$  estimated for the Chicago ridesharing time series using standard least squares. The resulting vector is dense, with both positive and negative coefficients. A large positive  $w_k$  at lag  $k = 168$  reflects the strong weekly periodicity of the time series. However, this dense representation makes it difficult to identify which lags are most important and to quantify periodic structure precisely. This motivates the need for sparse and interpretable AR models.

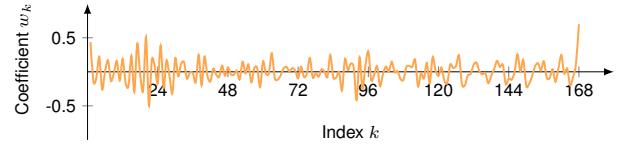


Fig. 2: Estimated AR coefficients  $\mathbf{w} \in \mathbb{R}^d$  for the Chicago ridesharing time series in Fig. 1, with order  $d = 168$ .

### 4 SPARSE AUTOREGRESSION

This section introduces Sparse Autoregression (SAR) with  $\ell_0$ -norm induced sparsity constraints for identifying dominant auto-correlations in time series. We first describe the modeling framework and associated statistical guarantees. We then present an MIO formulation for solving the resulting optimization problem exactly and compare solution quality across algorithms.

#### 4.1 Model Description

The  $\ell_0$ -norm of a vector  $\mathbf{w} \in \mathbb{R}^d$  is defined as  $\|\mathbf{w}\|_0 = |\text{supp}(\mathbf{w})| \leq d$ , i.e., the number of nonzero entries. While the least squares estimator in Eq. (2) produces dense solutions, it does not highlight dominant auto-correlations, making it difficult to quantify periodicity or seasonality.

To address this, we impose sparsity and non-negativity constraints on the coefficient vector  $\mathbf{w}$ , yielding the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & 0 \leq \mathbf{w} \leq \mathbf{M}, \quad \|\mathbf{w}\|_0 \leq \tau, \end{aligned} \quad (4)$$

where  $\mathbf{M} \in \mathbb{R}^d$  is a vector with all entries equal to a sufficiently large constant  $M \in \mathbb{R}$ , and  $\tau \in \mathbb{Z}^+$  controls the maximum number of nonzero coefficients. The non-negativity constraint encourages interpretability by focusing on positive auto-correlations, as is typical for periodicity in time series.

Next, we study the statistical learnability of a more general version of Problem (4)—excluding the non-negativity constraint. The following result, which we prove in Appendix A, relies on standard arguments [31].

**Theorem 1** (Excess Risk Bound for SAR). *Let  $\tilde{\mathbf{x}} = \mathbf{A}\mathbf{w}^* + \epsilon$ , where:*

- $\mathbf{A} \in \mathbb{R}^{(T-d) \times d}$  is fixed and deterministic,
- $\mathbf{w}^* \in \mathbb{R}^d$  satisfies  $\|\mathbf{w}^*\|_0 \leq \tau < d/2$ ,
- $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_d)$  is i.i.d. Gaussian noise.

Let  $\hat{\mathbf{w}}$  be the solution to Problem (4). Then the expected in-sample prediction error satisfies:

$$\mathbb{E} \left[ \frac{\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2}{T-d} \right] \lesssim \sigma^2 \cdot \frac{\tau}{T-d} \left( \log \left( \frac{d}{\tau} \right) + 1 \right).$$

Theorem 1 demonstrates that the error scales gracefully with sparsity level  $\tau$  and AR order  $d$ .

## 4.2 MIO Reformulation

Due to the combinatorial nature of the  $\ell_0$ -norm induced constraint, Problem (4) cannot be directly solved by MIO solvers. However, we can equivalently express the problem as a mixed-integer quadratic optimization problem by introducing binary variables that encode the support of  $\mathbf{w}$ . Let  $\mathbf{z} \in \{0, 1\}^d$  be a vector of binary variables, where  $z_k = 1$  indicates that  $w_k$  is allowed to be nonzero. We rewrite Problem (4) as:

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{z}} \quad & \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & 0 \leq w_k \leq M \cdot z_k, \quad \forall k \in [d], \\ & \sum_{k=1}^d z_k \leq \tau, \\ & z_k \in \{0, 1\}, \quad \forall k \in [d]. \end{aligned} \quad (5)$$

The binary support constraint  $\sum_{k=1}^d z_k \leq \tau$  ensures that at most  $\tau$  coefficients in  $\mathbf{w}$  are nonzero.

Problem (5) is nonconvex due to the  $\ell_0$ -norm induced constraint and is generally NP-hard to solve exactly. In principle, enumerating all support sets of cardinality  $\tau$  would incur a combinatorial cost of  $\mathcal{O}(d^\tau)$ , making exact search infeasible for large  $d$  or  $\tau$ . Modern MIO solvers bypass this challenge using branch-and-bound techniques and cutting-plane methods (e.g., [6]), which allow them to find globally optimal solutions efficiently in practice. Compared to greedy methods such as subspace pursuit, MIO yields

higher-quality solutions with provable optimality guarantees. The use of binary variables enables precise support set control, which is essential for model interpretability in settings where dominant lag selection matters.

## 4.3 Empirical Comparison of Solution Quality

The estimation method has a significant impact on the interpretability and fidelity of SAR models. To highlight this, we compare two approaches for solving Problem (4): (i) a greedy non-negative subspace pursuit (NNSP) algorithm [7], and (ii) the exact MIO formulation from Eq. (5). We use the time series from Fig. 1, with AR order  $d = 168$  (one week of hourly lags) and sparsity level  $\tau = 2$ . The solution returned by NNSP is:

$$\mathbf{w} = (0, \dots, 0, \underbrace{0.02}_{k=53}, 0, \dots, 0, \underbrace{0.96}_{k=168})^\top,$$

with objective value  $f(\mathbf{w}) = 8.32 \times 10^7$ . In contrast, the MIO solver yields:

$$\mathbf{w} = (\underbrace{0.22}_{k=1}, 0, \dots, 0, \underbrace{0.77}_{k=168})^\top,$$

with objective value  $f(\mathbf{w}) = 6.25 \times 10^7$ . The MIO solution is both quantitatively superior (lower error) and qualitatively more interpretable: it identifies lag  $k = 1$  (local autocorrelation) and lag  $k = 168$  (weekly periodicity), consistent with domain knowledge. The NNSP solution, by contrast, includes a spurious lag at  $k = 53$  with negligible weight.

When increasing the sparsity level to  $\tau = 3$ , both solvers return:

$$\mathbf{w} = (\underbrace{0.33}_{k=1}, 0, \dots, 0, \underbrace{0.20}_{k=167}, \underbrace{0.46}_{k=168})^\top,$$

highlighting strong lags near daily and weekly cycles.

Fig. 3 summarizes how the selected support sets and coefficient magnitudes evolve as a function of  $\tau$ . These results confirm that high-quality solutions are critical for interpretability and that MIO offers robust, principled support recovery even when greedy methods fail.

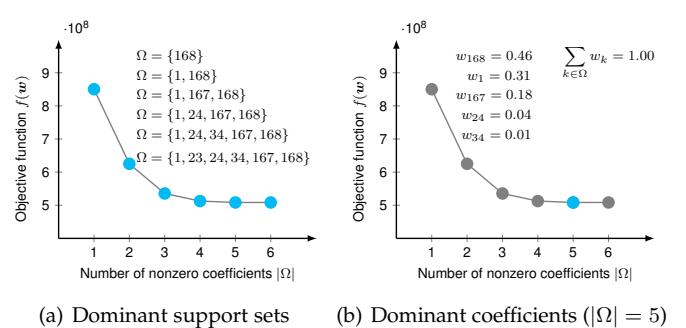


Fig. 3: Illustration of the dominant coefficients of SAR on the Chicago ridesharing trip time series (see Fig. 1) at different sparsity levels. The support set and the number of nonzero coefficients are denoted by  $\Omega$  and  $|\Omega|$ , respectively.

## 5 TIME-VARYING SPARSE AUTOREGRESSION

In this section, we now extend the SAR formulation to capture non-stationary dynamics by allowing the coefficients to vary across prescribed time segments. This yields the Time-Varying Sparse Autoregression (TV-SAR) model. To solve the associated MIO problem more efficiently, we also introduce a decomposition-based DVP strategy.

### 5.1 Model Description

Time-varying AR models allow the AR coefficients to evolve across time, capturing structural shifts in the data-generating process. We partition the time series into  $\Gamma \in \mathbb{Z}^+$  time segments, where each segment  $\gamma \in [\Gamma]$  contains  $T_\gamma \in \mathbb{Z}^+$  time steps. Let  $\mathbf{x}_\gamma = (x_{\gamma,1}, x_{\gamma,2}, \dots, x_{\gamma,T_\gamma})^\top \in \mathbb{R}^{T_\gamma}$  denote the segment corresponding to time segment  $\gamma$ . We model each segment with an independent  $d$ th-order AR process, with potentially different coefficient vectors  $\mathbf{w}_\gamma \in \mathbb{R}^d$ :

$$x_{\gamma,t} = \sum_{k=1}^d w_{\gamma,k} x_{\gamma,t-k} + \epsilon_{\gamma,t}, \quad \forall t \in [d+1, T_\gamma], \gamma \in [\Gamma],$$

where  $\epsilon_{\gamma,t} \sim \mathcal{N}(0, \sigma^2)$ . To ensure interpretability and temporal consistency, we impose the following structural constraints: (i) each  $\mathbf{w}_\gamma$  is sparse and non-negative, and (ii) all  $\mathbf{w}_\gamma$  share the same support set. Formally, we let  $\Phi = \{\mathbf{w}_\gamma\}_{\gamma=1}^\Gamma$  represent the set of coefficient vectors, referring to the decision variables in the optimization problem of TV-SAR such that

$$\begin{aligned} \min_{\Phi} \quad & \sum_{\gamma=1}^{\Gamma} \|\tilde{\mathbf{x}}_\gamma - \mathbf{A}_\gamma \mathbf{w}_\gamma\|_2^2 \\ \text{s.t.} \quad & 0 \leq \mathbf{w}_\gamma \leq \mathbf{M}, \quad \|\mathbf{w}_\gamma\|_0 \leq \tau, \quad \forall \gamma \in [\Gamma], \\ & \text{supp}(\mathbf{w}_\gamma) = \text{supp}(\mathbf{w}_{\gamma+1}), \quad \forall \gamma \in [\Gamma-1], \end{aligned} \quad (6)$$

where  $\mathbf{A}_\gamma \in \mathbb{R}^{(T_\gamma-d) \times d}$  is the design matrix,  $\tilde{\mathbf{x}}_\gamma \in \mathbb{R}^{T_\gamma-d}$  is the response vector for time segment  $\gamma$ ,  $\mathbf{M} \in \mathbb{R}^{d \times d}$  consists of all sufficiently large constant  $\mathcal{M} > 0$ , and  $\tau \in \mathbb{Z}^+$  controls global sparsity.

**Remark 1.** The constraint of the same support set enforces temporal smoothness by requiring identical sparsity patterns across time segments. This is a special case of the sparsely-varying support set constraint proposed in [6], where the symmetric difference between support sets is bounded:

$$|\text{supp}(\mathbf{w}_\gamma) \cup \text{supp}(\mathbf{w}_{\gamma+1})| - |\text{supp}(\mathbf{w}_\gamma) \cap \text{supp}(\mathbf{w}_{\gamma+1})| \leq \tilde{\tau}.$$

Setting  $\tilde{\tau} = 0$  recovers the shared support set case.

To encode Eq. (6) as an MIO problem, we again introduce binary variables  $\mathbf{z} \in \{0, 1\}^d$  to represent the global support set. The resulting MIO formulation is given by

$$\begin{aligned} \min_{\Phi, \mathbf{z}} \quad & \sum_{\gamma=1}^{\Gamma} \|\tilde{\mathbf{x}}_\gamma - \mathbf{A}_\gamma \mathbf{w}_\gamma\|_2^2 \\ \text{s.t.} \quad & 0 \leq \mathbf{w}_\gamma \leq \mathcal{M} \cdot \mathbf{z}, \quad \forall \gamma \in [\Gamma], \\ & \sum_{k=1}^d z_k \leq \tau, \\ & z_k \in \{0, 1\}, \quad \forall k \in [d]. \end{aligned} \quad (7)$$

Here, the binary variable  $z_k = 1$  if lag  $k$  is selected for any  $\gamma$ , enforcing a common support set across all time segments. TV-SAR thus extends SAR to non-stationary settings, allowing the autoregressive weights to vary across time segments while preserving interpretability through global sparsity and support set consistency.

### 5.2 Acceleration with Decision Variable Pruning

The computational cost of solving TV-SAR with MIO increases with the AR order  $d$ , due to the total number of decision variables  $(\Gamma + 1)d$ , which includes  $\Gamma d$  real-valued variables for the AR coefficients  $\mathbf{w}_\gamma$  and  $d$  binary variables in  $\mathbf{z}$ . When the sparsity level  $\tau$  is much smaller than  $d$ , most of these variables are expected to be zero. We exploit this by introducing a *Decision Variable Pruning (DVP)* strategy, which leverages subspace pursuit to pre-select a reduced candidate set of variables before solving the MIO. The strategy proceeds in three steps (illustrated in Fig. 4):

- **Step 1: Relax and Decompose TV-SAR.** We relax the TV-SAR formulation in Eq. (6), keeping only non-negativity and a looser sparsity constraint  $\tau_0 > \tau$ . This yields  $\Gamma$  decomposable subproblems:

$$\begin{aligned} \min_{\mathbf{w}_\gamma} \quad & \|\tilde{\mathbf{x}}_\gamma - \mathbf{A}_\gamma \mathbf{w}_\gamma\|_2^2 \\ \text{s.t.} \quad & \mathbf{w}_\gamma \geq 0, \quad \|\mathbf{w}_\gamma\|_0 \leq \tau_0. \end{aligned} \quad (8)$$

- **Step 2: Implement Subspace Pursuit.** We first solve each subproblem using NNSP. Let  $S_\gamma$  denote the support set of the resulting solution for time segment  $\gamma \in [\Gamma]$ , see Algorithm 1. We then form the global candidate set

$$\tilde{S} = \bigcup_{\gamma=1}^{\Gamma} S_\gamma.$$

- **Step 3: Solve the Reduced MIO.** We solve the original MIO problem in Eq. (7), restricting all decision variables to the reduced index set  $\tilde{S}$ . This results in a significantly smaller problem with  $(\Gamma + 1) \cdot |\tilde{S}|$  variables.

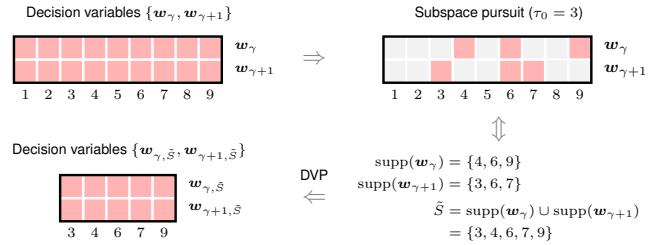


Fig. 4: Illustration of the DVP strategy. Subspace pursuit is used to select an index set  $\tilde{S}$  of candidate lags. The final MIO is then solved on this reduced support set, where the number of coefficients is reduced from  $2d$  to  $2|\tilde{S}|$  (e.g., from 18 coefficients to 10 coefficients in the illustration).

We integrate the DVP strategy into an MIO solver and refer to the resulting hybrid method as MIO-DVP, parameterized by  $\tau_0$ . The approach significantly reduces the search space—especially for large  $d$  (e.g.,  $d = 168$  for weekly

---

**Algorithm 1** Decision Variable Pruning via Non-Negative Subspace Pursuit

---

- 1: **Input:** Time series  $\mathbf{x}_\gamma \in \mathbb{R}^T$ ,  $\gamma \in [\Gamma]$ ; AR order  $d$ ; relaxed sparsity  $\tau_0 > \tau$ .
- 2: **for**  $\gamma \in [\Gamma]$  **do**
- 3:   Construct  $\tilde{\mathbf{x}}_\gamma \in \mathbb{R}^{T-d}$  and design matrix  $\mathbf{A}_\gamma \in \mathbb{R}^{(T-d) \times d}$ .
- 4:   Initialize  $\mathbf{w}_\gamma := \mathbf{0}$ ,  $S_\gamma := \emptyset$ , and residual  $\mathbf{r} := \tilde{\mathbf{x}}_\gamma$ .
- 5:   **while** not converged **do**
- 6:     Identify  $\ell$ : index set of  $\tau$  largest entries in  $|\mathbf{A}_\gamma^\top \mathbf{r}|$ .
- 7:      $S_\gamma \leftarrow S_\gamma \cup \ell$
- 8:     Solve  $\mathbf{w}_{\gamma, S_\gamma} := \arg \min_{\mathbf{v} \geq 0} \|\tilde{\mathbf{x}}_\gamma - \mathbf{A}_{\gamma, S_\gamma} \mathbf{v}\|_2^2$ .
- 9:     Keep  $\tau$  largest entries of  $\mathbf{w}_\gamma$ , zero out the rest.
- 10:   Update  $S_\gamma$  and  $\mathbf{w}_{\gamma, S_\gamma}$ .
- 11:   Update residual:  $\mathbf{r} \leftarrow \tilde{\mathbf{x}}_\gamma - \mathbf{A}_{\gamma, S_\gamma} \mathbf{w}_{\gamma, S_\gamma}$ .
- 12:   **end while**
- 13: **end for**
- 14: Return  $\tilde{S} := \bigcup_{\gamma=1}^{\Gamma} S_\gamma$ .

---

periodicity)—and makes MIO tractable at scale. MIO-DVP can be viewed as a *backbone-type algorithm* [32]: an iterative two-stage method that first screens candidate features and then solves the final problem over this reduced support set. Such screening strategies have been developed from both statistical (e.g., sure screening [33]) and optimization (e.g., safe screening [34]) perspectives.

## 6 SPATIALLY AND TIME-VARYING SPARSE AUTOREGRESSION

In this section, we introduce the Spatially and Time-Varying Sparse Autoregression (STV-SAR) model, a generalization of TV-SAR that accounts for both temporal and spatial variations in AR behavior. While TV-SAR captures time-varying dynamics within a single univariate time series, STV-SAR is designed for settings involving large spatiotemporal panels of time series (e.g., satellite-based climate data), where thousands of spatial locations exhibit their own local dynamics. This additional spatial dimension leads to a more expressive—but also more complex—model. In particular, when the number of spatial locations (expressed in terms of their latitudes and longitudes)  $M \times N = 1$ , STV-SAR reduces to TV-SAR. Conversely, STV-SAR enables us to model heterogeneity across locations while leveraging global sparsity and seasonality structures. To solve the associated MIO problem efficiently, we describe a two-stage optimization scheme that separates the learning processes of the global support set and individual coefficient vectors.

### 6.1 Model Description

Considering a collection of time series arranged over an  $M \times N$  spatial grid, we let  $\mathbf{X}_{\gamma, t} \in \mathbb{R}^{M \times N}$  denote the spatial matrix at time  $t \in [T_\gamma]$  in time segment  $\gamma \in [\Gamma]$ . Each grid cell  $(m, n)$  contains a multivariate time series  $\{x_{m, n, \gamma, t}\}_{t \in [T_\gamma], \gamma \in [\Gamma]}$ . Following the same logic as in Section 5, we model each time series with a  $d$ -order AR process, with coefficient vectors  $\mathbf{w}_{m, n, \gamma} \in \mathbb{R}^d$ , and impose a shared global

support set across all spatial locations and time segments. Let  $\tilde{\mathbf{x}}_{m, n, \gamma}$  and  $\mathbf{A}_{m, n, \gamma}$  be the lagged response vector and design matrix (as in Eq. (3)), respectively, then the MIO problem is formulated as follows,

$$\begin{aligned} \min_{\Phi, \mathbf{z}} \quad & \sum_{m, n, \gamma} \|\tilde{\mathbf{x}}_{m, n, \gamma} - \mathbf{A}_{m, n, \gamma} \mathbf{w}_{m, n, \gamma}\|_2^2 \\ \text{s.t.} \quad & 0 \leq \mathbf{w}_{m, n, \gamma} \leq \mathcal{M} \cdot \mathbf{z}, \quad \forall m, n, \gamma, \\ & \sum_{k=1}^d z_k \leq \tau, \\ & z_k \in \{0, 1\}, \quad \forall k \in [d], \end{aligned} \quad (9)$$

where  $\Phi = \{\mathbf{w}_{m, n, \gamma}\}_{m \in [M], n \in [N], \gamma \in [\Gamma]}$  denotes the set of coefficient vectors. The binary vector  $\mathbf{z} \in \{0, 1\}^d$  encodes the shared support set,  $\mathcal{M} > 0$  is a sufficiently large upper bound on the coefficients, and  $\tau \in \mathbb{Z}^+$  is a global sparsity budget.

This formulation assumes that only a few autoregressive lags drive the spatiotemporal dynamics across the entire system. For instance, in climate systems, monthly time series such as temperature or precipitation often exhibit strong seasonal structure. By setting  $d = 12$ , the model can select from lags corresponding to past months in the year. If, for example,  $z_{12} = 1$  in the optimal solution, it indicates that the data exhibits strong yearly periodicity—i.e., each month's value is relevant to its value one year ago. Other selected lags (e.g.,  $z_1, z_3$ , etc.) can be interpreted as short-term auto-correlations or sub-seasonal effects.

### 6.2 Acceleration via Global Support Estimation

When dealing with large spatiotemporal systems—such as climate datasets covering thousands of grid cells across multiple decades—estimating a separate SAR model for each individual time series becomes computationally infeasible. However, these time series often share common underlying periodicity and auto-correlation patterns. To exploit this structure, we first estimate a global SAR structure by fitting a single sparse coefficient vector  $\mathbf{w} \in \mathbb{R}^d$  across  $M \times N \times \Gamma$  time series. This global sparsity pattern can be reused to simplify subsequent localized coefficient estimation and identify a shared support set that generalizes across space and time. According to the property of matrix trace, we rewrite the objective function as follows,

$$\begin{aligned} f(\mathbf{w}) &= \sum_{m, n, \gamma} \|\tilde{\mathbf{x}}_{m, n, \gamma} - \mathbf{A}_{m, n, \gamma} \mathbf{w}\|_2^2 \\ &= \sum_{m, n, \gamma} (\tilde{\mathbf{x}}_{m, n, \gamma} - \mathbf{A}_{m, n, \gamma} \mathbf{w})^\top (\tilde{\mathbf{x}}_{m, n, \gamma} - \mathbf{A}_{m, n, \gamma} \mathbf{w}) \\ &= \sum_{m, n, \gamma} \left( \text{tr}(\mathbf{w} \mathbf{w}^\top \mathbf{A}_{m, n, \gamma}^\top \mathbf{A}_{m, n, \gamma}) \right. \\ &\quad \left. - 2\mathbf{w}^\top \mathbf{A}_{m, n, \gamma}^\top \tilde{\mathbf{x}}_{m, n, \gamma} \right) + C \\ &= \text{tr} \left( \mathbf{w} \mathbf{w}^\top \sum_{m, n, \gamma} \mathbf{A}_{m, n, \gamma}^\top \mathbf{A}_{m, n, \gamma} \right) \\ &\quad - 2\mathbf{w}^\top \sum_{m, n, \gamma} \mathbf{A}_{m, n, \gamma}^\top \tilde{\mathbf{x}}_{m, n, \gamma} + C, \end{aligned} \quad (10)$$

where  $C$  is the constant term. In particular, if one defines the following matrix and vector:

$$\mathbf{P} := \sum_{m,n,\gamma} \mathbf{A}_{m,n,\gamma}^\top \mathbf{A}_{m,n,\gamma}, \quad \mathbf{q} := \sum_{m,n,\gamma} \mathbf{A}_{m,n,\gamma}^\top \tilde{\mathbf{x}}_{m,n,\gamma}. \quad (11)$$

Then the objective can be simplified to

$$f(\mathbf{w}) = \text{tr}(\mathbf{w}\mathbf{w}^\top \mathbf{P}) - 2\mathbf{w}^\top \mathbf{q} + C. \quad (12)$$

Following that form, we encode the sparsity using binary variables  $\mathbf{z} \in \{0, 1\}^d$  and rewrite the MIO as follows,

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{z}} \quad & \text{tr}(\mathbf{w}\mathbf{w}^\top \mathbf{P}) - 2\mathbf{w}^\top \mathbf{q} \\ \text{s.t.} \quad & 0 \leq \mathbf{w} \leq \mathcal{M} \cdot \mathbf{z}, \quad \forall m, n, \gamma, \\ & \sum_{k=1}^d z_k \leq \tau, \\ & z_k \in \{0, 1\}, \quad \forall k \in [d]. \end{aligned} \quad (13)$$

Thus, the resulting global support set is given by

$$\Omega := \{k \in [d] \mid w_k > 0\} = \text{supp}(\mathbf{w}) = \text{supp}(\mathbf{z}). \quad (14)$$

### 6.3 Estimating Individual Coefficient Vectors

In practice, learning the global support set is advantageous because (i) the underlying periodicity structure can be quantified across  $M \times N$  spatial locations and  $\Gamma$  time segments, and (ii) the dominant indices of auto-correlations can be identified and estimated efficiently without having to find them explicitly for each individual coefficient vector. To learn the sparse coefficient vectors  $\{\mathbf{w}_{m,n,\gamma}\}_{m \in [M], n \in [N], \gamma \in [\Gamma]}$  within the given support set  $\Omega$  in Eq. (14), the optimization problem now becomes

$$\begin{aligned} \min_{\mathbf{w}_{m,n,\gamma}} \quad & \|\tilde{\mathbf{x}}_{m,n,\gamma} - \mathbf{A}_{m,n,\gamma} \mathbf{w}_{m,n,\gamma}\|_2^2 \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{w}_{m,n,\gamma}) \geq 0, \quad \mathcal{P}_\Omega^\perp(\mathbf{w}_{m,n,\gamma}) = 0, \end{aligned} \quad (15)$$

for all  $m \in [M], n \in [N], \gamma \in [\Gamma]$ . Here,  $\mathcal{P}_\Omega(\cdot)$  denotes the orthogonal projection supported on  $\Omega$ . For any vector  $\mathbf{w} \in \mathbb{R}^d$  with entries  $\{w_k\}_{k \in [d]}$ , the orthogonal projection takes  $[\mathcal{P}_\Omega(\mathbf{w})]_k = w_k$  if  $k \in \Omega$ ; otherwise,  $[\mathcal{P}_\Omega^\perp(\mathbf{w})]_k = 0$  for any  $k \notin \Omega$  in the complement of  $\Omega$ . That means that the  $k$ th entry of  $\mathbf{w}_{m,n,\gamma}$  is zero when satisfying  $k \in [d]$  and  $k \notin \Omega$  simultaneously. Thus, we only have  $|\Omega|$  entries in  $\mathbf{w}_{m,n,\gamma}$  to estimate. By doing so, we can solve this optimization problem by quadratic optimization with linear constraints.

## 7 EXPERIMENTS

In this section, we conduct extensive experiments to evaluate the proposed SAR models on real-world time series data. The time-varying ridesharing trip time series dataset in NYC allows one to discover daily and weekly periodicities and their changes across different periods from 2019 to 2023. The climate variable time series datasets, including North America climate variables and sea surface temperature, are spatially- and time-varying, enabling the tasks such as the spatial pattern discovery of yearly seasonality and the identification of system evolution related to climate dynamics.

### 7.1 Human Mobility Periodicity

#### 7.1.1 Ridesharing Trip Time Series

In this work, we examine the proposed TV-SAR model on the ridesharing trip data in NYC, ranging from February 2019 to December 2023. In terms of human mobility, trips to airports or leaving airports are closely related to the flight schedule, which usually demonstrate strong daily and weekly periodicities. Fig. 5 shows the pickup and dropoff trip time series of John F. Kennedy International Airport in NYC. In Figs. 5(a) and 5(b), the trip time series of each week is visualized as a row of the heatmap, and there are around 260 weeks from 2019 to 2023 in total. As can be seen, pickup trips show perturbation and fluctuation due to the influential factors such as flight delay, baggage claim, and travel time to ridesharing pickup zones. The peak time demanding ridesharing services is the nighttime. In contrast, dropoff trips to the departure area of the airport are uniform across different weeks. There are morning and afternoon peak hours with remarkable dropoff trips to the airport. Figs. 5(c) and 5(d) show the daily periodicity of aggregated time series of pickup and dropoff trips, in which the dropoff trips are more daily and weekly periodic than the pickup trips.

TABLE 2: Objective function  $f(\mathbf{w})$  in Eq. (4) on the ridesharing pickup trip time series in John F. Kennedy International Airport. The solution algorithms include NNSP, MIO-DVP, and MIO. The unit of objective function values is  $\times 10^7$ . Note that the lowest objective function values are emphasized in bold fonts. The last two rows present the average computational times (in seconds) of the algorithm.

Data	Sparsity	NNSP	MIO-DVP ( $\tau_0 = 5$ )	MIO-DVP ( $\tau_0 = 10$ )	MIO
2019	$\tau = 4$	8.48	8.48	<b>8.24</b>	<b>8.24</b>
	$\tau = 6$	8.41	-	<b>8.07</b>	<b>8.07</b>
2020	$\tau = 4$	2.12	2.12	<b>1.90</b>	<b>1.90</b>
	$\tau = 6$	2.03	-	<b>1.86</b>	<b>1.86</b>
2021	$\tau = 4$	3.11	3.11	<b>3.06</b>	<b>3.06</b>
	$\tau = 6$	3.06	-	<b>2.97</b>	<b>2.97</b>
2022	$\tau = 4$	6.85	6.76	<b>6.49</b>	<b>6.49</b>
	$\tau = 6$	6.69	-	<b>6.34</b>	<b>6.34</b>
2023	$\tau = 4$	8.59	8.45	<b>8.14</b>	<b>8.14</b>
	$\tau = 6$	8.39	-	<b>7.95</b>	<b>7.95</b>
Cost	$\tau = 4$	0.03 s	0.32 s	0.67 s	221 s
	$\tau = 6$	0.04 s	-	0.66 s	223 s

#### 7.1.2 Comparison among NNSP, MIO-DVP, and MIO

Basically, we examine three solution algorithms—NNSP, MIO-DVP, and MIO—for addressing Problem (4) on the ridesharing pickup trip time series across five years. The sparsity levels in these algorithms are set as  $\tau = 4, 6$ . In particular, we test the MIO-DVP algorithm with different sparsity levels  $\tau_0 = 5, 10$  for pruning decision variables, while NNSP is an important baseline for measuring the performance gain. Table 2 shows the objective function values of SAR with different algorithmic settings. As can be seen, the NNSP algorithm fails to find the solution as good as the MIO algorithm. The MIO-DVP algorithm performs better than NNSP if a well-designed pruning strategy such as NNSP with  $\tau_0 = 10$  is available. Herein, the sparsity

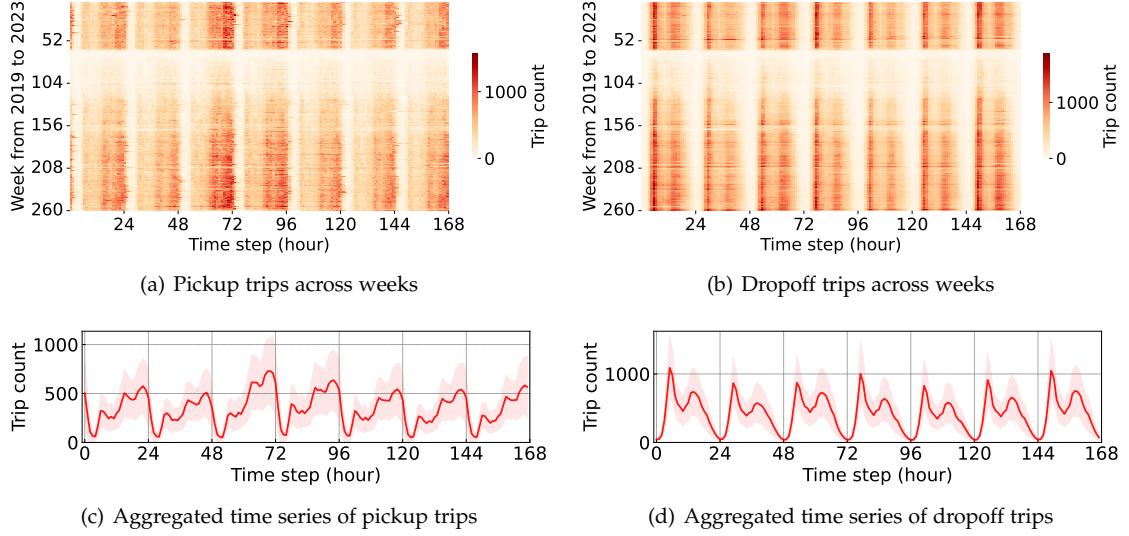


Fig. 5: Hourly time series of the aggregated ridesharing trip counts of John F. Kennedy International Airport in NYC from 2019 to 2023. (a-b) The row of each heatmap refers to the ridesharing trip time series of each week. (c-d) The time series refers to the average ridesharing trips of each hour within a week window, while the standard deviations are also presented.

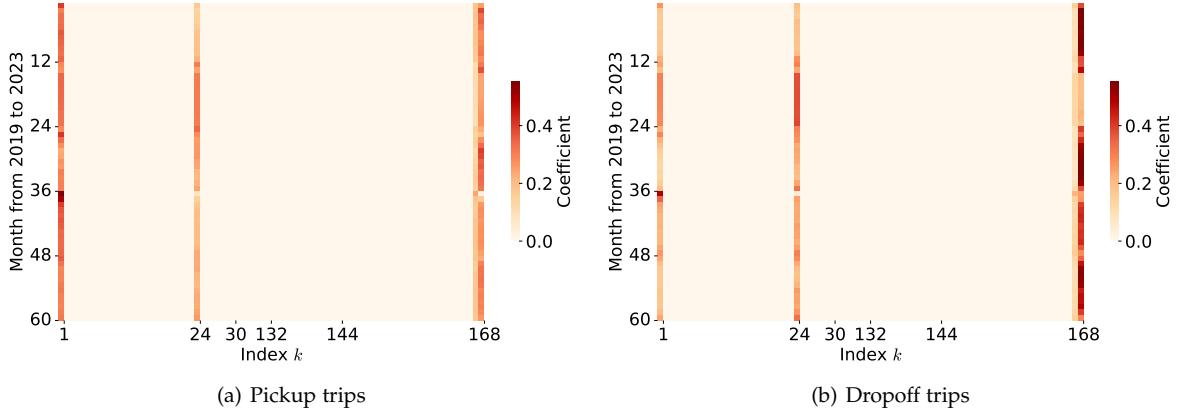


Fig. 6: Sparse coefficient vectors  $w_\gamma \in \mathbb{R}^{168}$ ,  $\gamma \in [2, 60]$  of TV-SAR on the ridesharing trip time series from February 2019 to December 2023, i.e., 59 months in total. Since each time series data corresponds to one month, there are 59 coefficient vectors that are represented as 59 rows in the heatmap. Each time series is accumulated from January to the end of the given month. Three most significant auto-correlations are revealed at  $k = 1, 24, 168$ , referring to local auto-correlation, daily periodicity, and weekly periodicity, respectively.

level  $\tau_0$  is an influential factor in the MIO-DVP algorithm for finding a good solution as MIO. In contrast to MIO, one remarkable advantage of MIO-DVP is the efficiency and scalability due to the reduction of search space.

### 7.1.3 Periodicity of Ridesharing Trips

As shown in Fig. 5, the ridesharing pickup/dropoff trip time series is time-varying in a long-term range. In the time series, one can observe the remarkable reduction of total trip counts in 2020 due to the COVID-19 pandemic, recovering slowly until the end of 2021. To identify such kind of time-varying system behaviors, we use the proposed TV-SAR model. The time series of each month is regarded as a time segment. The order of autoregression is set as  $d = 168$ , referring to a weekly cycle. The sparsity level is

set as  $\tau = 4$  which aims to cover the indices of daily and weekly periodicities at  $k = 24$  and  $k = 168$ , respectively.

Of the results in Fig. 6, the proposed model identifies the support set as  $\text{supp}(w_\gamma) = \{1, 24, 167, 168\}$ ,  $\forall \gamma \in [2, 60]$ , corresponding to local, daily, and weekly auto-correlations. Fig. 6(a) shows the periodicity pattern changes of pickup trips from 2019 to 2023. As can be seen, the strength reduction of weekly periodicity and the increasing strength of daily periodicity are remarkable in 2020. The similar results are also demonstrated in Fig. 6(b) for dropoff trips, namely, the decreasing strength of weekly periodicity and the increasing strength of daily periodicity in 2020. Another finding from Fig. 6(b) is the increasing local auto-correlations at  $k = 1$  in 2020, implying increasing variability in the dropoff trips. Observing the coefficients at  $k = 168$ ,

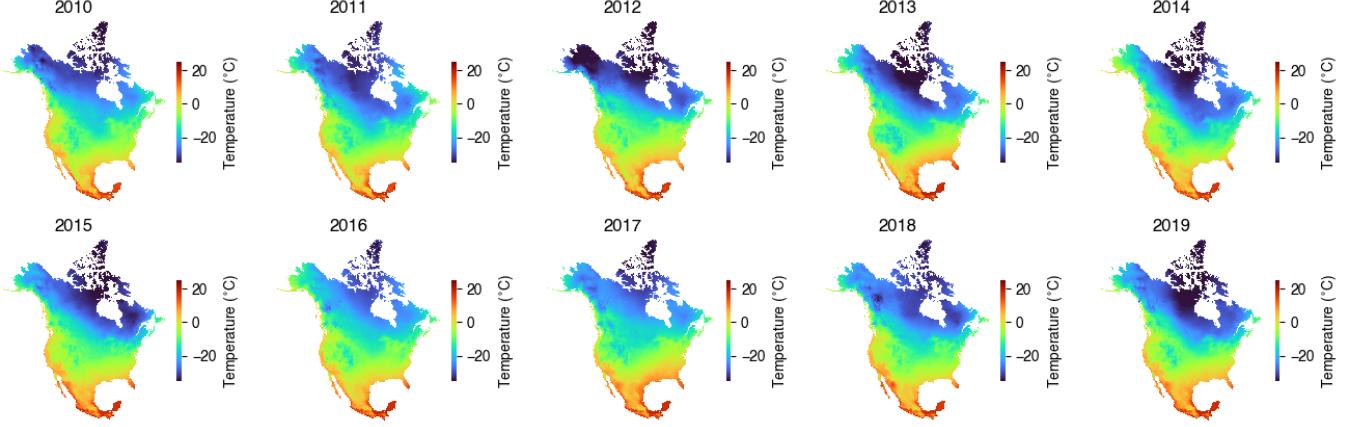


Fig. 7: Monthly aggregated minimum temperature across North America of January from 2010 to 2019. The color scale represents the temperature in degrees Celsius, with blue tones indicating lower temperatures and red tones indicating higher temperatures.

the weekly periodicity of dropoff trips in 2019, 2021, 2022, and 2023 are stronger than pickup trips, and such evidence is consistent with pickup and dropoff trip time series in Fig. 5. Although the periodicity patterns in 2020 is very different from the years before and after COVID-19, the local and nonlocal auto-correlations in 2019, 2021, 2022, and 2023 demonstrate no remarkable difference. This implies that the regularity of human mobility after COVID-19 has recovered to 2019.

## 7.2 Seasonality Patterns of North America Climate Variables

Daymet provides monthly climate variables (e.g., minimum/maximum temperature, and precipitation) at a spatial resolution of  $1\text{ km} \times 1\text{ km}$  [35].<sup>1</sup> In this work, we consider the data spanning from 1980 to 2019, which is in the form of multidimensional time series. Fig. 7 shows the year-to-year variations and the temporal variability of the minimum temperature data of January in the past decade. These panels reveal seasonal temperature patterns and changes over the decade, demonstrating regional climate trends.

### 7.2.1 Robustness of Seasonality Patterns with Different Spatial Resolutions

Since the original dataset is overly high-resolution, we aggregate the data with different spatial resolutions, including  $5\text{ km} \times 5\text{ km}$ ,  $10\text{ km} \times 10\text{ km}$ , and  $20\text{ km} \times 20\text{ km}$ . To quantify yearly seasonality of these climate time series, we use the proposed STV-SAR model with a prescribed sparsity level  $\tau = 3$ . For instance, on the minimum temperature dataset, we build the monthly time series across  $M \times N$  grids of each decade. The global support set is optimized as  $\Omega = \{1, 11, 12\}$  with the sparsity level  $\tau = 3$ , and the coefficients at the index  $k = 12$  are used to quantify the yearly seasonality of monthly time series. In particular, we claim that the global support set can characterize temporal structures of climate variable time series. On the minimum temperature data of  $20\text{ km} \times 20\text{ km}$  resolution, Fig. 8 shows

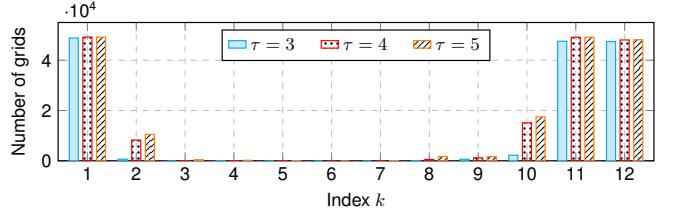


Fig. 8: Support sets of independent SAR models with different sparsity levels tested on the minimum temperature data in 2010s across 49,180 grids.

support sets of independent SAR models with sparsity levels  $\tau = 3, 4, 5$  across 49,180 time series. Although there is no constraint for the consistent support set, the dominant indices of sparsity level  $\tau = 3$  include  $\{1, 11, 12\}$  which are the same as the global support set  $\Omega$  as mentioned above. With the increase of sparsity levels, the indices include  $\{1, 2, 10, 11, 12\}$  and the dominant indices are also  $\{1, 11, 12\}$ . Therefore, the implication of learning global support set can be verified, instead of only mentioning the efficiency and scalability benefits of that process.

Fig. 9 shows that the strengths of yearly seasonality with different spatial resolutions are consistent, in which a relatively higher coefficient implies stronger seasonality. In 2010s, both spatial patterns of yearly seasonality highlight some very seasonal regions such as north areas of Canada, western areas of USA, and southeastern areas of USA. The temperature of central regions in North America is extremely less seasonal than other regions. As shown in Fig. 9, temperature data of Mexico demonstrate remarkable variability and inconsistency among nearby areas, one can observe detailed spatial patterns of temperature seasonality from Fig. 10(a).

To address the computational challenges posed by the large-scale MIO problem in Eq. (9), we propose a two-stage optimization scheme aimed at improving scalability in the presence of thousands to millions of decision variables. Table 3 summarizes the model's running times at three spatial resolutions. As the results show, the proposed scheme

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

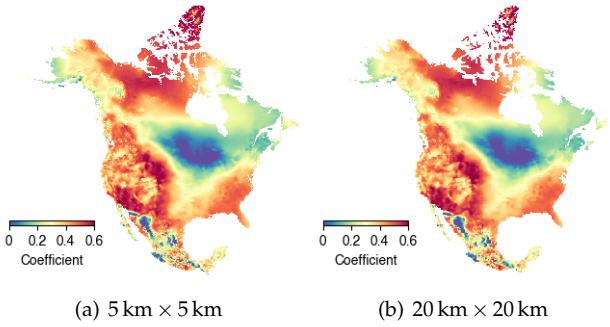


Fig. 9: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the minimum temperature data in 2010s.

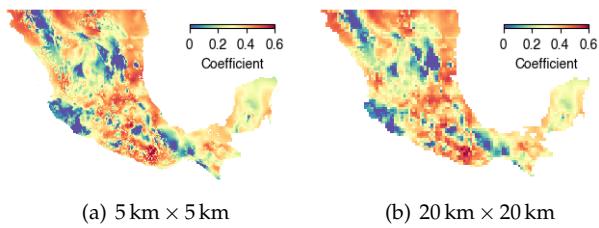


Fig. 10: Spatial patterns of the strengths of yearly seasonality on the minimum temperature data in 2010s within Mexico.

enables efficient computation even for datasets with up to 3 million time series, making it well-suited for large-scale, real-world applications.

TABLE 3: Running times of STV-SAR on the North America (minimum) temperature dataset with different spatial resolutions. Note that  $\nu$  denotes the number of time series.

	5 km × 5 km	10 km × 10 km	20 km × 20 km
Number $\nu$	3,343,628	816,612	196,720
Cost	48.55 s	11.77 s	2.88 s

### 7.2.2 Variations of Seasonality Patterns across Four Decades

Discovering the variations of seasonality patterns allows one to quantify the evolution of dynamical climate systems. Fig. 11 shows yearly seasonality of monthly time series data in North America across past four decades with a spatial resolution of 10 km × 10 km. Recall that higher coefficients in the panels indicate regions with stronger yearly seasonality. We first summarize the findings from the seasonality patterns of minimum and maximum temperature data. Over the decades, regions with stronger yearly seasonality are predominantly located in northern and high-latitude areas such as Canada and Alaska. The southern regions such as Mexico seem to be less seasonal in the past four decades than northern regions, and the variation of yearly seasonality in Mexico is remarkable among relatively small areas. A remarkable expansion of highly seasonal regions occurs in 2000s compared to earlier decades, demonstrating an intensification of seasonal temperature variations. Such kind of

expansion of regions with a strong yearly seasonality highlights area where temperature-driven seasonal cycles are more predictable. In 2010s, two most remarkable regions are highlighted, i.e., north territory of Canada (e.g., Nunavut) and western states of USA. While the seasonality patterns in 1980s and 1990s are quite consistent, the changing seasonality patterns in 2000s and 2010s are truly remarkable. The shifts in seasonality patterns in the past four decades provide evidence of changing climate dynamics in North America with implications for ecological systems.

Then, we compare the seasonality patterns among different climate variables in Fig. 11. The yearly seasonality patterns of precipitation demonstrate remarkable variations in the past four decades even among nearby small regions. The seasonality patterns of minimum and maximum temperature in 1980s and 1990s are quite consistent. However, the seasonality pattern of the minimum temperature in 2000s shows more seasonal regions than the maximum temperature. In 2010s, the maximum temperature of Mexico is less seasonal than the minimum temperature. the maximum temperature of western regions of USA and central regions of North America is more seasonal than the minimum temperature. In contrast, the maximum temperature of northern regions of Canada is less seasonal than the minimum temperature.

### 7.2.3 Sensitivity Analysis of Phase Segmentation

As mentioned above, we discover the yearly seasonality from monthly climate variable time series with each phase corresponding to each decade. In what follows, we evaluate the model with a 5-year phase on the minimum temperature data from 2000 to 2019. Fig. 12(a) shows the seasonality pattern that is more consistent with Fig. 11(a) for 1990s. Fig. 12(b) depicts the seasonality pattern from 2005 to 2009 that is dominant in 2000s, see Fig. 11(a). Observing the seasonality pattern in Fig. 12(c), more regions around the central part of North America show less seasonal minimum temperature from 2010 to 2014, comparing with the seasonality pattern in the next 5 years as shown in Fig. 12(d). Notably, as the minimum temperature of Mexico from 2015 to 2019 is more seasonal than the periods from 2000 to 2014, it is difficult to see such an evidence from Fig. 11.

## 7.3 Seasonality Patterns of Sea Surface Temperature

In this work, we consider the sea surface temperature dataset that covers monthly averages of temperatures on the spatial resolution of a (0.25 degree latitude, 0.25 degree longitude)-grid, and there are 720 × 1440 global grids (i.e., 1,036,800 cells) in total.<sup>2</sup> The dataset spans a past four-decade period from January 1982 to December 2019. Fig. 13 shows both monthly and yearly temperature trends of the global sea surface temperature, demonstrating a gradual warming trend as evident from the increase in yearly average temperatures. These results highlight the increasing influence of climate change on ocean temperatures. The higher fluctuations in the monthly time series reveal seasonal variations. Thus, quantifying yearly seasonality is

2. <https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2.1/access/avhrr/>

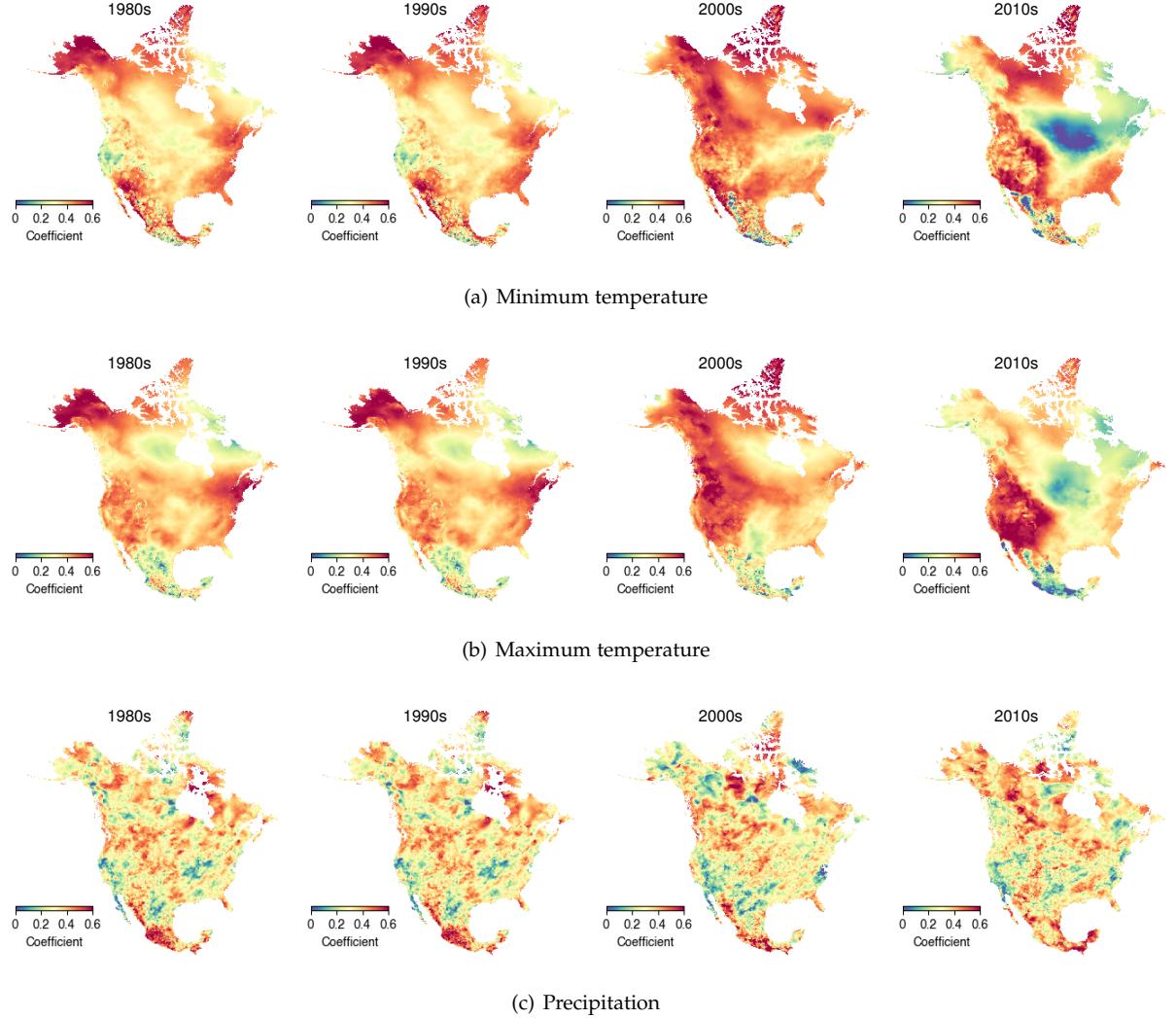


Fig. 11: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the North America climate data across the past four decades.

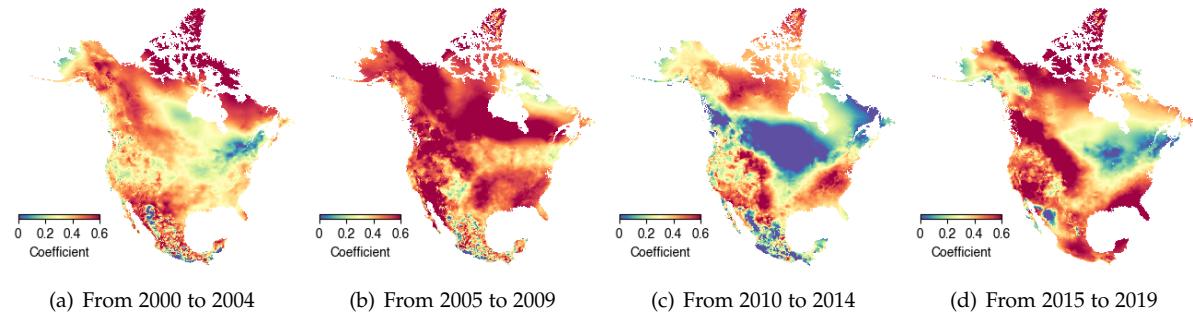


Fig. 12: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the minimum temperature data with a 5-year phase (i.e.,  $T_\gamma = 60$  months).

important for measuring the climate change and understanding the dynamical system because stronger seasonality of temperatures implies less variations in the time series.

In the experiment, we first set the order and the sparse level as  $d = 12$  and  $\tau = 3$ , respectively. We examine the STV-SAR model on the sea surface temperature data

of different decades independently and compare the coefficients of yearly seasonality of temperatures. Fig. 14 explores the yearly periodicity of sea surface temperature through a spatial representation of the coefficients associated with yearly seasonality. These results show the intensity of yearly sea surface temperature variations, with higher coefficients

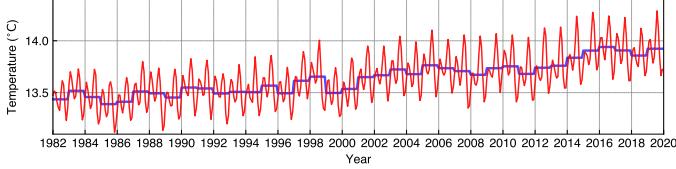


Fig. 13: Average values of the monthly sea surface temperature data from January 1982 to December 2019. The average temperature values in 1980s, 1990s, 2000s, and 2010s are  $13.46^{\circ}\text{C}$ ,  $13.54^{\circ}\text{C}$ ,  $13.69^{\circ}\text{C}$ , and  $13.83^{\circ}\text{C}$ , respectively. The blue and red curves correspond to yearly and monthly average temperature values, respectively.

indicating stronger periodicity. High values, marked in red, are concentrated in regions close to Eastern/Middle Asia, North America, Europe, and North Africa, where seasonal temperature changes are more pronounced. In contrast, the equatorial regions such as the area of El Niño exhibit lower coefficients, demonstrating less seasonal and predictable sea surface temperature. Although there is a consistent pattern of strong periodicity across the decades, the minor shifts indicating changes in oceanic dynamics can also be identified. For instance, as shown in Fig. 15, these regions close to Canada have temperature variations and seasonality shift in the past three decades. One can see from the coefficients that the temperature of Arctic Ocean became less seasonal in the past two decades, potentially influenced by factors such as oceanic circulation changes, global warming, or variability in atmospheric patterns.

## 8 CONCLUSION

In this work, we provide a unified SAR framework for quantifying time series periodicity in real-world spatiotemporal systems. We advance the development of TV-SAR and STV-SAR models from an interpretable machine learning perspective by incorporating spatially- and time-varying sparsity constraints. To address the MIO problems, we present efficient and scalable algorithmic implementation by designing a DVP strategy and a two-stage optimization scheme for TV-SAR and STV-SAR, respectively. Our findings discovered by the proposed models from ridesharing trip data and climate variable data empirically demonstrate the interpretable components related to periodicity and seasonality of time series. The spatial and temporal dynamics and patterns revealed by time series periodicity also justify remarkable real-world implications of periodicity quantification to spatiotemporal systems.

## APPENDIX A PROOF OF THEOREM 1

*Proof.* Let  $\mathcal{F}_\tau = \{\mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\|_0 \leq \tau\}$  be the class of  $\tau$ -sparse vectors. Since  $\hat{\mathbf{w}}$  minimizes the residual error over  $\mathcal{F}_\tau$ , we have:

$$\|\tilde{\mathbf{x}} - \mathbf{A}\hat{\mathbf{w}}\|_2^2 \leq \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}^*\|_2^2.$$

Substituting  $\tilde{\mathbf{x}} = \mathbf{A}\mathbf{w}^* + \boldsymbol{\epsilon}$ , this becomes:

$$\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*) - \boldsymbol{\epsilon}\|_2^2 \leq \|\boldsymbol{\epsilon}\|_2^2.$$

Expanding the left-hand side:

$$\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2 - 2\langle \boldsymbol{\epsilon}, \mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*) \rangle + \|\boldsymbol{\epsilon}\|_2^2 \leq \|\boldsymbol{\epsilon}\|_2^2.$$

Cancelling  $\|\boldsymbol{\epsilon}\|_2^2$  from both sides:

$$\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2 \leq 2\langle \boldsymbol{\epsilon}, \mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*) \rangle.$$

Now take expectations on both sides:

$$\mathbb{E} [\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2] \leq 2 \cdot \mathbb{E} \left[ \sup_{\mathbf{w} \in \mathcal{F}_\tau} \langle \boldsymbol{\epsilon}, \mathbf{A}(\mathbf{w} - \mathbf{w}^*) \rangle \right].$$

Let  $\mathcal{S}_\tau = \{\mathbf{A}(\mathbf{w} - \mathbf{w}^*) : \mathbf{w} \in \mathcal{F}_\tau\}$ . We now apply a Gaussian complexity bound:

$$\mathbb{E} \left[ \sup_{\mathbf{v} \in \mathcal{S}_\tau} \langle \boldsymbol{\epsilon}, \mathbf{v} \rangle \right] \leq \sigma \cdot \sqrt{2 \log(|\mathcal{F}_\tau|)} \cdot \sup_{\mathbf{v} \in \mathcal{S}_\tau} \|\mathbf{v}\|_2.$$

The number of  $\tau$ -sparse supports is bounded by:

$$|\mathcal{F}_\tau| \leq \binom{d}{\tau} \leq \left( \frac{ed}{\tau} \right)^\tau.$$

Taking logarithms:

$$\log |\mathcal{F}_\tau| \leq \tau \log \left( \frac{ed}{\tau} \right) \leq \tau \left( \log \left( \frac{d}{\tau} \right) + 1 \right).$$

Putting everything together:

$$\mathbb{E} [\|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2] \leq C \cdot \sigma^2 \cdot \tau \left( \log \left( \frac{d}{\tau} \right) + 1 \right),$$

for some universal constant  $C$ . Finally, divide both sides by  $T - d$  to normalize by the number of samples:

$$\mathbb{E} \left[ \frac{1}{T-d} \|\mathbf{A}(\hat{\mathbf{w}} - \mathbf{w}^*)\|_2^2 \right] \leq C \cdot \sigma^2 \cdot \frac{\tau}{T-d} \left( \log \left( \frac{d}{\tau} \right) + 1 \right). \quad \square$$

## ACKNOWLEDGMENTS

This research is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Vehicle Technology Program Award Number DE-EE0009211 and DE-EE0011186. The Mens, Manus, and Machina (M3S) is an interdisciplinary research group (IRG) of the Singapore MIT Alliance for Research and Technology (SMART) center.

## REFERENCES

- [1] S. Pezzulli, D. Stephenson, and A. Hannachi, "The variability of seasonality," *Journal of Climate*, vol. 18, no. 1, pp. 71–88, 2005.
- [2] X. Feng, A. Porporato, and I. Rodriguez-Iturbe, "Changes in rainfall seasonality in the tropics," *Nature Climate Change*, vol. 3, no. 9, pp. 811–815, 2013.
- [3] J. D. Tonkin, M. T. Bogan, N. Bonada, B. Rios-Touma, and D. A. Lytle, "Seasonality and predictability shape temporal species diversity," *Ecology*, vol. 98, no. 5, pp. 1201–1216, 2017.
- [4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [5] J. D. Hamilton, *Time series analysis*. Princeton university press, 2020.
- [6] D. Bertsimas, V. Digalakis Jr, M. L. Li, and O. S. Lami, "Slowly varying regression under sparsity," *Operations Research*, 2024.
- [7] X. Chen, H. Cai, F. Liu, and J. Zhao, "Correlating time series with interpretable convolutional kernels," *IEEE Transactions on Knowledge and Data Engineering*, 2025.

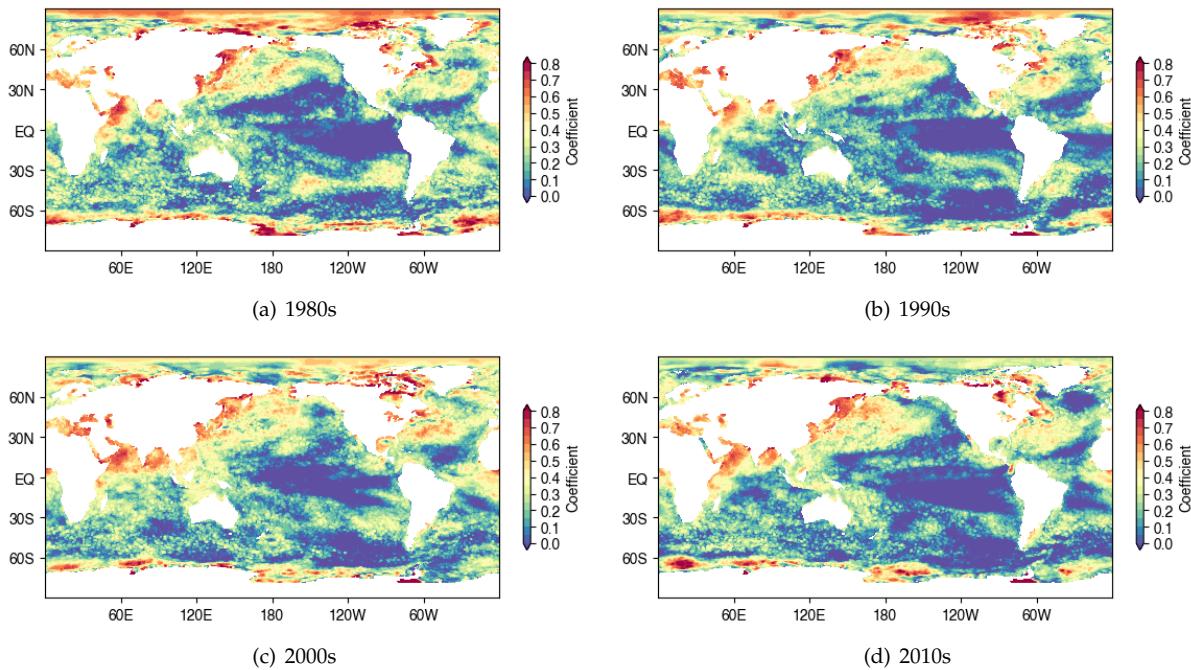


Fig. 14: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the sea surface temperature data in the past four decades.

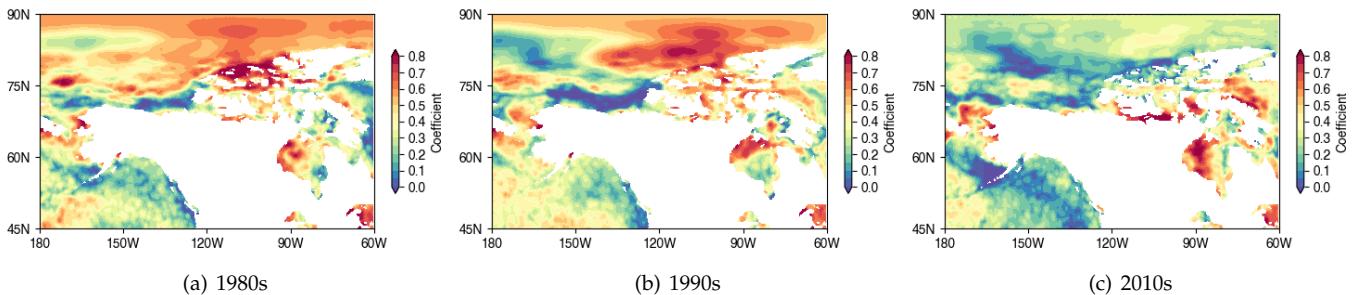


Fig. 15: Spatial patterns of the strengths of yearly seasonality on the sea surface temperature data with highlighted areas around Canada.

- [8] X. Chen, C. Zhang, X. Chen, N. Saunier, and L. Sun, "Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 504–517, 2024.
- [9] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.
- [10] J. H. Cochrane, "Time series for macroeconomics and finance," 1997.
- [11] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. Springer science & business media, 1991.
- [12] J. M. Haslbeck, L. F. Bringmann, and L. J. Waldorp, "A tutorial on estimating time-varying vector autoregressive models," *Multivariate behavioral research*, vol. 56, no. 1, pp. 120–149, 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] L. F. Bringmann, E. L. Hamaker, D. E. Vigo, A. Aubert, D. Borsboom, and F. Tuerlinckx, "Changing dynamics: Time-varying autoregressive models using generalized additive modeling," *Psychological methods*, vol. 22, no. 3, p. 409, 2017.
- [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] Y. Nardi and A. Rinaldo, "Autoregressive process modeling via the lasso procedure," *Journal of Multivariate Analysis*, vol. 102, no. 3, pp. 528–549, 2011.
- [17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.
- [18] R. A. Davis, P. Zang, and T. Zheng, "Sparse vector autoregressive modeling," *Journal of Computational and Graphical Statistics*, vol. 25, no. 4, pp. 1077–1096, 2016.
- [19] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the national academy of sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.
- [20] D. Bertsimas and W. Gurnee, "Learning sparse nonlinear dynamics via mixed-integer optimization," *Nonlinear Dynamics*, vol. 111, no. 7, pp. 6585–6604, 2023.
- [21] J. Liu, S. Rosen, C. Zhong, and C. Rudin, "Okr ridge: Scalable optimal k-sparse ridge regression," *Advances in neural information processing systems*, vol. 36, 2024.
- [22] E. Carrizosa, A. V. Olivaress-Nadal, and P. Ramírez-Cobo, "A sparsity-controlled vector autoregressive model," *Biostatistics*, vol. 18, no. 2, pp. 244–259, 2017.
- [23] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with appli-

- cations to wavelet decomposition," in *Proceedings of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.
- [24] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [25] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [26] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens, "A simple new approach to variable selection in regression, with application to genetic fine mapping," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 82, no. 5, pp. 1273–1300, 2020.
- [27] D. Bertsimas, A. King, and R. Mazumder, "Best subset selection via a modern optimization lens," 2016.
- [28] D. Bertsimas and B. Van Parys, "Sparse high-dimensional regression," *The Annals of Statistics*, vol. 48, no. 1, pp. 300–323, 2020.
- [29] H. Hazimeh, R. Mazumder, and A. Saab, "Sparse regression at scale: Branch-and-bound rooted in first-order optimization," *Mathematical Programming*, vol. 196, no. 1, pp. 347–388, 2022.
- [30] A. Atamturk and A. Gomez, "Rank-one convexification for sparse regression," *Journal of Machine Learning Research*, vol. 26, no. 35, pp. 1–50, 2025.
- [31] F. Bach, *Learning theory from first principles*. MIT press, 2024.
- [32] D. Bertsimas and V. Digalakis Jr, "The backbone method for ultrahigh dimensional sparse machine learning," *Machine Learning*, vol. 111, no. 6, pp. 2161–2212, 2022.
- [33] J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 70, no. 5, pp. 849–911, 2008.
- [34] A. Atamturk and A. Gómez, "Safe screening rules for l0-regression from perspective relaxations," in *International conference on machine learning*. PMLR, 2020, pp. 421–430.
- [35] M. Thornton, R. Shrestha, Y. Wei, P. Thornton, and S.-C. Kao, "Daymet: Monthly climate summaries on a 1-km grid for north america, version 4 r1," 2022. [Online]. Available: [https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds\\_id=2131](https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2131)



**Lijun Ding** is an Assistant Professor at the Department of Mathematics at the University of California, San Diego. He earned his PhD in Operations Research from Cornell University in 2021. He specializes in continuous optimization, especially in semidefinite programming and applications of optimization in statistics.



**Dingyi Zhuang** is a Ph.D. student in Transportation Engineering at MIT Urban Mobility Lab. He received a B.S. degree in Mechanical Engineering from Shanghai Jiao Tong University in 2019, and an M.Eng. degree in Transportation Engineering from McGill University, in 2021. His research interests lie in deep learning, urban computing, and spatiotemporal data modeling.



**Jinhua Zhao** is currently the Professor of Cities and Transportation at MIT. He brings behavioral science and transportation technology together to shape travel behavior, design mobility systems, and reform urban policies. He directs the MIT Urban Mobility Laboratory and Public Transit Laboratory.



**Xinyu Chen** is a Postdoctoral Associate at Massachusetts Institute of Technology, Cambridge, MA, United States. He received his Ph.D. degree from the University of Montreal, Montreal, QC, Canada. His current research centers on machine learning, spatiotemporal data modeling, intelligent transportation systems, and urban science.



**Vassilis Digalakis Jr** is an Assistant Professor of Operations & Technology Management at the O&T Management Department at Boston University's Questrom School of Business. He received his Ph.D. in Operations Research at the Operations Research Center at MIT. His research develops methods in trustworthy AI and machine learning, with applications in healthcare and sustainability.