

Interpretable Time Series Autoregression for Periodicity Quantification

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

Abstract—Time series autoregression (AR) is a classical statistical tool for modeling auto-correlations and periodic structures in real-world systems. We revisit this model from an interpretable machine learning perspective by introducing sparse autoregression (SAR), where ℓ_0 -norm constraints are used to isolate dominant periodicities. We formulate exact mixed-integer optimization (MIO) approaches for both stationary and non-stationary settings and introduce two scalable extensions: a decision variable pruning (DVP) strategy for temporally-varying SAR (TV-SAR), and a two-stage optimization scheme for spatially- and temporally-varying SAR (STV-SAR). These models enable efficient inference on real-world spatiotemporal datasets. We validate our framework on large-scale mobility and climate time series. On NYC ridesharing data, TV-SAR reveals interpretable daily and weekly cycles as well as long-term shifts due to COVID-19. On climate datasets, STV-SAR uncovers the evolving spatial structure of temperature and precipitation seasonality across four decades in North America and detects global sea surface temperature dynamics, including El Niño. Together, our results demonstrate the interpretability, flexibility, and scalability of sparse autoregression for periodicity quantification in complex 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 and temporal regularities, 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 such as policy interventions, demand changes, service disruption, environmental variability, global events, and extreme climate phenomena, making interpretable, adaptive, and data-driven methods indispensable. A central challenge is to automatically identify dominant periodic components from spatially- and temporally-varying systems, track their evolution over a long-term time period, and distinguish true structural changes from random variability.

Urban transportation systems display strong temporal regularity driven by commuting travels, business cycles, and travel demand. As a common sense, mobility regularity patterns are often governed by the daily, weekly, and multiple periodic cycles. However, such patterns are not fixed—they evolve with 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 temporally-varying system?* Addressing these questions is critical for planning, forecasting, and adaptive resource allocation.

Climate systems also exhibit periodic and seasonal patterns that shape temperature, precipitation, 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 extreme events, and supporting policy-making. This calls for *interpretable and 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 quantifying spatially- and temporally-varying periodicity and structural shifts in real-world time series data. Our models build on classical autoregression (AR) [4], [5] but incorporate sparse structure via ℓ_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:

- 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)

- **Sparse Autoregression (SAR):** We introduce an interpretable framework for identifying dominant auto-correlations from time series by reformulating AR with ℓ_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.
- **Temporally-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 Temporally-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 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 Temporally-Varying Autoregression

Time series modeling is a foundational tool in transportation, climate science, econometrics, and other fields [8], [9], [10]. Classical models such as 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 temporally-varying systems. Temporally-varying AR allows 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 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 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 ℓ_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 (OMP) 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 ℓ_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 real-world large-scale 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 $\{i, i+1, \dots, j\}, i < j$
$\ \cdot\ _0$	ℓ_0 -norm (number of nonzero entries)
$\ \cdot\ _1$	ℓ_1 -norm (sum of absolute values)
$\ \cdot\ _2$	ℓ_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, leading to the notion as AR(d), and $\mathbf{w} = (w_1, w_2, \dots, w_d)^\top \in \mathbb{R}^d$ is the coefficient vector. As the residual, ϵ_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:

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

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 Δt , 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 Δt manifests as a large positive coefficient $w_{\Delta t}$, allowing us to infer dominant cycles directly from the estimated sparse coefficient vector \mathbf{w} .

Figure 1(a) shows the periodicity of ridesharing activity in Chicago, highlighting its weekly mobility patterns. On the given time series, Figure 1(b) illustrates the coefficient vector \mathbf{w} estimated 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.

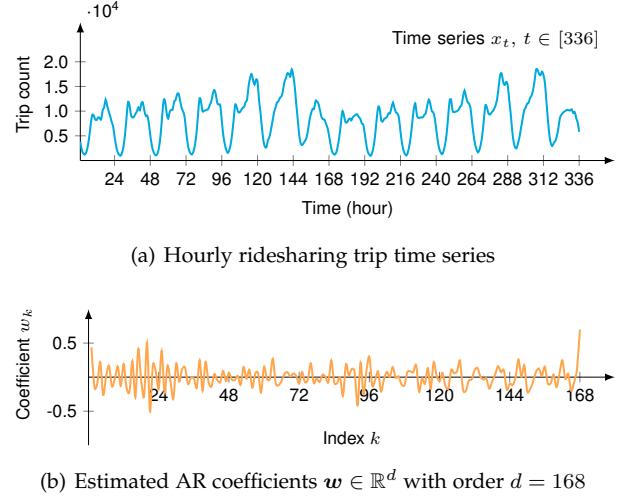


Fig. 1: Hourly time series and AR coefficients of aggregated ridesharing trip counts in Chicago during the first two weeks (336 hours) starting April 1, 2024. The data exhibits strong periodicity with a weekly cycle $\Delta t = 7 \times 24 = 168$.

4 SPARSE AUTOREGRESSION

This section introduces Sparse Autoregression (SAR) with ℓ_0 -norm induced sparsity constraints for identifying dominant auto-correlations in time series. We first describe the modeling framework; then, we present an MIO formulation for solving the resulting optimization problem exactly and compare solution quality across algorithms.

4.1 Model Description

The ℓ_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 as the upper bound of AR coefficients, and $\tau \in \mathbb{Z}^+$ controls the maximum number of nonzero coefficients. The non-negativity constraint

encourages interpretability by focusing on positive autocorrelations, as is typical for periodicity in time series.

To ensure statistical validity and numerical stability, we select the constant \mathcal{M} in Eq. (4) so that the underlying AR(d) model is stationary. Stationarity requires all roots of the characteristic polynomial $\phi(\zeta) = 1 - w_1\zeta - w_2\zeta^2 - \dots - w_d\zeta^d$ to lie strictly outside the unit circle, i.e., $|r_i| > 1$ for all $i \in [d]$. Under the non-negativity constraint $w_k \geq 0$, this already imposes a useful necessary condition: plugging in $\zeta = 1$, we obtain $\phi(1) = 1 - \sum_{k=1}^d w_k$. Hence, to avoid unit roots, it is necessary that $\sum_{k=1}^d w_k < 1$, so we set $\mathcal{M} = 1$.

For low-order models, we can derive more refined coefficient bounds using classical results on polynomial root geometry. To do so, we consider the reciprocal polynomial $\tilde{\phi}(\zeta) = w_d + w_{d-1}\zeta + \dots + w_1\zeta^{d-1} - \zeta^d$, whose roots are the reciprocals of those of $\phi(\zeta)$. Thus, the condition $|r_i| > 1$ on the original polynomial is equivalent to $|\tilde{r}_i| < 1$ for all roots \tilde{r}_i of $\tilde{\phi}$. To ensure this, we apply Fujiwara's bound [31], which provides an upper bound on the absolute values of the roots of a degree- d polynomial:

$$|\tilde{r}_i| \leq 2 \max_{1 \leq k \leq d} |w_k|^{1/k}.$$

To ensure $|\tilde{r}_i| < 1$, it suffices that $\max_k |w_k|^{1/k} < 2^{-1}$, or equivalently: $|w_k| < 2^{-k}$, for all $k \in [d]$. We therefore set $\mathcal{M} = (2^{-1}, 2^{-2}, \dots, 2^{-d})^\top \in \mathbb{R}^d$, which guarantees that all estimated models satisfy the stationarity condition. However, such bounds—derived from worst-case root behavior—can be overly conservative, especially for high-order AR models.

In the next section, we relax the global stationarity assumption by segmenting non-stationary time series into locally stationary phases, thereby enabling more flexible and data-adaptive modeling.

4.2 MIO Reformulation

Due to the combinatorial nature of the ℓ_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 set 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:

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

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

Problem (5) is nonconvex due to the ℓ_0 -norm induced sparsity constraint and is generally NP-hard to solve exactly. In principle, enumerating all support sets of cardinality τ would incur a combinatorial cost of $\mathcal{O}(d^\tau)$, making exact

search infeasible for large d or τ . 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 Figure 1(a), with AR order $d = 168$ (one week of hourly lags) and sparsity level $\tau = 2$. The solution returned by NNSP is:

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

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

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

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. As can be seen, the choice of solution algorithms affect not only the objective function of optimization problem but the interpretability of sparse coefficient vector as well.

When increasing the sparsity level to $\tau = 3$, both solvers return the coefficient vector such that

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

highlighting strong lags near a weekly cycle.

Figure 2 summarizes how the selected support sets and coefficient magnitudes evolve as a function of τ . These results confirm that high-quality solutions are critical for interpretability and that MIO offers robust, principled support recovery even when greedy methods fail.

5 TEMPORALLY-VARYING SPARSE AUTOREGRESSION

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

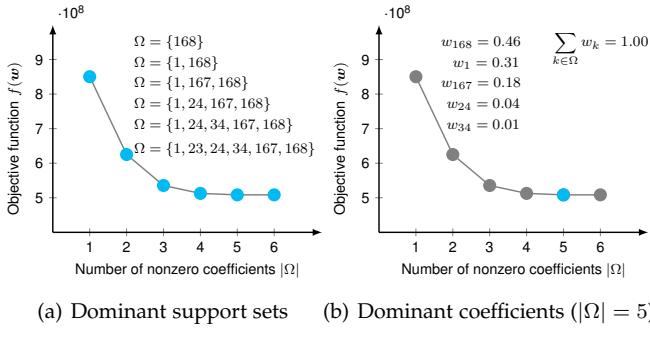


Fig. 2: Illustration of the dominant coefficients of SAR on the Chicago ridesharing trip time series (see Figure 1(a)) at different sparsity levels. The support set and the number of nonzero coefficients are denoted by Ω and $|\Omega|$, respectively.

5.1 Model Description

Temporally-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 time series corresponding to time segment γ . We model each time series with an independent AR(d) process, with potentially different coefficient vectors $\mathbf{w}_\gamma \in \mathbb{R}^d$:

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

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}_γ is sparse and non-negative, and (ii) all \mathbf{w}_γ 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 γ , $\mathbf{M} \in \mathbb{R}^d$ consists of all sufficiently large constant $\mathcal{M} > 0$, and $\tau \in \mathbb{Z}^+$ controls global sparsity. Here, $\text{supp}(\cdot)$ denotes the support set of vector, namely, the index set of nonzero entries in the vector. The cardinality of $\text{supp}(\mathbf{w})$ for any $\mathbf{w} \in \mathbb{R}^d$ is defined as $|\text{supp}(\mathbf{w})| = \sum_{k=1}^d \mathbb{1}\{\mathbf{w}_k \neq 0\} = \|\mathbf{w}\|_0$.

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 γ , 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 Γd real-valued variables for the AR coefficients \mathbf{w}_γ and d binary variables in \mathbf{z} . When the sparsity level τ 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 preselect a reduced candidate set of variables before solving the MIO. The strategy proceeds in three steps (illustrated in Figure 3):

- **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 Γ 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_γ 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.

We integrate the DVP strategy into an MIO solver and refer to the resulting hybrid method as MIO-DVP, parameterized by τ_0 . The approach significantly reduces the search space—especially for large d (e.g., $d = 168$ for weekly 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.

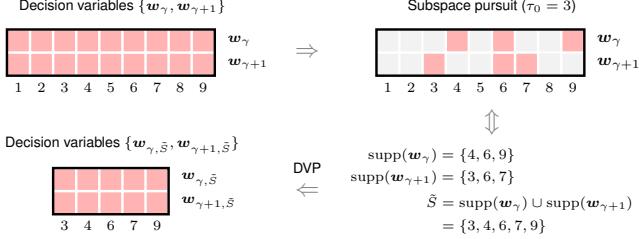


Fig. 3: Illustration of the DVP strategy. Subspace pursuit is used to select an index set \tilde{S} of candidate lags. The final MIO is then solved on this reduced support set, where the number of coefficients is reduced from $2d$ to $2|\tilde{S}|$.

Algorithm 1 DVP 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 $\mathbf{A}_\gamma \in \mathbb{R}^{(T-d) \times d}$.
- 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 ℓ : index set of τ largest entries in $|\mathbf{A}_\gamma^\top \mathbf{r}|$.
- 7: Update support set: $S_\gamma \leftarrow S_\gamma \cup \ell$.
- 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 τ largest entries of \mathbf{w}_γ , zero out the rest.
- 10: Update S_γ 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 $\tilde{S} := \bigcup_{\gamma=1}^{\Gamma} S_\gamma$.

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 TEMPORALLY-VARYING SPARSE AUTOREGRESSION

In this section, we introduce the Spatially and Temporally-Varying Sparse Autoregression (STV-SAR) model, a generalization of TV-SAR that accounts for both temporal and spatial variations in AR. While TV-SAR captures temporally-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 $\{x_{m,n,\gamma,t}\}_{t \in [T_\gamma], \gamma \in [\Gamma]}$, in which any γ th univariate time series is of length T_γ . Following the same logic as in Section 5, we model each time series with an AR(d) 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 with sparsity level τ 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, and \mathcal{M} is the upper bound.

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 have

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

where C is the constant term. In particular, we define the following matrix and vector:

$$\begin{aligned} \mathbf{P} &:= \sum_{m,n,\gamma} \mathbf{A}_{m,n,\gamma}^\top \mathbf{A}_{m,n,\gamma} \in \mathbb{R}^{d \times d}, \\ \mathbf{q} &:= \sum_{m,n,\gamma} \mathbf{A}_{m,n,\gamma}^\top \tilde{\mathbf{x}}_{m,n,\gamma} \in \mathbb{R}^d, \end{aligned} \quad (11)$$

to simplify the objective function.

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

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{z}} \quad & \text{tr}(\mathbf{w}\mathbf{w}^\top \mathbf{P}) - 2\mathbf{w}^\top \mathbf{q} \\ \text{s.t.} \quad & 0 \leq \mathbf{w} \leq \mathcal{M} \cdot \mathbf{z}, \\ & \sum_{k=1}^d z_k \leq \tau, \\ & z_k \in \{0, 1\}, \quad \forall k \in [d]. \end{aligned} \quad (12)$$

Thus, the resulting global support set is given by

$$\Omega := \{k \in [d] \mid w_k > 0\} = \text{supp}(\mathbf{w}) = \text{supp}(\mathbf{z}). \quad (13)$$

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 Γ 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 Ω in Eq. (13), the optimization problem now becomes

$$\begin{aligned} \min_{\mathbf{w}_{m,n,\gamma}} \quad & \|\tilde{\mathbf{x}}_{m,n,\gamma} - \mathbf{A}_{m,n,\gamma} \mathbf{w}_{m,n,\gamma}\|_2^2 \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{w}_{m,n,\gamma}) \geq 0, \quad \mathcal{P}_\Omega^\perp(\mathbf{w}_{m,n,\gamma}) = 0, \end{aligned} \quad (14)$$

for all $m \in [M], n \in [N], \gamma \in [\Gamma]$. Here, $\mathcal{P}_\Omega(\cdot)$ denotes the orthogonal projection supported on Ω . 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 Ω . 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 evaluate the proposed sparse autoregression (SAR) models through a series of experiments on real-world time series datasets. We consider two distinct domains: urban mobility and climate science. First, we analyze a temporally-varying dataset of ridesharing trips in NYC, which exhibits rich daily and weekly periodicity. This dataset spans multiple years (2019–2023), enabling us to detect how temporal patterns evolve across pre-pandemic, pandemic, and post-pandemic periods. Second, we study large-scale climate datasets, including North American atmospheric variables and global sea surface temperatures. These datasets are both spatially and temporally structured, allowing us to reveal the geographical distribution of seasonal effects and track shifts in climate patterns over time.

7.1 Human Mobility Periodicity

This case study focuses on identifying and tracking periodic patterns in human mobility using ridesharing data in NYC from 2019 to 2023. Due to daily and weekly commuting cycles and disruptions such as COVID-19, the underlying temporal structure is both periodic and non-stationary. We demonstrate how the proposed TV-SAR model captures this evolving structure and compare alternative solution approaches for model estimation.

7.1.1 Dataset and Motivation for TV-SAR

We apply the proposed TV-SAR model to NYC ridesharing data spanning February 2019 to December 2023, focusing on mobility patterns at John F. Kennedy (JFK) International Airport. Airport-related trips are known to follow strong daily and weekly rhythms due to fixed flight schedules.

Figures 4(a) and 4(b) visualize the weekly evolution of pickup and dropoff activities, with each heatmap row corresponding to one week (approximately 260 weeks in total). Pickup trips show higher variability due to factors such as flight delays, baggage collection, and access constraints to pickup zones; demand peaks in the evening hours. In contrast, dropoff trips exhibit clearer regularity, with consistent morning and afternoon peaks aligned with flight departures. Aggregated time series in Figures 4(c) and 4(d) further confirm that dropoff trips display stronger daily and weekly periodicity compared to pickup trips.

Furthermore, as shown in Figure 4, the ridesharing pickup and dropoff time series display long-term temporal variability. Notably, there is a sharp decline in trip volumes in 2020 due to the COVID-19 pandemic, followed by a gradual recovery through the end of 2021. These observations suggest that the periodic structure of the ridesharing time series is not only rich but also subject to external disruptions. This motivates the use of a temporally-varying model such as TV-SAR to flexibly capture changes in autoregressive patterns across months and years.

7.1.2 Model Training: Comparison of Estimation Algorithms

We compare three algorithms for solving Problem (4) on the ridesharing pickup time series over five years: NNSP, MIO-DVP, and MIO. All methods are tested at two sparsity levels, $\tau = 4$ and $\tau = 6$. For MIO-DVP, we additionally conduct experiments with different pruning thresholds $\tau_0 = 5$ and $\tau_0 = 10$, and examine how the preselected support set size influences solution quality.

Table 2 reports the objective function values under each setting. As expected, the MIO algorithm consistently achieves the best performance, finding globally optimal solutions. NNSP serves as a baseline and exhibits inferior solution quality, particularly at higher sparsity levels. The performance of MIO-DVP lies in between: with an effective pruning threshold (e.g., $\tau_0 = 10$), it nearly matches the performance of full MIO while incurring significantly lower computational cost.

This highlights an important tradeoff: although MIO guarantees the best solution, MIO-DVP offers a scalable alternative with much faster runtimes due to its reduced search space. The pruning parameter τ_0 plays a critical role—aggressive pruning can degrade performance,

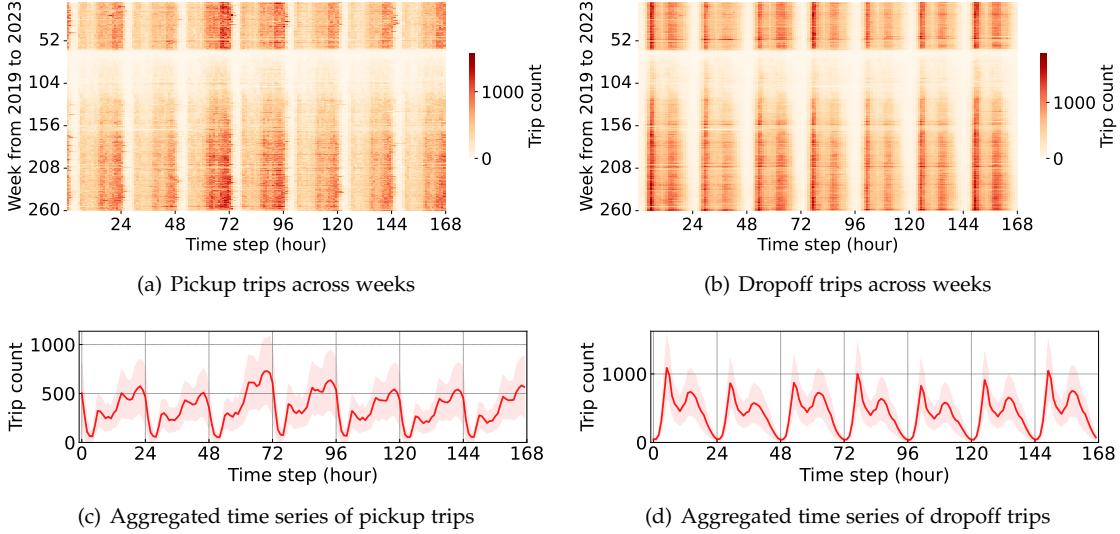


Fig. 4: Hourly time series of the aggregated ridesharing trip counts at JFK International Airport from 2019 to 2023. (a-b) The row of each heatmap refers to the ridesharing trip time series of each week. (c-d) The time series refers to the average ridesharing trips of each hour within a week window, while the standard deviations are also presented.

TABLE 2: Objective function $f(\mathbf{w})$ in Eq. (4) on the ridesharing pickup trip time series at JFK. 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).

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

whereas moderate values (e.g., $\tau_0 = 10$) enable a favorable balance between efficiency and accuracy.

7.1.3 Results: Periodicity of Ridesharing Trips

To capture the temporally-varying dynamics present in the dataset, we apply the proposed TV-SAR model, treating each month as an independent time segment. In order to keep stability of sparse coefficients, we use the training data of each time segment as the incremental set from January to the current month of each year. The autoregressive order is set to $d = 168$, corresponding to one week of hourly lags. We use a sparsity level of $\tau = 4$ to allow the model to select dominant lags, particularly targeting daily ($k = 24$) and weekly ($k = 168$) periodicity.

Figure 5 shows that the model consistently recovers the support set $\text{supp}(\mathbf{w}_\gamma) = 1, 24, 167, 168$ for all $\gamma \in [2, 60]$, reflecting the presence of strong local, daily, and weekly auto-correlations. Figure 5(a) visualizes how the periodic structure of pickup trips evolves over time. In 2020, the strength of the weekly component decreases remarkably, while the daily component grows stronger. This pattern is similarly observed in the dropoff coefficients (Figure 5(b)), where we also note a sharp increase in the local coefficient at $k = 1$ during 2020—suggesting increased short-term variability in dropoff behavior.

Looking across all years, the weekly periodicity at $k = 168$ is generally more pronounced in dropoff trips than in pickup trips. This is consistent with the patterns seen in Figure 4, where dropoff activity follows more regular flight schedules, while pickup trips are affected by greater uncertainty (e.g., baggage claim, ridesharing wait times). Although periodicity in 2020 differs substantially from the adjacent years, the autoregressive patterns in 2019, 2021, 2022, and 2023 show no significant differences. This indicates that by 2021–2023, human mobility behavior had largely returned to its pre-pandemic structure.

7.2 Seasonality Patterns of North America Climate Variables

This case study investigates large-scale seasonal patterns in climate variables across North America using long-term temperature and precipitation data (1980–2019). These data exhibit strong yearly seasonality influenced by geography and climate change, and are structured both spatially and temporally. Our goal is to quantify the strength and variability of these seasonal patterns over space and time. We use the STV-SAR model to uncover the geographical organization of seasonality, track its evolution over decades, and assess the robustness of the learned patterns across spatial resolutions and segmentation schemes.

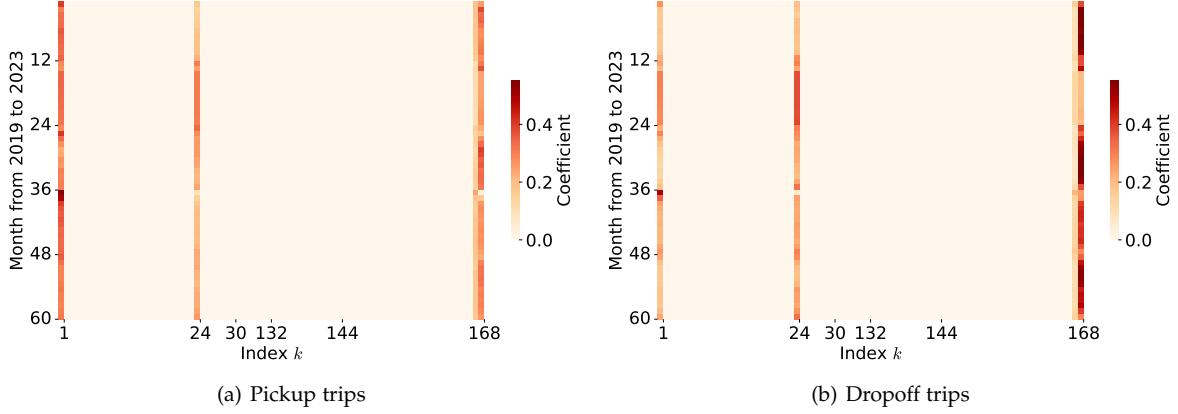


Fig. 5: Sparse coefficient vectors $w_\gamma \in \mathbb{R}^{168}$, $\gamma \in [2, 60]$ of TV-SAR on the ridesharing trip time series from February 2019 to December 2023, i.e., 59 months in total. Since each time series data corresponds to one month, there are 59 coefficient vectors that are represented 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.

7.2.1 Dataset and Motivation for STV-SAR

We study the Daymet dataset, which provides monthly climate variables—including minimum/maximum temperature and precipitation—at a high spatial resolution of $1\text{km} \times 1\text{km}$ across North America [35].¹ Our analysis focuses on data from 1980 to 2019, resulting in long, spatially-distributed multidimensional time series. Figure 6 illustrates the minimum temperature data for January over the past decade. These panels reveal strong seasonal structure and regional climate variability, suggesting the need for models that can capture recurring periodic patterns as well as their changes across space and time.

Given the spatial scale and long-term coverage of this dataset, we aim to (i) quantify yearly seasonality and its regional variation, and (ii) examine how such patterns evolve across decades and spatial resolutions. These goals motivate the use of our proposed STV-SAR model, which is well-suited for discovering global periodic structures in large-scale, multidimensional time series.

7.2.2 Global Support Verification

To enable scalable model training on high-resolution spatiotemporal data, we aggregate the original $1\text{km} \times 1\text{km}$ dataset to a coarser resolution of $10\text{km} \times 10\text{km}$. Even at this level, the STV-SAR model of Eq. (9) must be fit across over 800,000 time series. To manage this computational burden, we employ the two-stage optimization scheme outlined in Sections 6.2 and 6.3. As shown in Table 3, our scheme trains the full model across all time series in under 12 seconds—demonstrating its practical scalability for large-scale climate datasets.

We apply STV-SAR with a prescribed sparsity level $\tau = 3$, aiming to capture dominant temporal dependencies that characterize yearly seasonality. For example, using the minimum temperature dataset, we construct monthly time series at each spatial cell over decadal segments (from 1980s to 2010s). Solving the global model yields a common

support set $\Omega = \{1, 11, 12\}$, which includes recent memory ($k = 1$), near-annual lags ($k = 11, 12$), and aligns with expected seasonal dynamics.

To assess the robustness of this global support set, we also fit independent SAR models to 204,153 time series at $10\text{km} \times 10\text{km}$ resolution using sparsity levels $\tau \in \{3, 4, 5\}$. As shown in Figure 7, even without enforcing any global constraint, the most frequently selected lags coincide with the global support set Ω . For $\tau = 3$, the dominant indices are $\{1, 11, 12\}$; for higher sparsity levels, the support expands to include nearby lags (e.g., $\{2, 10\}$), but $\{1, 11, 12\}$ remain consistently prominent. This confirms that the global support set not only enhances computational efficiency but also meaningfully captures shared temporal structures across space.

7.2.3 Results: Seasonality Patterns across Four Decades

Understanding how seasonality patterns evolve over time allows us to quantify the dynamics of climate systems. Figure 8 visualizes the yearly seasonality of monthly climate time series across North America from 1980 to 2019, using a spatial resolution of $10\text{km} \times 10\text{km}$. Higher coefficients correspond to regions with stronger yearly periodicity.

We begin by analyzing the seasonality patterns of minimum/maximum temperature. Across the four decades, areas with stronger yearly seasonality are consistently concentrated in high-latitude regions, such as Canada and Alaska. In contrast, southern regions like Mexico exhibit weaker and more spatially variable seasonality. A notable expansion of highly seasonal regions occurs in the 2000s, suggesting an intensification of temperature-driven seasonal cycles. In the 2010s, two regions stand out: northern Canada (e.g., Nunavut) and the western United States. While the patterns in the 1980s and 1990s are relatively stable, the 2000s and 2010s show substantial shifts—highlighting evolving climate dynamics with potential ecological implications.

Next, we compare seasonal structures across different climate variables. The seasonality of precipitation exhibits significant spatial heterogeneity throughout the decades,

¹ <https://daac.ornl.gov/DAYMET>

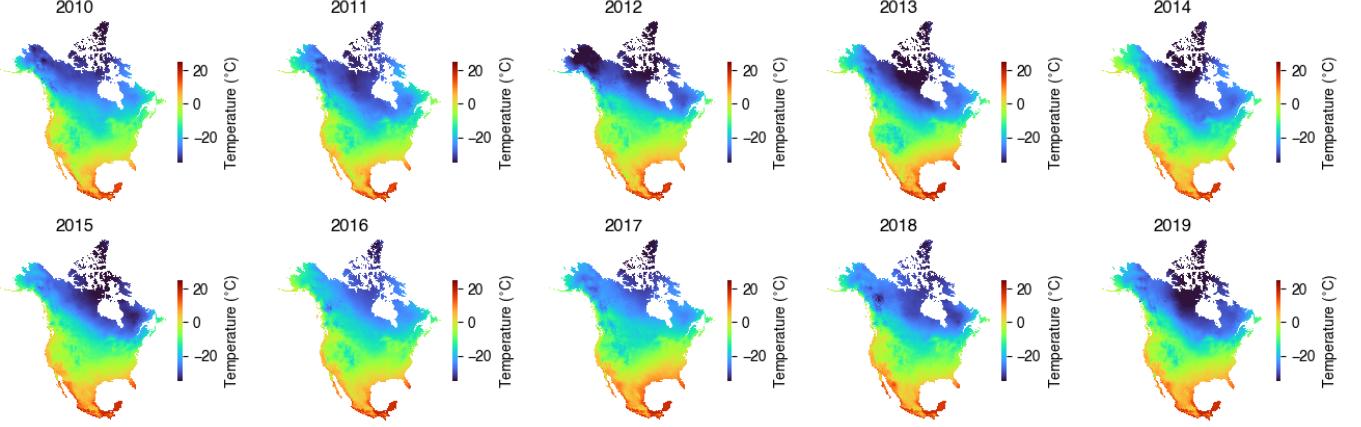


Fig. 6: Monthly aggregated minimum temperature across North America of January from 2010 to 2019. The color scale represents the temperature in degrees Celsius, with blue tones indicating lower temperatures and red tones indicating higher temperatures.

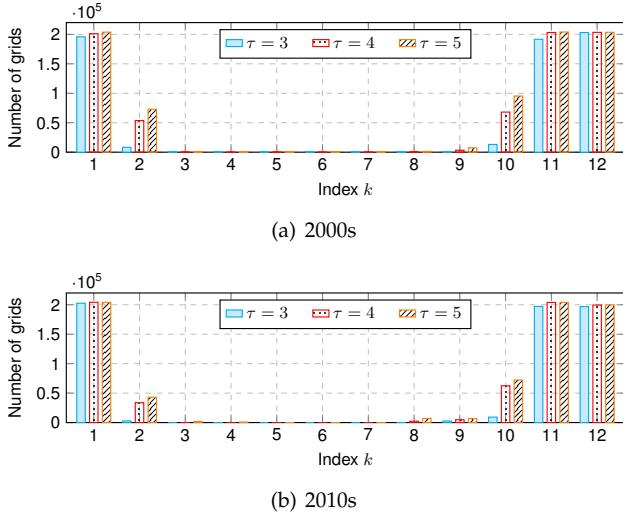


Fig. 7: Support sets of independent SAR models with different sparsity levels tested on the minimum temperature data in 2000s and 2010s across 204,153 grids.

even across small neighboring regions. Meanwhile, the seasonality of minimum and maximum temperatures remains relatively stable in the 1980s and 1990s. However, in the 2000s, minimum temperature seasonality intensifies and expands more broadly than that of maximum temperature. In the 2010s, Mexico shows higher seasonality in minimum temperature compared to maximum temperature, whereas the reverse holds in the western United States and central North America. Notably, in northern Canada, maximum temperature is less seasonal than minimum temperature. These contrasting patterns suggest differing drivers and sensitivities of seasonal cycles across variables and geographies.

7.2.4 Sensitivity Analysis: Spatial Resolution

We now investigate the effect of spatial aggregation by comparing results at three different resolutions: $5\text{ km} \times 5\text{ km}$, $10\text{ km} \times 10\text{ km}$, and $20\text{ km} \times 20\text{ km}$. Figure 9 illustrates that

the spatial distribution of yearly seasonality remains consistent across these resolutions. In particular, the northern regions of Canada, the western United States, and the southeastern United States exhibit strong seasonal signals, while central North America displays substantially weaker seasonality. Temperature seasonality in Mexico shows high spatial variability and inconsistency among nearby areas, as highlighted in Figure 10. Fine-grained resolution (e.g., 5 km) reveals localized structure that may be smoothed out at coarser scales. This demonstrates that our STV-SAR model captures robust patterns across resolutions while still preserving high-resolution detail where available.

To address the computational challenges of large-scale training, we employ a two-stage optimization scheme that efficiently solves the STV-SAR problem across millions of time series. Table 3 reports runtime statistics at each resolution. Remarkably, the model computes over 3 million time series at $5\text{ km} \times 5\text{ km}$ resolution within one minute, underscoring the scalability of the proposed method for real-world geospatial applications.

TABLE 3: Runtime of STV-SAR on the North America (minimum) temperature dataset across four decades with different spatial resolutions. Note that ν denotes the number of time series.

	$5\text{ km} \times 5\text{ km}$	$10\text{ km} \times 10\text{ km}$	$20\text{ km} \times 20\text{ km}$
Number ν	3,343,628	816,612	196,720
Runtime (s)	48.55	11.77	2.88

7.2.5 Sensitivity Analysis: Phase Segmentation

In our main analysis, we segmented the climate time series into decades to capture long-term trends. We now evaluate the sensitivity of our findings to the choice of segmentation granularity by re-estimating seasonality patterns using shorter, 5-year phases on the minimum temperature data from 2000 to 2019. Figure 11(a) illustrates the spatial distribution of yearly seasonality from 2000 to 2004, which closely resembles the 1990s pattern shown in Figure 8(a). In the 2005–2009 window (Figure 11(b)), seasonality intensifies

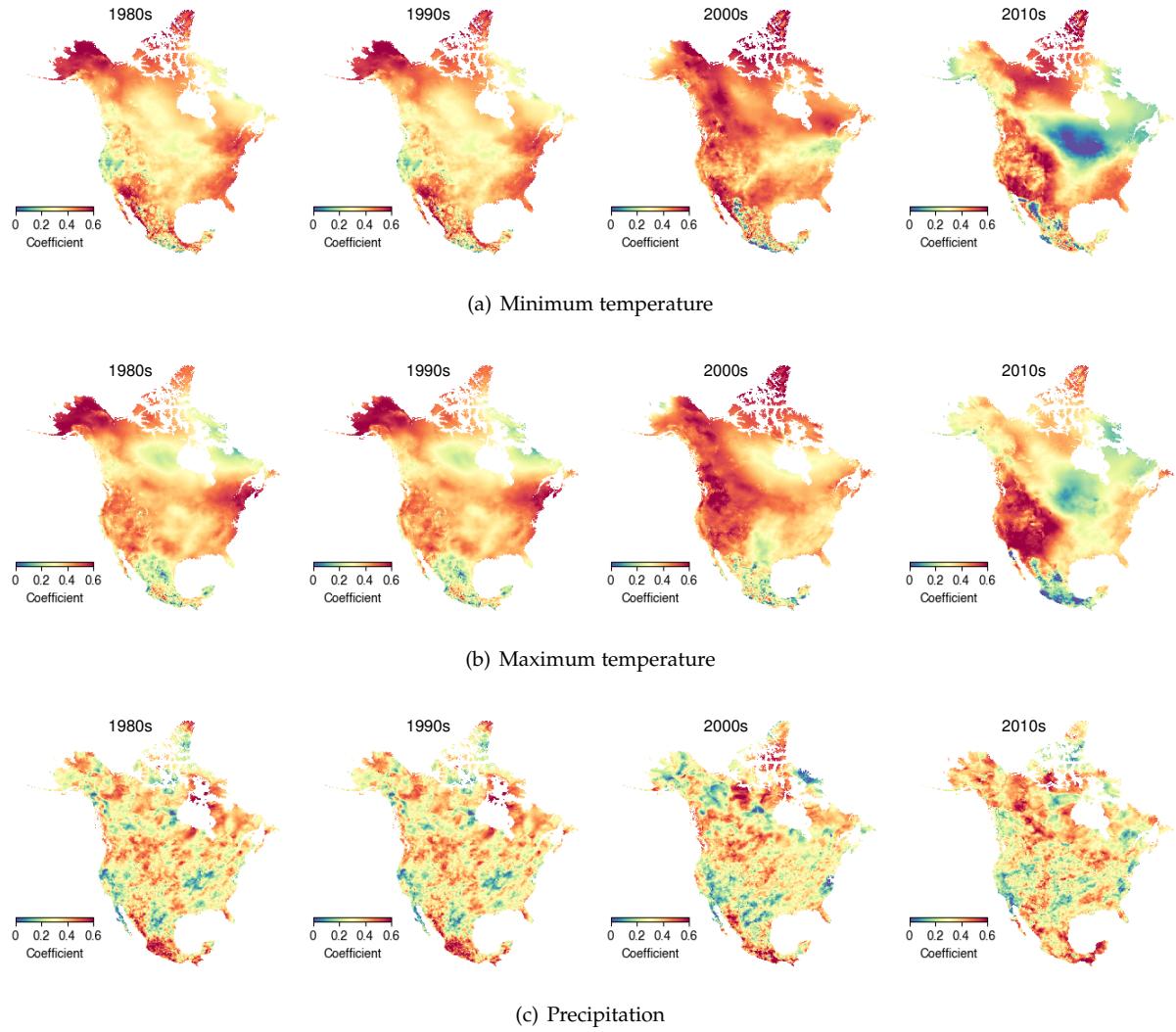


Fig. 8: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the North America climate data across the past four decades.

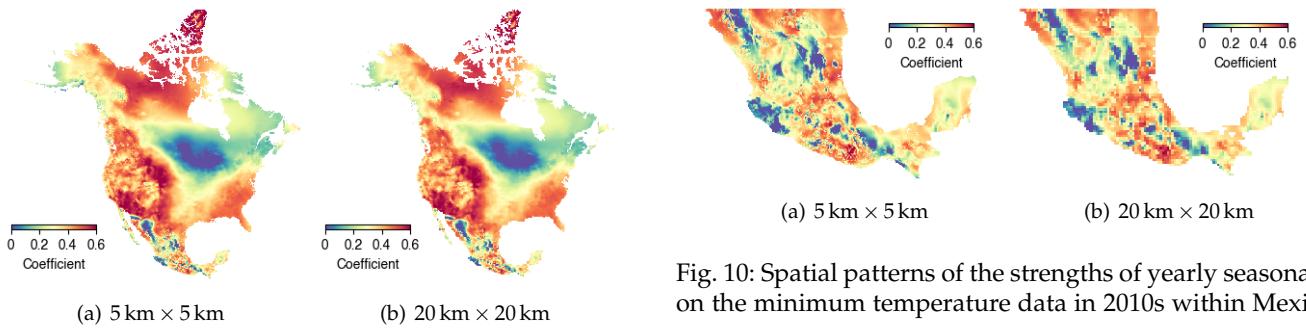


Fig. 9: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the minimum temperature data in 2010s.

in several regions, aligning with the dominant pattern of the 2000s. From 2010 to 2014 (Figure 11(c)), central North America shows noticeably weaker seasonal signals, whereas the period from 2015 to 2019 (Figure 11(d)) marks a rebound

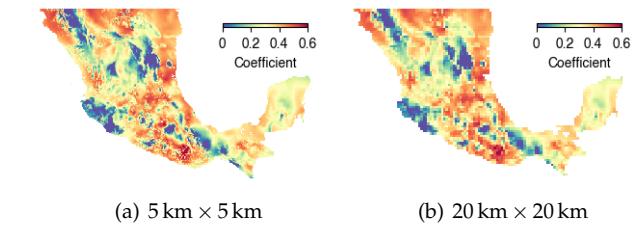


Fig. 10: Spatial patterns of the strengths of yearly seasonality on the minimum temperature data in 2010s within Mexico.

in seasonality strength, particularly in Mexico. These subtle temporal shifts are less evident in the decade-level summary (Figure 8), highlighting the value of finer segmentation for detecting short-term variations in seasonality patterns.

7.3 Seasonality Patterns of Sea Surface Temperature

This case study focuses on identifying and tracking yearly periodic patterns in global sea surface temperatures from

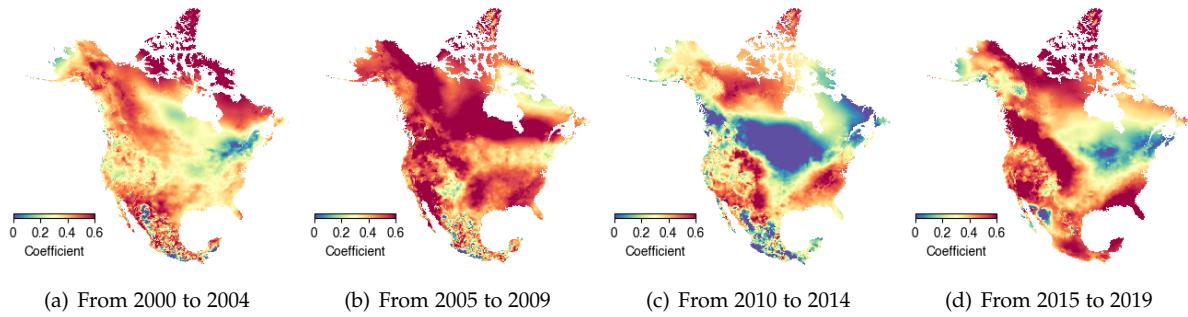


Fig. 11: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the minimum temperature data with a 5-year phase (i.e., $T_\gamma = 60$ months).

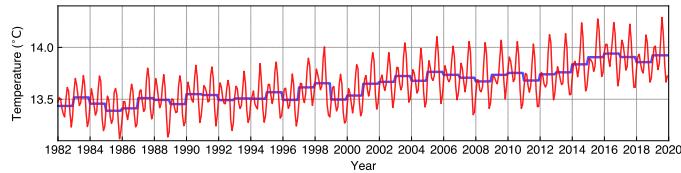


Fig. 12: Average values of the monthly sea surface temperature data from January 1982 to December 2019. The average temperature values in 1980s, 1990s, 2000s, and 2010s are 13.46°C , 13.54°C , 13.69°C , and 13.83°C , respectively. The blue and red curves correspond to yearly and monthly average temperature values, respectively.

1982 to 2019. Sea surface temperature is a key indicator of climate variability, with dynamics influenced by geography, ocean circulation, and anthropogenic warming. The data exhibit both strong seasonality and long-term trends, making them ideal for evaluating the ability of STV-SAR to capture structured temporal variation at scale. We use the model to uncover geographic patterns of seasonal strength and examine how these patterns evolve across decades and spatial regions.

7.3.1 Dataset and Motivation for STV-SAR

We study the NOAA Optimum Interpolation Sea Surface Temperature dataset, which provides monthly average temperatures over a $0.25^\circ \times 0.25^\circ$ global grid, comprising 720×1440 cells (i.e., 1,036,800 spatial locations).² The dataset spans four decades, from January 1982 to December 2019. Figure 12 presents both monthly and yearly average sea surface temperatures across this period. The monthly time series reveals strong seasonal fluctuations, while the yearly average trend shows a gradual warming signal consistent with global climate change. These dynamics highlight the importance of quantifying yearly seasonality—not only to understand recurring oceanic patterns but also to characterize the variability and potential instability in the climate system. In particular, stronger seasonality implies more predictable, cyclic dynamics, whereas weaker seasonality can indicate structural shifts or increased variability.

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

7.3.2 Results: Spatial and Temporal Structure of Sea Surface Seasonality

To analyze the temporal structure of sea surface temperatures, we fit the STV-SAR model separately on each decade of data. We fix the model order at $d = 12$ and sparsity level $\tau = 3$, which allows us to isolate and interpret coefficients corresponding to yearly periodicity. Figure 13 shows the spatial distribution of the coefficient corresponding to lag $k = 12$, which serves as a proxy for yearly seasonality strength. Higher values (in red) indicate regions with stronger temperature seasonality—most notably near Eastern and Central Asia, North America, Europe, and North Africa. In contrast, equatorial regions such as the Pacific Ocean around the El Niño zone exhibit lower coefficients, reflecting weaker seasonality and more irregular variation.

Despite the broad stability of global seasonal patterns across decades, we observe subtle temporal changes. For example, as shown in Figure 14, the regions near Canada and the Arctic Ocean exhibit decreasing seasonality over time. This decline may signal evolving oceanic dynamics due to changes in ocean circulation, warming trends, or atmospheric variability. The Arctic Ocean, in particular, appears to have become notably less seasonal over the past two decades, potentially reflecting broader climate-induced disruptions in polar ocean systems.

8 CONCLUSION

This work introduces a unified SAR framework for quantifying periodicity in real-world spatiotemporal time series. We develop interpretable extensions—TV-SAR and STV-SAR—that incorporate temporally and spatially-varying sparsity constraints to reveal structured periodic components. To address the computational challenges posed by the underlying MIO problems, we propose scalable algorithmic solutions: a DVP-based method for TV-SAR and a two-stage optimization scheme for STV-SAR. Applied to both ridesharing mobility data and climate time series, our models successfully uncover meaningful, interpretable seasonal patterns and periodic structures. The learned spatial and temporal dynamics illustrate the practical relevance of periodicity quantification for analyzing complex systems.

We point out the following directions for future improvement. The global support estimation strategy used in STV-SAR is particularly effective when temporal structures

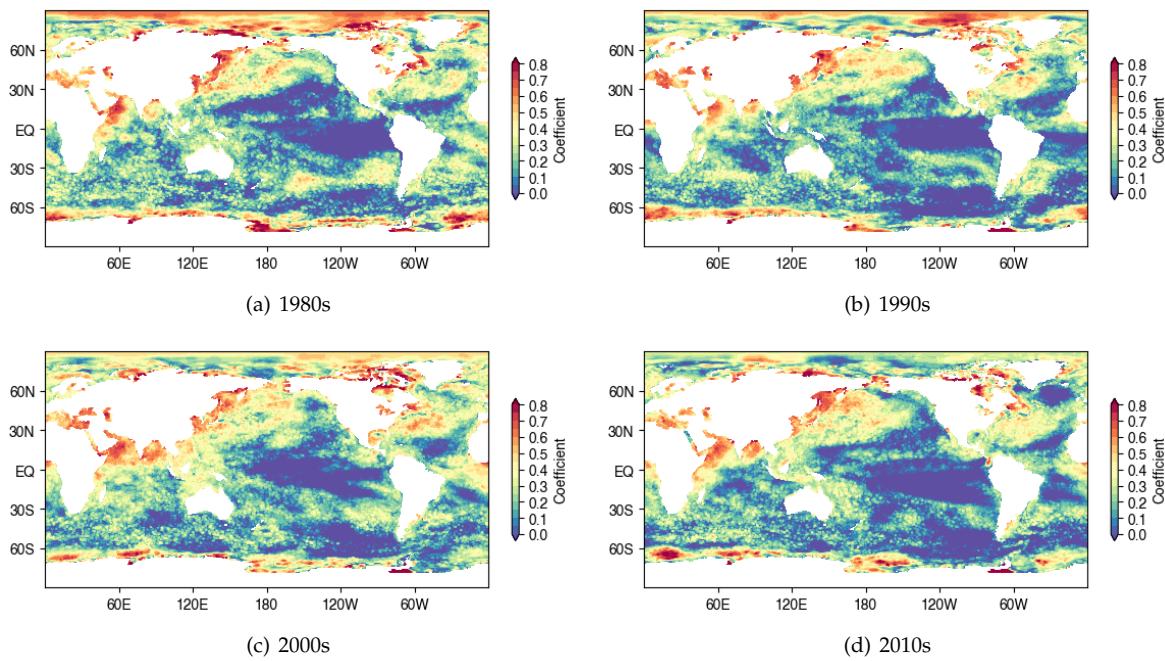


Fig. 13: Spatial patterns of the strengths of yearly seasonality quantified by STV-SAR on the sea surface temperature data in the past four decades.

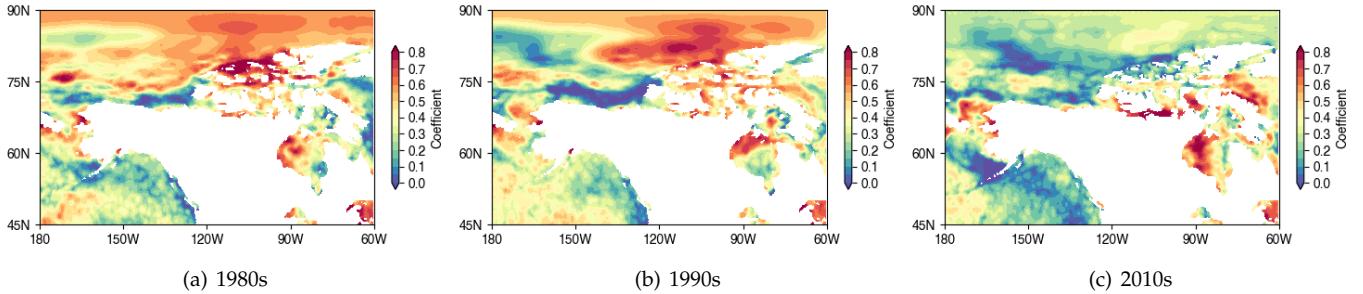


Fig. 14: Spatial patterns of the strengths of yearly seasonality on the sea surface temperature data with highlighted areas around Canada.

are shared across time series, as in climate datasets (e.g., Figure 7). However, for heterogeneous datasets with diverse or complex dynamics, the approximation introduced by the two-stage optimization may lead to estimation bias. Addressing this gap—potentially by integrating adaptive or hierarchical modeling techniques—could further enhance the robustness and generality of the framework.

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.
 - [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, "Okridge: Scalable optimal k-sparse ridge regression," *Advances in neural information processing systems*, vol. 36, 2024.
- [22] E. Carriosa, A. V. Olivares-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 applications 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] M. Fujiwara, "Über die obere grenzen des absoluten wertes des wurzels an algebraic equation," *Tohoku Mathematical Journal, First Series*, vol. 10, pp. 167–171, 1916.
- [32] D. Bertsimas and V. Digalakis Jr, "The backbone method for ultra-high 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



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.



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.