



UNIVERSITY OF  
CENTRAL FLORIDA

# Machine Learning and Optimization for Understanding Spatiotemporal Systems

Time Series Imputation & Periodicity Quantification

**Xinyu Chen**

Postdoctoral Associate, MIT

May 22, 2025

Orlando, USA

# Spatiotemporal Data

- Transport & mobility application scenarios



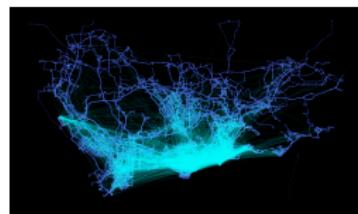
Highway (Portland)



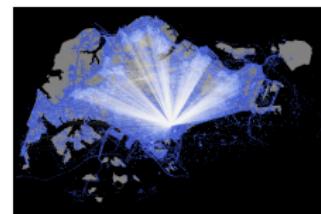
Uber movement (NYC)



Uber movement (Seattle)



Taxi trajectory (Shenzhen)



Human movement (Singapore)

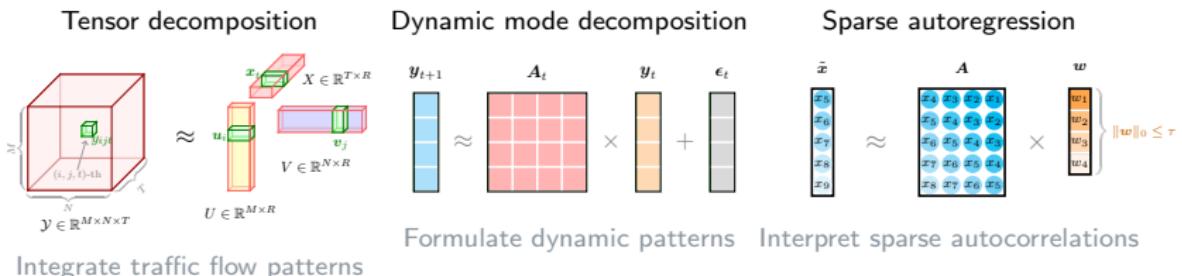
- Challenges: Sparsity, high-dimensionality (network-scale), and multi-dimensionality (complicated data structure), time-varying systems

# Research Contributions

- (Practical) Formulating challenging spatiotemporal problems



- (Methodological) Advancing ML development



# Reproducible Research

- The last mile of AI for spatiotemporal data computing

Human mobility & smart cities  
Data-driven transport analytics  
Spatiotemporal data modeling  
Tensor decomposition for ML  
Optimization for interpretable ML

...

Directions & Topics



- Advancing ML development with open-source research

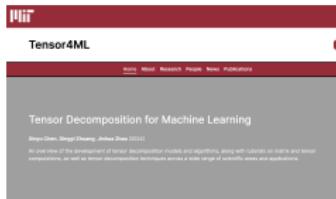


**transdim**

(1,200+ GitHub stars)

ML for Transport Data Imputation

<https://github.com/xinychen/transdim>



**Tensor Decomposition for ML**

(ML project)

Math & ML Tutorials

<https://sites.mit.edu/tensor4ml>



**Spatiotemporal Data Computing**

(Data valorization project)

Model Development of ML & Data Science

<https://spatiotemporal-data.github.io>

# Spatiotemporal Data Imputation

- Convolution     Fast Fourier transform     Optimization w/  $\ell_1$ -norm
- Time series imputation     Speed field reconstruction



Xinyu Chen  
UdeM → MIT



Zhanhong Cheng  
McGill → UF



HanQin Cai  
UCF



Nicolas Saunier  
PolyMtl



Lijun Sun  
McGill

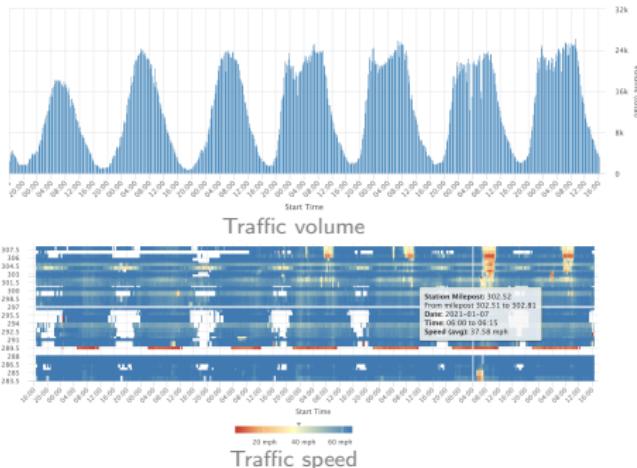
- **Xinyu Chen**, Zhanhong Chen, HanQin Cai, Nicolas Saunier, Lijun Sun (2024). “Laplacian Convolutional Representation for Traffic Time Series Imputation”. *IEEE Transactions on Knowledge and Data Engineering*, 36 (11): 6490–6502.
- Blog post: Understanding time series convolution.  
[https://spatiotemporal-data.github.io/posts/ts\\_conv](https://spatiotemporal-data.github.io/posts/ts_conv)

# Motivation

- Portland highway traffic data<sup>1</sup>



Highway network & sensor locations

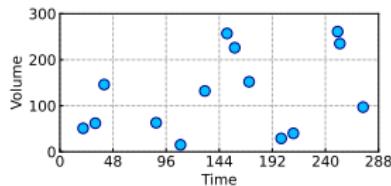


- $\mathbf{X} \in \mathbb{R}^{N \times T}$  with  $N$  spatial locations  $\times T$  time steps
- Traffic volume/speed shows strong spatial/temporal dependencies
- Missing data are there, how to improve data quality?

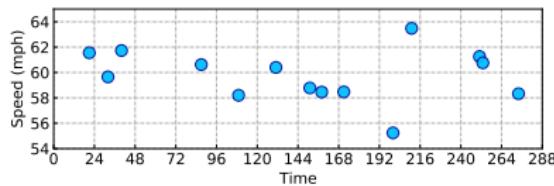
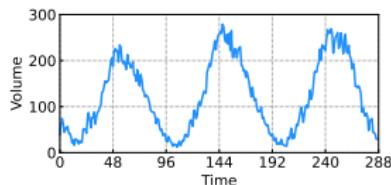
---

<sup>1</sup><https://portal.its.pdx.edu/home>

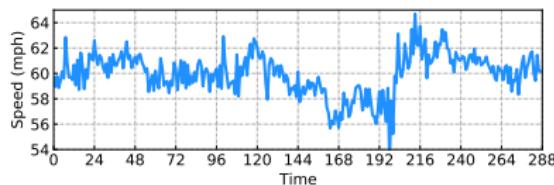
# Motivation



↓  
Reconstruct  
traffic volume?

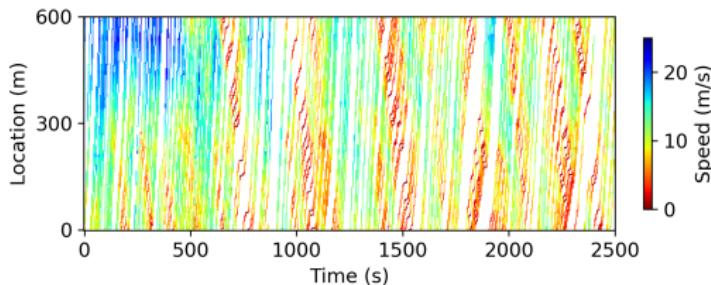


↓  
Reconstruct  
traffic speed?



- How to utilize the global trends of traffic time series?
- How to produce local consistency of traffic data?

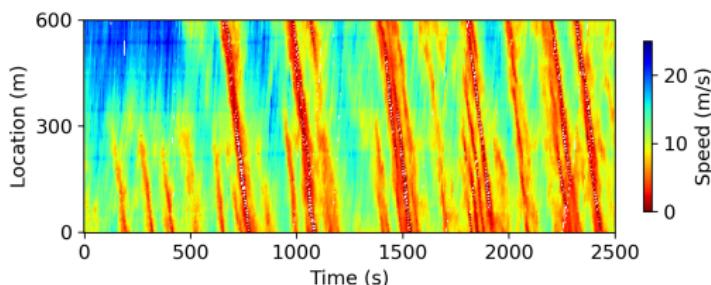
# Motivation



200-by-500 matrix  
(NGSIM)



Reconstruct speed field from  
20% sparse trajectories?

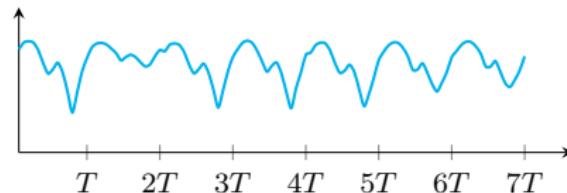


- How to learn from sparse spatiotemporal data?
- How to characterize spatial/temporal local dependencies?

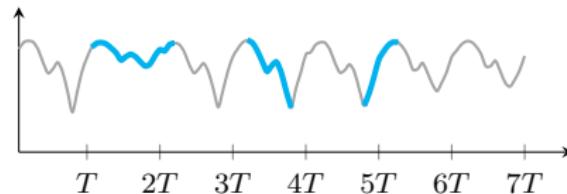
# Time Series Imputation

Global/local trends in sparse data?

- Global trends (e.g., daily/weekly periodicity):



- Local trends (e.g., short-term time series trends):



# Local Trend Modeling

- Intuition of Laplacian matrix

Undirected and circulant graph

Modeling  $\longrightarrow$

$$\mathbf{L} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

- Laplacian kernel:

$$\boldsymbol{\ell} \triangleq \underbrace{(2, -1, 0, 0, -1)}_{\text{first column of } \mathbf{L}}^\top$$

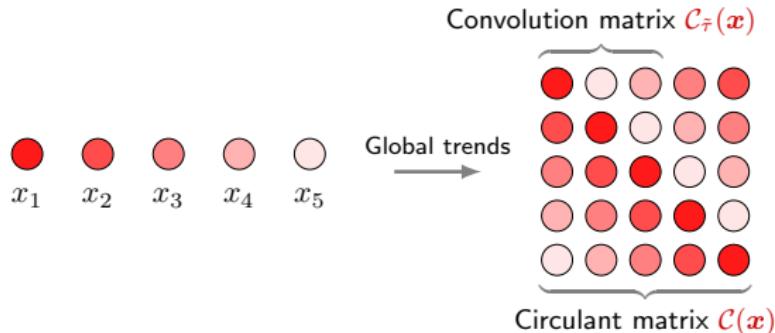
extending to the degree  $2\tau$  (i.e., graph connectivity) for  $\mathbf{x} \in \mathbb{R}^T$ .

- Temporal regularization:

$$\mathcal{R}(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{L}\mathbf{x}\|_2^2}_{\text{mat-vec mul.}} = \underbrace{\frac{1}{2} \|\boldsymbol{\ell} * \mathbf{x}\|_2^2}_{\text{convolution}*}$$

# Global Trend Modeling

Circulant matrix  $\mathcal{C}(\mathbf{x})$  vs. convolution matrix  $\mathcal{C}_{\tilde{\tau}}(\mathbf{x})$



- Circulant/Convolution nuclear norm minimization
  - A balance between global and local trends modeling?

CircNNM (Liu'22, Liu & Zhang'23)

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathcal{C}(\mathbf{x})\|_* \\ \text{s.t. } & \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon \end{aligned}$$

on data  $\mathbf{y}$  w/ observed index set  $\Omega$ .

ConvNNM (Liu'22, Liu & Zhang'23)

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathcal{C}_{\tilde{\tau}}(\mathbf{x})\|_* \\ \text{s.t. } & \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon \end{aligned}$$

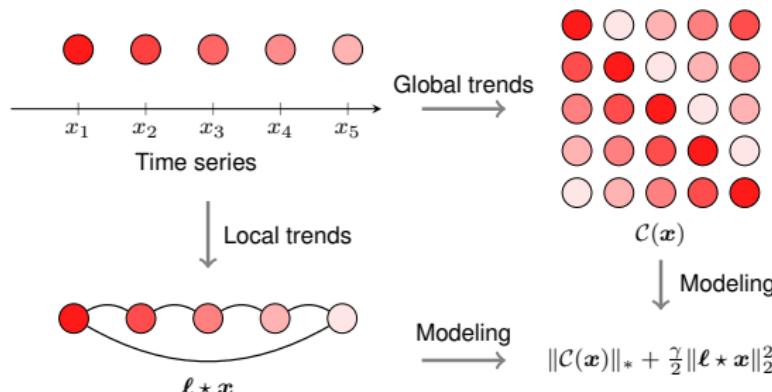
on data  $\mathbf{y}$  w/ observed index set  $\Omega$ .

# Global + Local Trends?

## Laplacian Convolutional Representation (LCR)

For any partially observed time series  $\mathbf{y} \in \mathbb{R}^T$  with observed index set  $\Omega$ , LCR utilizes **circulant matrix** and **Laplacian kernel** to characterize global/local time series trends:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \underbrace{\|\mathcal{C}(\mathbf{x})\|_*}_{\text{global}} + \frac{\gamma}{2} \underbrace{\|\ell * \mathbf{x}\|_2^2}_{\text{local}} \\ \text{s.t. } & \|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon \end{aligned}$$



## Laplacian Convolutional Representation

- LCR model:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2} \|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2 \\ \text{s.t. } & \|\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq \epsilon \end{aligned}$$

$$\implies \min_{\boldsymbol{x}} \underbrace{\|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2} \|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2}_{\text{global} + \text{local}} + \underbrace{\frac{\eta}{2} \|\mathcal{P}_\Omega(\boldsymbol{z} - \boldsymbol{y})\|_2^2}_{\text{regularization}}$$

s.t.  $\boldsymbol{z} = \boldsymbol{x}$

“The alternating direction method of multipliers (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle.”

— Source: <https://stanford.edu/~boyd/admm.html>

# Laplacian Convolutional Representation

- Augmented Lagrangian function:

$$\mathcal{L} = \underbrace{\|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2}\|\ell * \mathbf{x}\|_2^2}_{\text{global + local}} + \underbrace{\frac{\lambda}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \langle \mathbf{w}, \mathbf{x} - \mathbf{z} \rangle}_{\text{Lagrangian multiplier } \mathbf{w}} + \underbrace{\frac{\eta}{2}\|\mathcal{P}_\Omega(\mathbf{z} - \mathbf{y})\|_2^2}_{\text{observations } \mathbf{y}}$$

- Optimize  $\mathbf{x}$  w/ FFT in  $\mathcal{O}(T \log T)$  time:

$$\begin{cases} \|\mathcal{C}(\mathbf{x})\|_* = \|\mathcal{F}(\mathbf{x})\|_1 = \|\hat{\mathbf{x}}\|_1 & (\text{circulant matrix}) \\ \frac{1}{2}\|\ell * \mathbf{x}\|_2^2 = \frac{1}{2T}\|\mathcal{F}(\ell) \circ \mathcal{F}(\mathbf{x})\|_2^2 = \frac{1}{2T}\|\hat{\ell} \circ \hat{\mathbf{x}}\|_2^2 & (\text{circular convolution}) \end{cases}$$

- Reformulate the optimization as  $\ell_1$ -norm minimization:

$$\begin{aligned} \mathbf{x} &:= \arg \min_{\mathbf{x}} \|\mathcal{C}(\mathbf{x})\|_* + \frac{\gamma}{2}\|\ell * \mathbf{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{x} - \mathbf{z} + \mathbf{w}/\lambda\|_2^2 \\ \implies \hat{\mathbf{x}} &:= \arg \min_{\hat{\mathbf{x}}} \underbrace{\|\hat{\mathbf{x}}\|_1}_{\ell_1\text{-norm}} + \frac{\gamma}{2T}\|\hat{\ell} \circ \hat{\mathbf{x}}\|_2^2 + \frac{\lambda}{2T}\|\hat{\mathbf{x}} - \hat{\mathbf{z}} + \hat{\mathbf{w}}/\lambda\|_2^2 \end{aligned}$$

# Laplacian Convolutional Representation

$\ell_1$ -norm Minimization (Liu & Zhang'23)

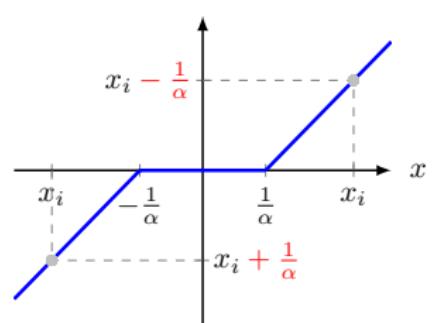
For any  $\hat{\mathbf{h}} \in \mathbb{C}^T$  and  $\delta \in \mathbb{R}$ :

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 + \frac{\delta}{2} \|\hat{\mathbf{x}} - \hat{\mathbf{h}}\|_2^2$$

The solution to  $\hat{\mathbf{x}}$ :

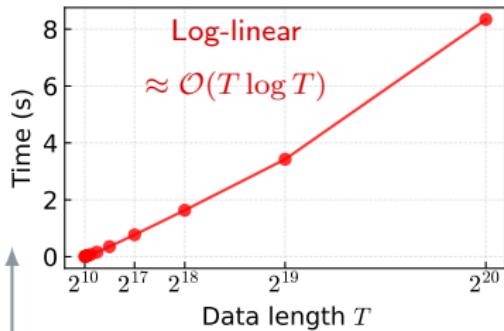
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \underbrace{\max\{0, |\hat{h}_t| - 1/\delta\}}_{\text{shrinkage (e.g., ReLU)}}, t \in [T]$$

$$y_i = \frac{x_i}{|x_i|} \cdot \max\{|x_i| - 1/\alpha, 0\}$$

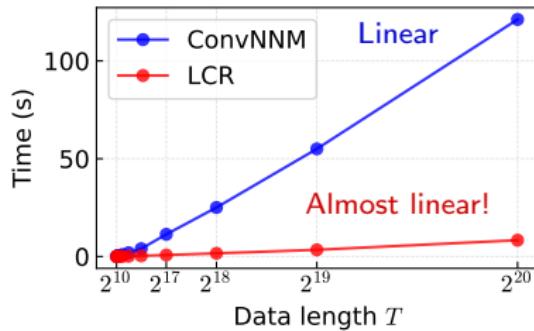


# Laplacian Convolutional Representation

Time complexity & scalability & efficiency?



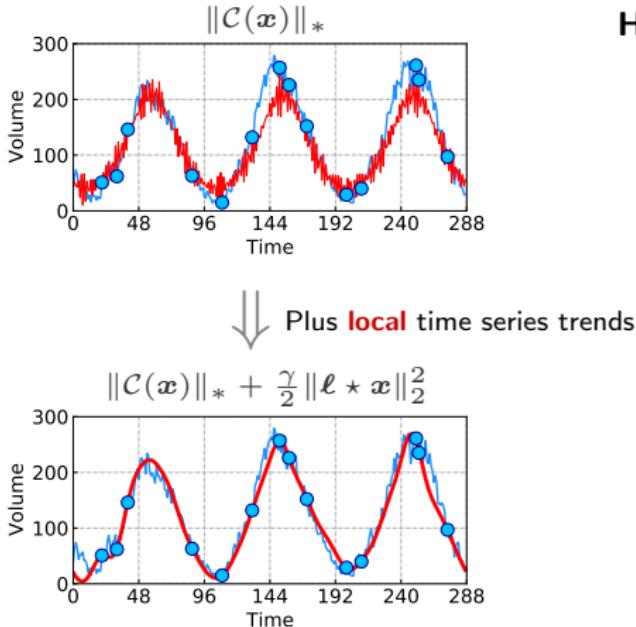
Empirical time complexity



On the synthetic data  $y \in \mathbb{R}^T$  with  
 $T \in \{2^{10}, 2^{11}, \dots, 2^{20}\}$

# Experiments

- Traffic speed imputation<sup>2</sup> (95% missing rate)



## Highlights:

- Rethink the importance of local trend modeling in traffic data imputation tasks.
- Find a unified global and local trend modeling framework whose optimization can be efficiently solved by **FFT**:

$$\min_{\mathbf{x}} \underbrace{\|\mathcal{C}(\mathbf{x})\|_*}_{\text{global}} + \frac{\gamma}{2} \underbrace{\|\ell * \mathbf{x}\|_2^2}_{\text{local}}$$

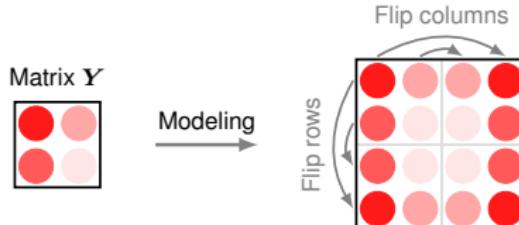
s. t.  $\|\mathcal{P}_\Omega(\mathbf{x} - \mathbf{y})\|_2 \leq \epsilon$

<sup>2</sup>Blue dot: partial observation; red line: imputation.

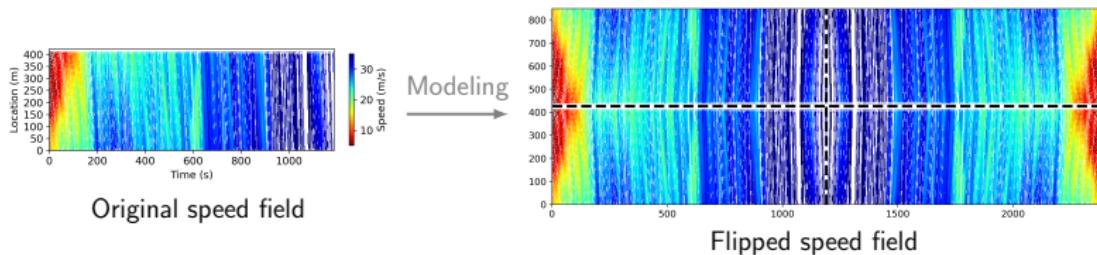
# Experiments

## Speed field reconstruction<sup>3</sup>

- Flipping operation on a matrix:



- Flipping operation on a speed field of vehicular traffic flow:



<sup>3</sup>Highway Drone (HighD) dataset at <https://www.hightd-dataset.com/>

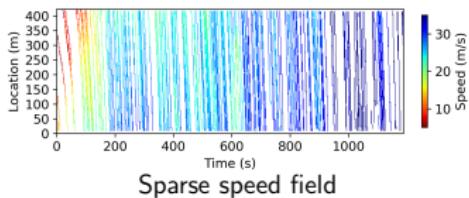
# Experiments

Speed field reconstruction in German highways<sup>4</sup>

- Scenario: Mask trajectories of 70% vehicles
- LCR-2D on partially observed  $\mathbf{Y} \in \mathbb{R}^{N \times T}$ :

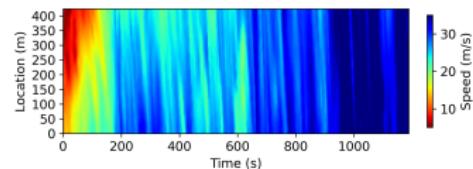
$$\min_{\mathbf{X}} \underbrace{\|\mathcal{C}(\mathbf{X})\|_*}_{\text{global trend}} + \frac{\gamma}{2} \underbrace{\|(\ell_s \ell^\top) \star \mathbf{X}\|_F^2}_{\text{local trend}}$$

s.t.  $\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \epsilon$

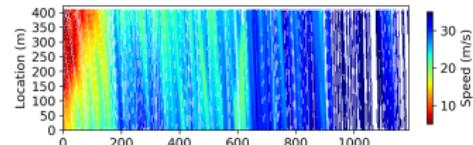


Sparse speed field

LCR-2D



Reconstructed speed field



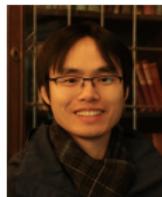
Ground-truth speed field

<sup>4</sup>Highway Drone (HighD) dataset at <https://www.hightd-dataset.com/>

# Quantifying Time Series Periodicity

(Ongoing Research)

- Interpretable ML     Optimization w/  $\ell_0$ -norm     Mixed-integer programming
- Human mobility regularity     Climate system seasonality



Xinyu Chen  
MIT



Dingyi Zhuang  
MIT



Yunhan Zheng  
MIT



Jinhua Zhao  
MIT



HanQin Cai  
UCF



Ryan Qi Wang  
Northeastern



Lijun Ding  
UCSD

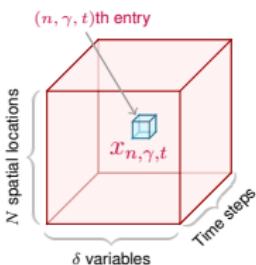


Vassilis Digalakis Jr  
HEC Paris

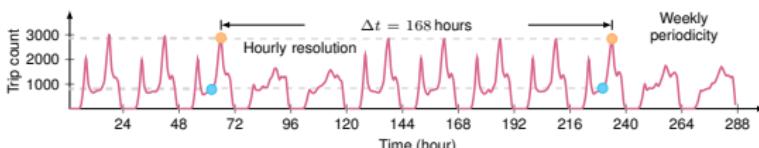
# Motivation

Human mobility data show daily/weekly regularity and periodicity?

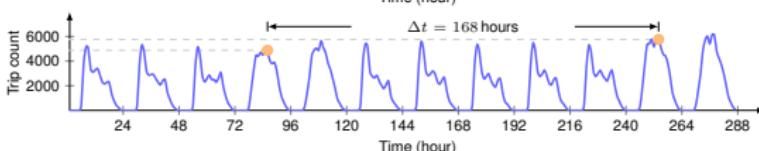
A



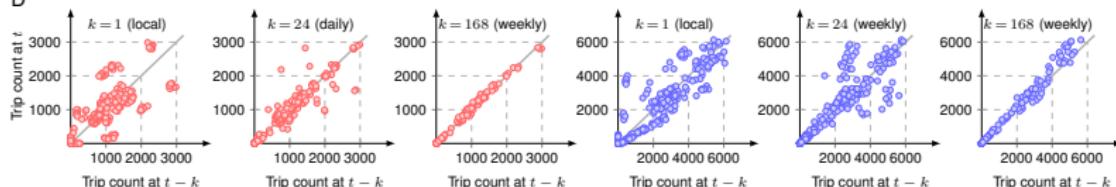
B



C



D

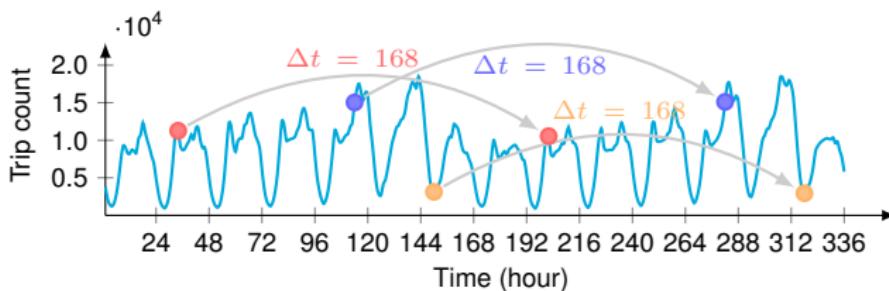


“Closeness” to the  
anti-diagonal  $y = x$

$x_t \approx x_{t-168}$  (weekly periodicity)

# Motivation

Weekly periodicity of ridesharing trip time series in Chicago



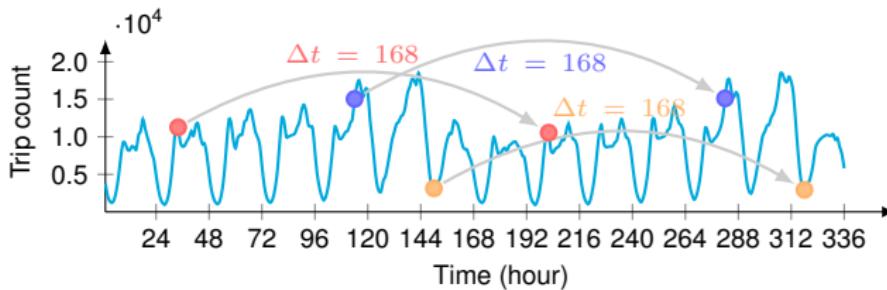
What motivate us most about periodicity?

- ① **Resilience and stability of systems:** Empirically measure the periodicity and predictability of urban systems.
- ② **Optimization of transport systems:** Optimize resources (e.g., public transit, taxi, rideshare, and micromobility) to meet transport demand efficiently.
- ③ **Design of sustainable transport & infrastructure:** Implement energy-efficient solutions tailored to peak hours.

## Motivation

- Time series autoregression on  $\mathbf{x} \in \mathbb{R}^T$

$$\mathbf{w} := \arg \min_{\{w_k\}_{k \in [d]}} \sum_{t \in [d+1, T]} \left( \mathbf{x}_t - \sum_{k \in [d]} w_k \mathbf{x}_{t-k} \right)^2$$



Periodicity of rideshare trip time series

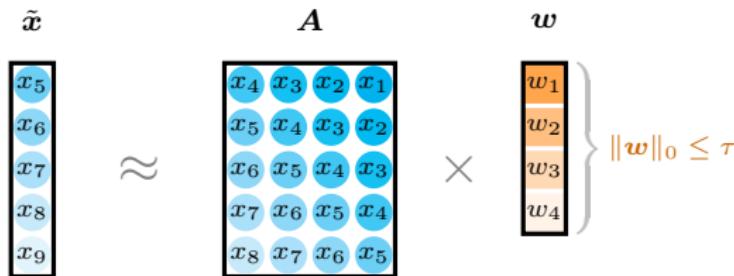
- Sparse coefficient vector  $\mapsto$  **Interpretability?**

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

# Valorizing Autoregression

- Time series autoregression

$$\begin{aligned} \mathbf{w} &:= \arg \min_{\{w_k\}_{k \in [d]}} \sum_{t \in [d+1, T]} \left( \mathbf{x}_t - \sum_{k \in [d]} w_k \mathbf{x}_{t-k} \right)^2 \\ &= \arg \min_{\mathbf{w}} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \end{aligned}$$



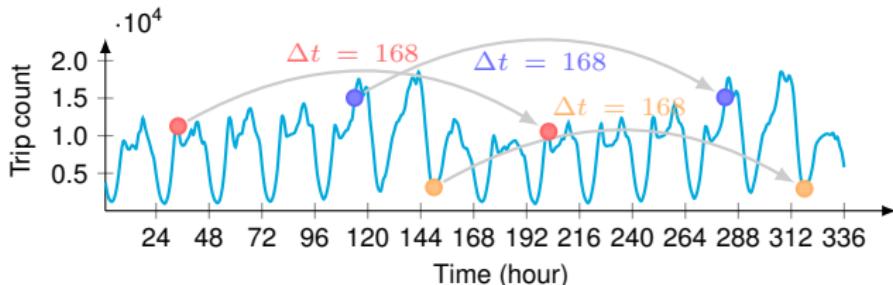
Autoregression on time series  $\mathbf{x} = (x_1, x_2, \dots, x_9)^\top$  w/ sparsity  $\tau \in \mathbb{Z}^+$

- Sparse autoregression

$$\begin{array}{ll} \min_{\mathbf{w} \geq 0} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 & \min_{\mathbf{w}, \boldsymbol{\beta}} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \\ \text{s.t. } \underbrace{\|\mathbf{w}\|_0 \leq \tau}_{\text{sparsity w/ } \ell_0\text{-norm}} & \iff \text{s.t. } \begin{cases} 0 \leq \mathbf{w} \leq \boldsymbol{\beta}, \boldsymbol{\beta} \in \{0, 1\}^d \\ \|\boldsymbol{\beta}\|_1 \leq \tau \end{cases} \end{array}$$

## Solution Quality

- Subspace pursuit (SP) sometimes fails

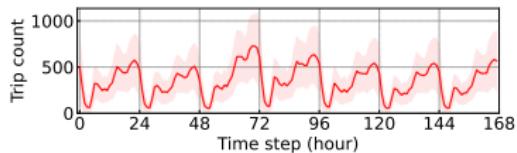
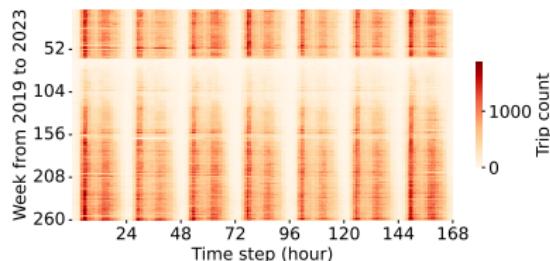
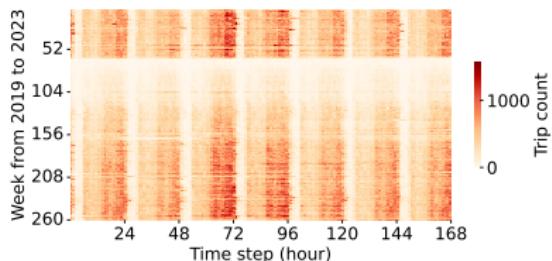


- Exact solution w/ mixed-integer programming (MIP)
- An intuitive example (sparsity  $\tau = 2$ ):

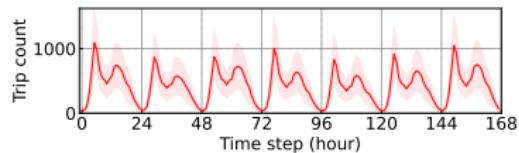
$$\underbrace{\boldsymbol{w} = (\dots, \underbrace{0.02}_{k=53}, \dots, \underbrace{0.96}_{k=168})^\top}_{\text{loss func. } = 8.32 \times 10^7 \text{ (SP)}} \quad \text{vs.} \quad \underbrace{\boldsymbol{w} = (\underbrace{0.22}_{k=1}, \dots, \underbrace{0.77}_{k=168})^\top}_{\text{loss func. } = 6.25 \times 10^7 \text{ (MIP)}}$$

# John F. Kennedy International Airport

- Pickup/Dropoff trips in airport
  - Pickup trips are relevant to flight delay, baggage claim, and other factors.
  - Dropoff trips to airport are highly related to flight schedules.



Pickup trips from airport



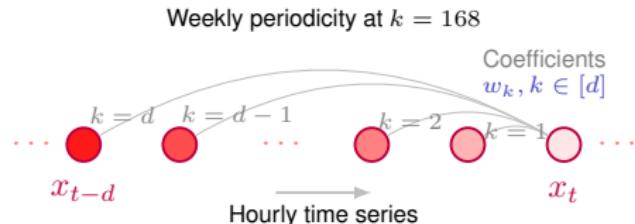
Dropoff trips to airport

- Sparse coefficient vectors (**sparsity  $\tau = 3$** ):

$$\mathbf{w} = (\underbrace{0.31}_{k=1}, \dots, \underbrace{0.28}_{k=24}, \dots, \underbrace{0.41}_{k=168})^\top \quad \text{vs.} \quad \mathbf{w} = (\underbrace{0.18}_{k=1}, \dots, \underbrace{0.35}_{k=24}, \dots, \underbrace{0.47}_{k=168})^\top$$

# Spatially- and Time-Varying Autoregression

## Univariate autoregression

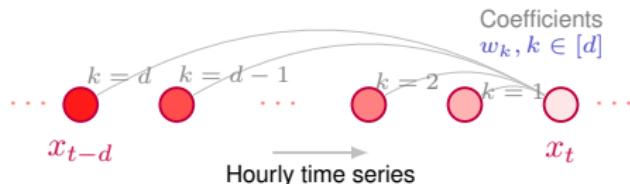


$$\min_t \sum_{k \in [d]} \left( x_t - \sum_{k \in [d]} w_k x_{t-k} \right)^2$$

# Spatially- and Time-Varying Autoregression

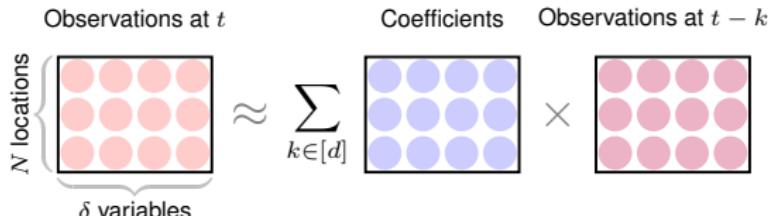
## Univariate autoregression

Weekly periodicity at  $k = 168$



$$\min_t \sum_{k \in [d]} \left( x_t - \sum_{k \in [d]} w_k x_{t-k} \right)^2$$

## Multidimensional autoregression



$$\min \sum_{n \in [N]} \sum_{\gamma \in [\delta]} \sum_t \left( x_{n, \gamma, t} - \sum_{k \in [d]} w_{n, \gamma, k} x_{n, \gamma, t-k} \right)^2$$

# Envisioning Human Mobility

- Ridesharing trip data  $\{x_{n,\gamma}\}$  across  $\gamma \in [\delta]$  years
- Reformulate sparse autoregression:

$$\min_{\{\mathbf{w}_{n,\gamma}\}, \boldsymbol{\beta}} \sum_{n \in [N]} \sum_{\gamma \in [\delta]} \sum_{t \in [d+1, T_\gamma]} \left( x_{n,\gamma,t} - \sum_{k \in [d]} w_{n,\gamma,k} x_{n,\gamma,t-k} \right)^2$$

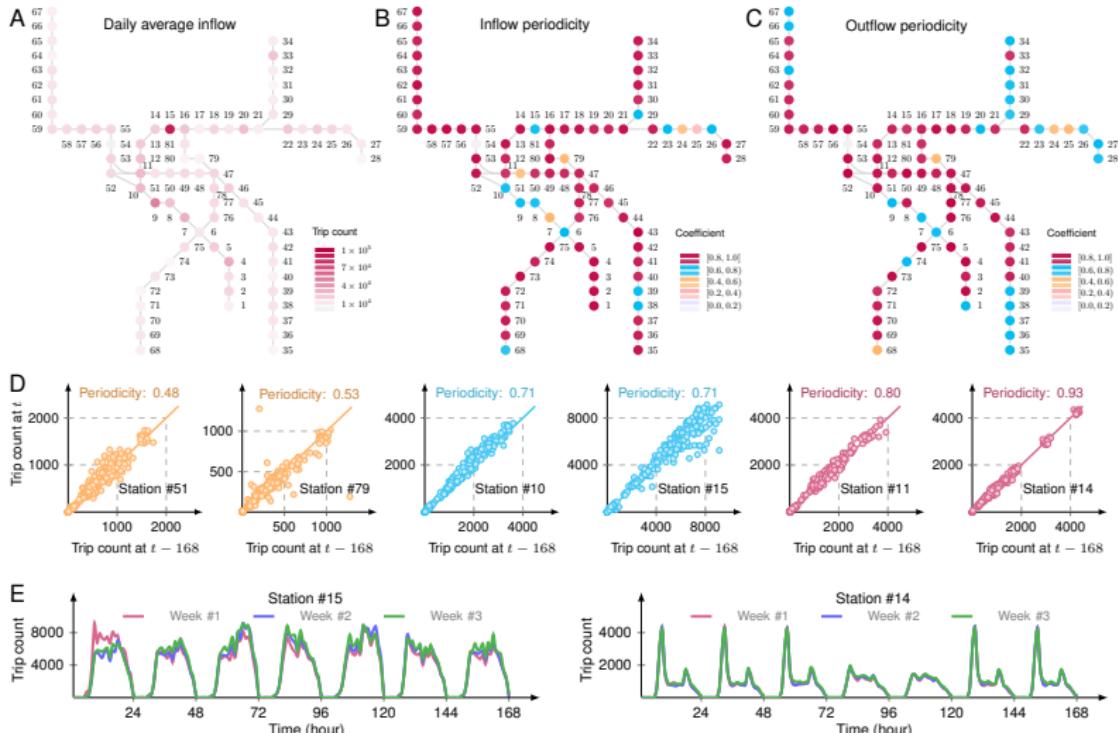
year  
spatial location      hourly time step

s.t.  $\underbrace{\boldsymbol{\beta} \in \{0, 1\}^d}_{\text{binary var.}}$      $\underbrace{0 \leq \mathbf{w}_{n,\gamma} \leq \boldsymbol{\beta}}_{\text{upper bound in } \{0, 1\}}$      $\underbrace{\|\boldsymbol{\beta}\|_1 \leq \tau}_{\text{sum of binary var.}}$

- MIP problem w/  $(N\delta + 1)d$  variables!
- How to handle thousands or millions of (e.g.,  $N\delta = 10^6$ ) time series?

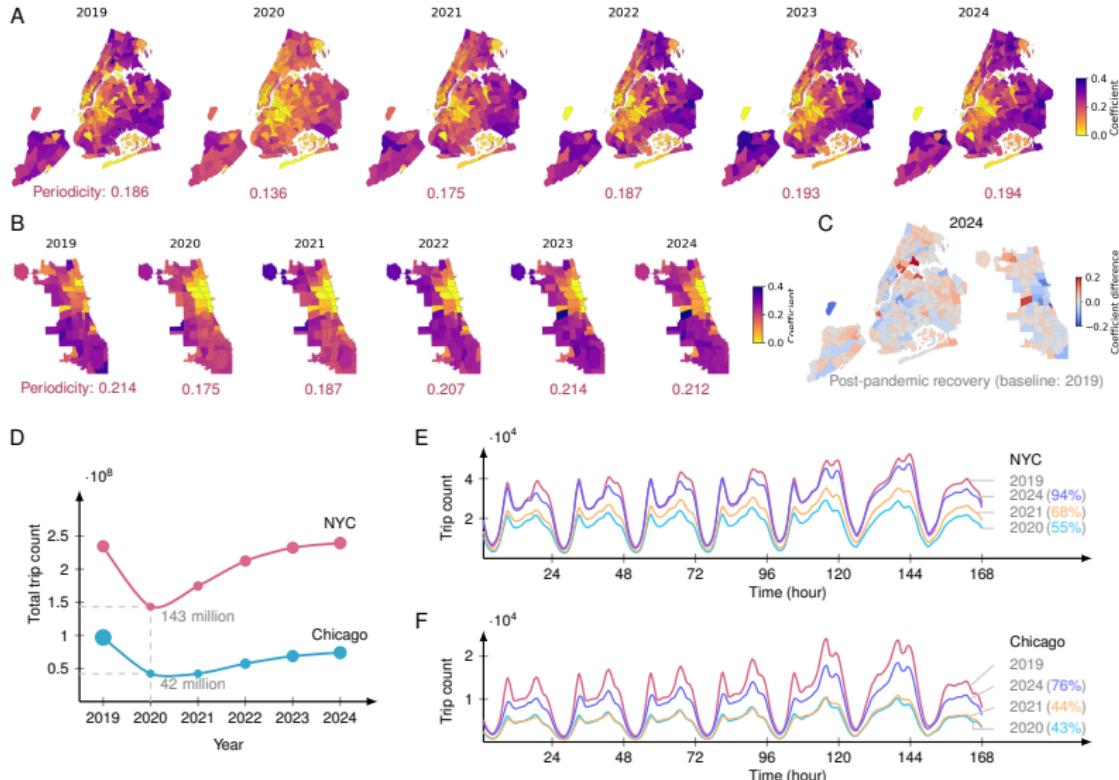
Quantify weekly periodicity by  $\{w_{n,\gamma,k}\}$  at index  $k = 168$

# Envisioning Human Mobility



Hangzhou metro passenger flow in January 2019

# Envisioning Human Mobility



Weekly periodicity reveals spatial patterns of ridesharing systems

# Understanding Climate Systems

Quantify yearly seasonality by  $\{w_{m,n,\gamma,k}\}$  at index  $k = 12$

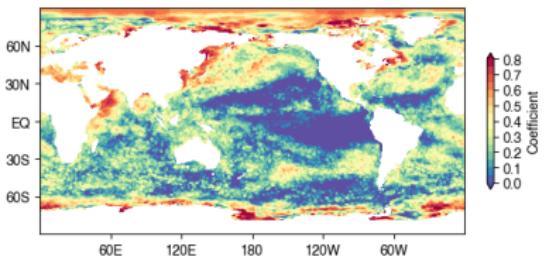
$$\min_{\{\boldsymbol{w}_{m,n,\gamma}\}, \boldsymbol{\beta}} \sum_{m \in [M]} \sum_{n \in [N]} \sum_{\gamma \in [\delta]} \sum_{t \in [d+1, T_\gamma]} \left( x_{m,n,\gamma,t} - \sum_{k \in [d]} w_{m,n,\gamma,k} x_{m,n,\gamma,t-k} \right)^2$$

s.t. 
$$\begin{cases} \boldsymbol{\beta} \in \{0, 1\}^d & \text{binary decision var.} \\ 0 \leq \boldsymbol{w}_{m,n,\gamma} \leq \boldsymbol{\beta}, \forall m, n, \gamma \\ \|\boldsymbol{\beta}\|_1 \leq \tau \\ \|\boldsymbol{w}_{m,n,\gamma}\|_1 = 1, \forall m, n, \gamma \end{cases}$$

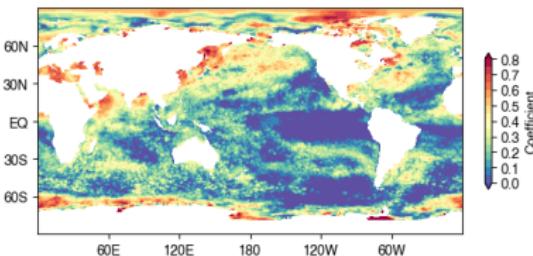
$\ell_1$ -normalization

longitude  
latitude    decade    monthly time step

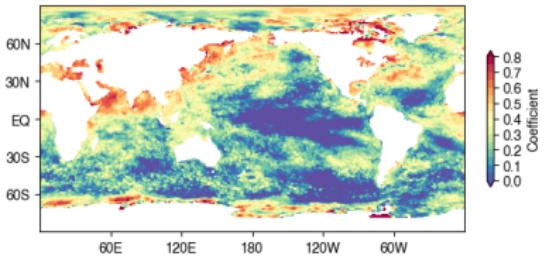
# Sea Surface Temperature



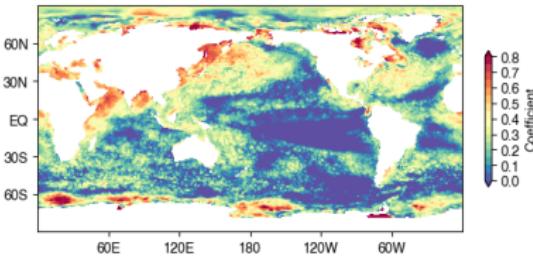
1980s



1990s



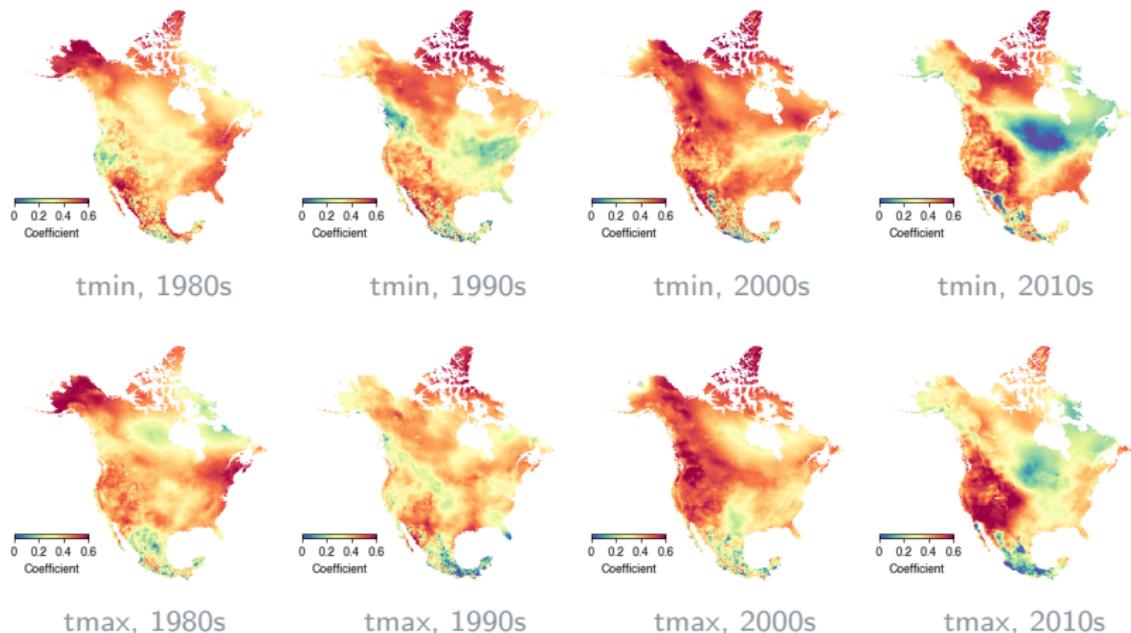
2000s



2010s

- Identify yearly periodicity at  $k = 12$  from SST data ( $\tau = 4$ )
  - ❶ The areas of El Niño events are less seasonal/predictable
  - ❷ Arctic becomes less seasonal/predictable in the past 20 years
- Insights into climate system monitoring

## North America Temperature



- Identify yearly periodicity at  $k = 12$  from temperature data ( $\tau = 4$ )
  - ❶ Stronger yearly seasonality in high-latitude areas
  - ❷ Less seasonal temperature in south areas (e.g., Mexico)
  - ❸ Seasonality patterns in 2000s & 2010s are different from 1980s & 1990s



UNIVERSITY OF  
CENTRAL FLORIDA

# Thanks for your attention!

## Any Questions?

Slides: [https://xinychen.github.io/slides/ml\\_opt\\_stsystem.pdf](https://xinychen.github.io/slides/ml_opt_stsystem.pdf)

Essential AR: [https://xinychen.github.io/slides/essential\\_ar.pdf](https://xinychen.github.io/slides/essential_ar.pdf)

### About me:

- 🏠 Homepage: <https://xinychen.github.io>
- 🏠 MIT sites: <https://sites.mit.edu/xinychen>
- ✉ How to reach me: [xinychen@mit.edu](mailto:xinychen@mit.edu)