



Machine Learning and Optimization for Data-Driven Transportation Analytics and Beyond

Tensor Decomposition, Interpretable ML, Mathematical Programming

Xinyu Chen

Postdoctoral Associate, MIT

February 7, 2025

Transport Data

- Transport & mobility application scenarios



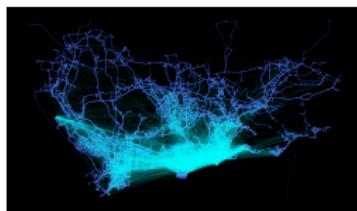
Highway (Portland)



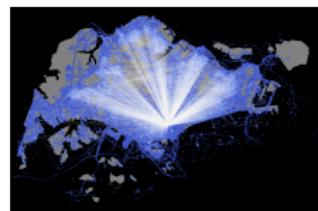
Uber movement (NYC)



Uber movement (Seattle)



Taxi trajectory (Shenzhen)



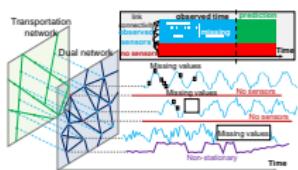
Human movement (Singapore)

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

Data-Driven Transportation Analytics

Part I

Traffic Imputation

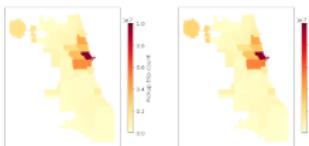


Chen et al.'24

Matrix/tensor completion
Circulant matrix
Circular convolution
Fast Fourier transform

Part II

Pattern Discovery

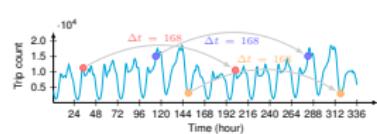


Chen et al.'24

Time series autoregression
Tensor decomposition
Conjugate gradient
Procrustes problems

Part III

Periodicity Qualification



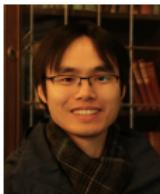
Working paper

Time series autoregression
Sparse learning
Greedy methods
Mixed-integer programming

ML + Optimization for Transportation

Spatiotemporal Traffic Data Imputation

(Matrix/Tensor Factorization)



Xinyu Chen
UdeM → MIT



Nicolas Saunier
PolyMtl

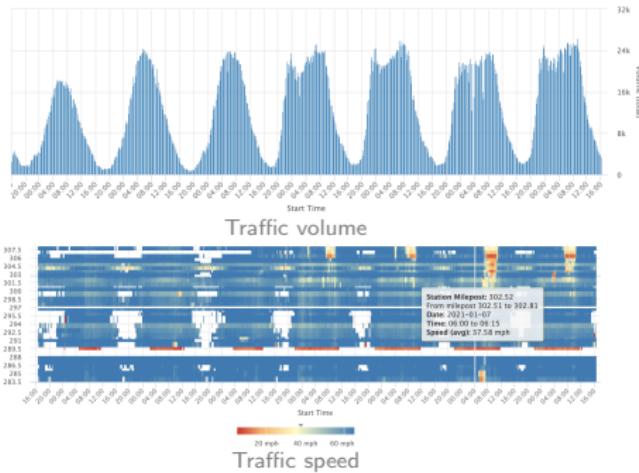


Lijun Sun
McGill

- **Chen & Sun (2022).** “Bayesian Temporal Factorization for Multidimensional Time Series Prediction”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (9): 4659–4673.
- **Chen et al. (2024).** “Laplacian Convolutional Representation for Traffic Time Series Imputation”. *IEEE Transactions on Knowledge and Data Engineering*, 36 (11): 6490–6502.
- **Chen et al. (2024).** “Forecasting Urban Traffic States with Sparse Data Using Hankel Temporal Matrix Factorization”. *INFORMS Journal on Computing*.

Traffic Flow Data

- Portland highway traffic data¹



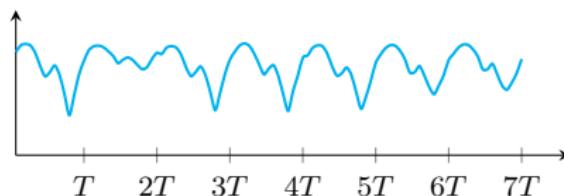
- $X \in \mathbb{R}^{N \times T}$ with N spatial locations $\times T$ time steps
- Traffic volume/speed shows strong spatial/temporal dependencies
- Missing data are there, how to improve data quality?

¹<https://portal.its.pdx.edu/home>

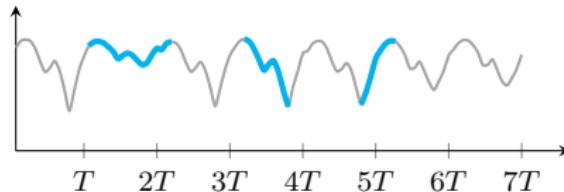
Time Series Imputation

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 →

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

(Circulant) Laplacian matrix

- Laplacian kernel:

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

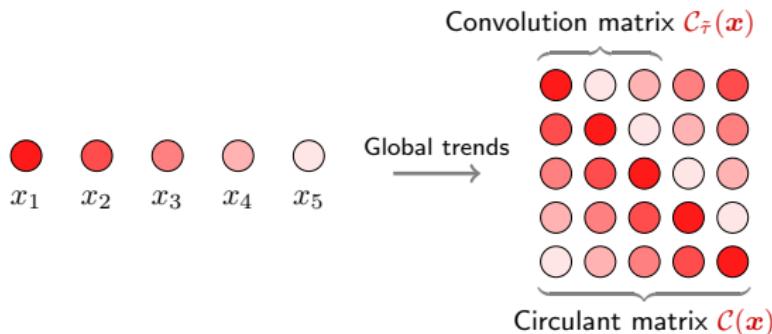
extending to the degree 2τ (i.e., graph connectivity) for $\mathbf{x} \in \mathbb{R}^T$.

- Temporal regularization:

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

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 Ω .

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 Ω .

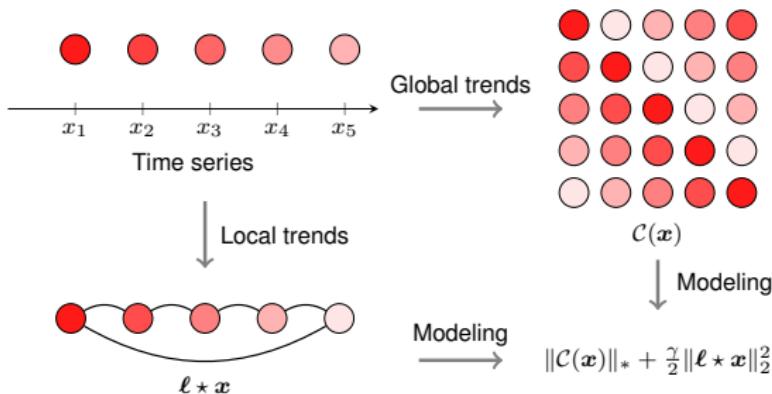
Global + Local Trends?

Laplacian Convolutional Representation (LCR)

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

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

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



Laplacian Convolutional Representation

- Augmented Lagrangian function:²

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

- Optimize \mathbf{x} w/ FFT (Properties of circulant matrix & circular convolution)

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

- Reformulate the optimization as ℓ_1 -norm minimization:

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

in $\mathcal{O}(T \log T)$ time.

² $\mathbf{w} \in \mathbb{R}^T$ (Lagrange multiplier); $\langle \cdot, \cdot \rangle$ (inner product).

Laplacian Convolutional Representation

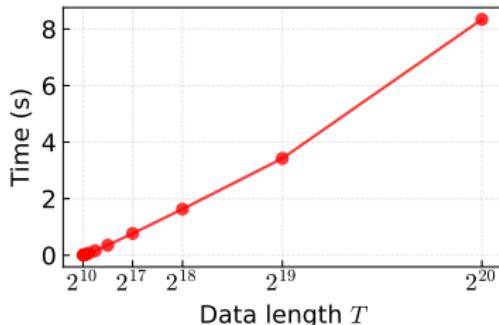
ℓ_1 -norm Minimization (Liu & Zhang'23)

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

with complex-valued $\hat{\mathbf{x}}, \hat{\mathbf{h}} \in \mathbb{C}^T$ and weight parameter δ , the solution is

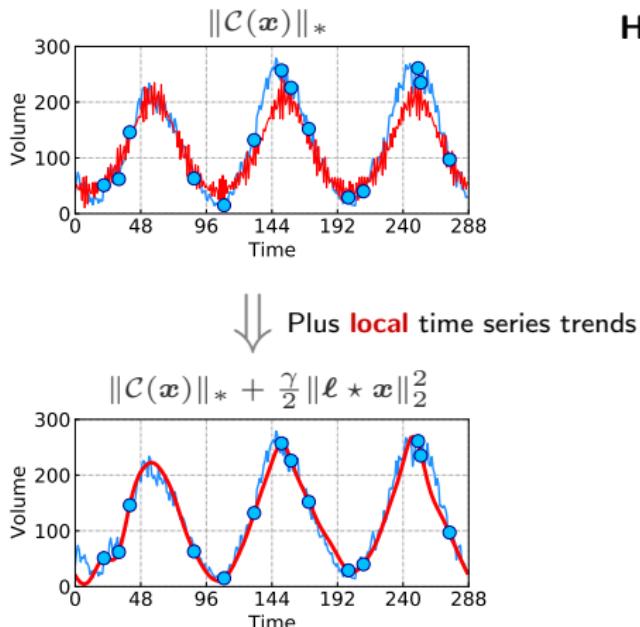
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \underbrace{\max\{0, |\hat{h}_t| - 1/\delta\}}_{\text{shrinkage (e.g., ReLU)}}, t \in [T].$$

- **Empirical time complexity of LCR:** On the synthetic data $\mathbf{y} \in \mathbb{R}^T$ with $T \in \{2^{10}, 2^{11}, \dots, 2^{20}\}$ (w.r.t. FFT in $\mathcal{O}(T \log T)$ time)



Experiments

- Traffic speed imputation³ (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:

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

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

³Blue dot: partial observation; red line: imputation.

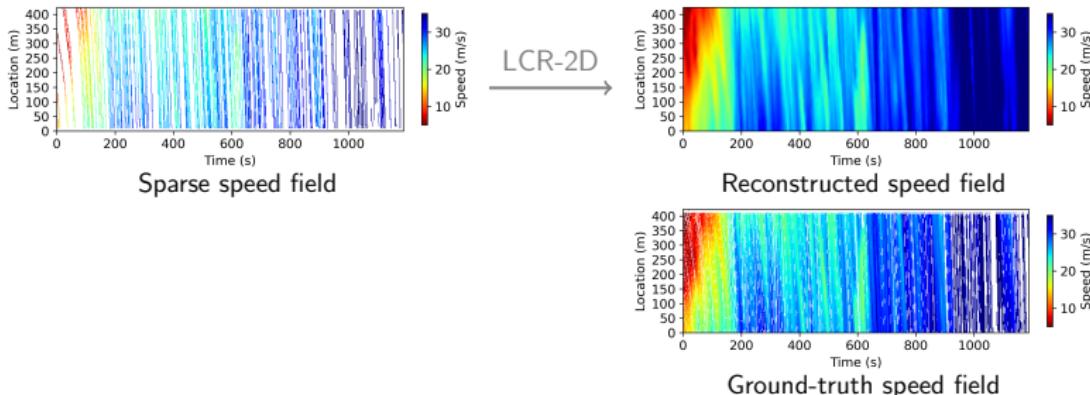
Experiments

Speed field reconstruction in German highways⁴

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

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

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



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

Discovering Dynamic Patterns from Spatiotemporal Systems



Xinyu Chen
UdeM → MIT



Nicolas Saunier
PolyMtl



Lijun Sun
McGill



Dingyi Zhuang
MIT



HanQin Cai
UCF



Shenhao Wang
UF

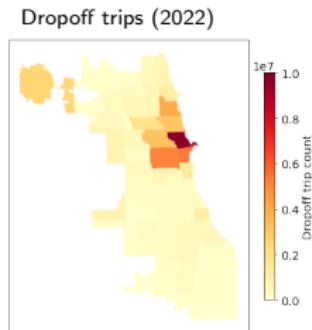
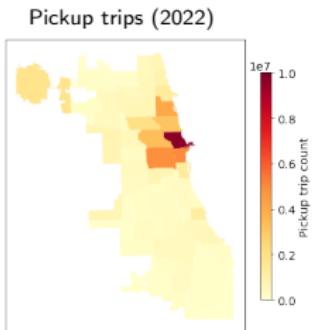
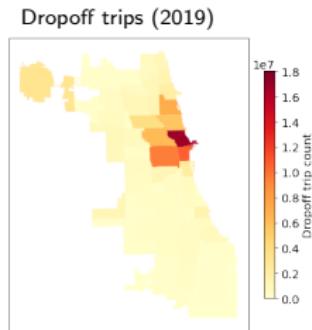
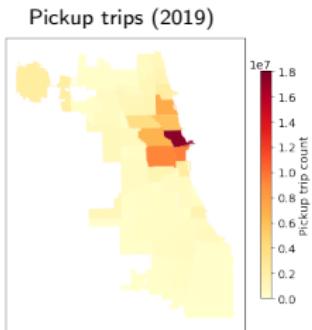


Jinhua Zhao
MIT

- **Chen et al. (2024).** "Discovering Dynamic Patterns from Spatiotemporal Data with Time-Varying Low-Rank Autoregression". *IEEE Transactions on Knowledge and Data Engineering*, 36 (2): 504–517.
- **Chen et al. (2025).** "Dynamic autoregressive tensor factorization for pattern discovery of spatiotemporal systems". Major revision at *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Motivation

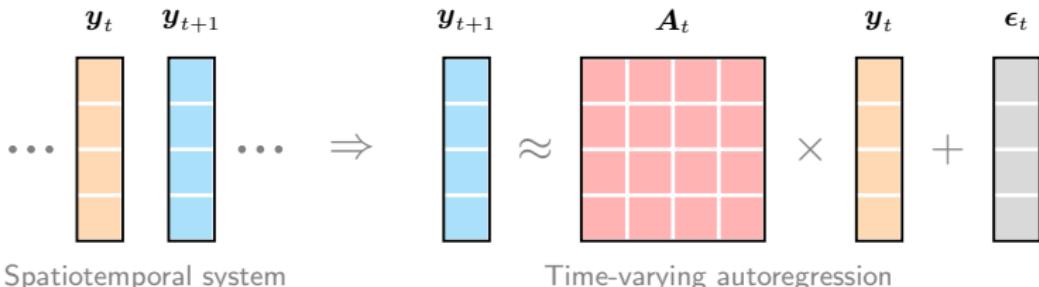
- Chicago rideshare: 96.6M trips (2019) vs. 57.3M trips (2022)



- Contradict:** No difference in aggregation. How about latent patterns?

Autoregression

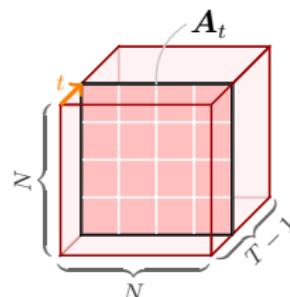
- How to characterize dynamical systems?

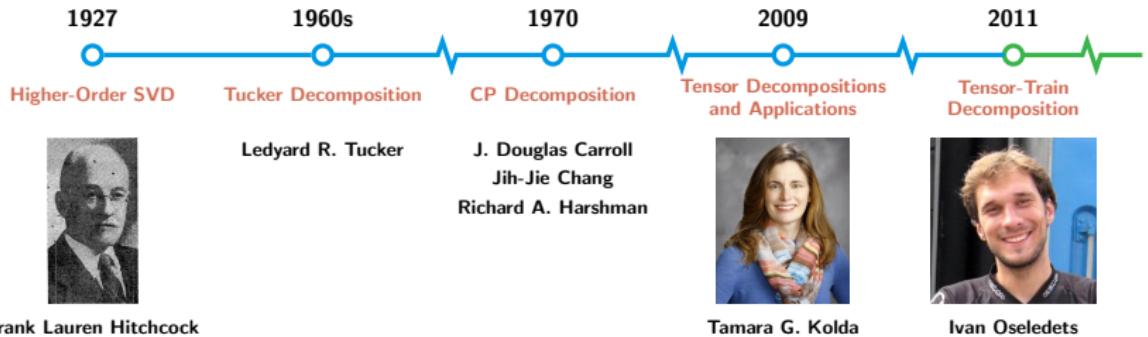


- On spatiotemporal data $\mathbf{Y} \in \mathbb{R}^{N \times T}$:

$$\underbrace{\mathbf{y}_{t+1} = \mathbf{A}\mathbf{y}_t + \epsilon_t}_{\text{time-invariant (e.g., DMD)}} \quad \text{v.s.} \quad \underbrace{\mathbf{y}_{t+1} = \mathbf{A}_t \mathbf{y}_t + \epsilon_t}_{\text{time-varying (ours)}}$$

- How to discover spatial/temporal modes (patterns) from the tensor $\mathcal{A} \triangleq \{\mathbf{A}_t\}_{t \in [T-1]}$?

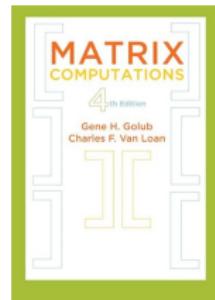




Dynamic Autoregressive Tensor Factorization

- Tensor factorization:

$$\mathcal{A} = \underbrace{\mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{V} \times_3 \mathbf{X}}_{\text{Tucker decomposition}}$$
$$\Downarrow$$
$$\mathbf{A}_t = \mathcal{G} \times_1 \underbrace{\mathbf{W}}_{\text{spatial modes}} \times_2 \mathbf{V} \times_3 \underbrace{\mathbf{x}_t^\top}_{\text{temporal modes}}$$



- Optimization problem:

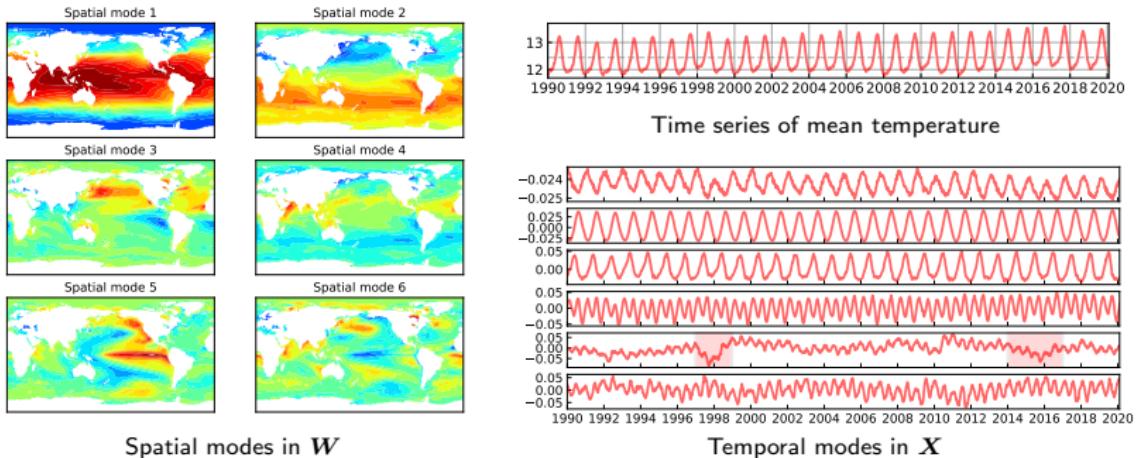
$$\min_{\mathcal{G}, \mathbf{W}, \mathbf{V}, \mathbf{X}} \frac{1}{2} \sum_{t \in [T-1]} \| \mathbf{y}_{t+1} - (\mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{V} \times_3 \mathbf{x}_t^\top) \mathbf{y}_t \|_2^2$$

s.t. $\underbrace{\mathbf{W}^\top \mathbf{W} = \mathbf{I}_R}_{\text{orthogonal spatial modes}}$

on spatiotemporal data $\mathbf{Y} \in \mathbb{R}^{N \times T}$.

Beyond Transport: Sea Surface Temperature

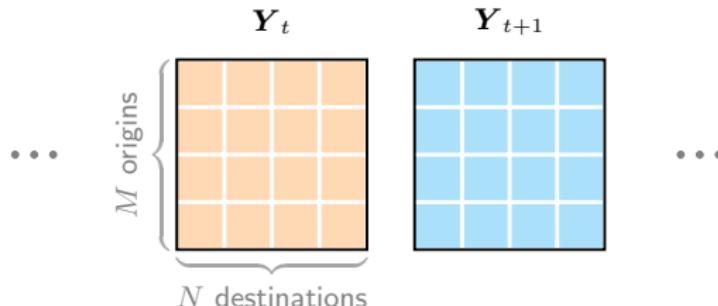
- Sea surface temperature (**SST**) dataset



- Identify two strongest El Nino events (on 1997-98 & 2014-16)
- Insights into climate change

Dynamic Autoregressive Tensor Factorization

- Origin-Destination (OD) matrices



- On spatiotemporal systems $\mathcal{Y} \in \mathbb{R}^{M \times N \times T}$:

$$\mathbf{y}_{n,t+1} = \underbrace{\mathbf{A}_{n,t} \mathbf{y}_{n,t} + \boldsymbol{\epsilon}_{n,t}}_{\text{time-varying \& destination-varying}}$$

- Optimization problem:

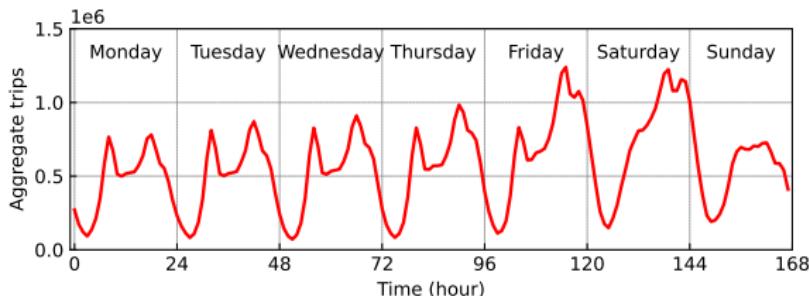
$$\min_{\mathcal{G}, \mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{x}} \frac{1}{2} \sum_{n \in [N]} \sum_{t \in [T-1]} \left\| \mathbf{y}_{n,t+1} - \underbrace{(\mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{U} \times_3 \mathbf{V} \times_4 \mathbf{x}_t^\top) \mathbf{y}_{n,t}}_{\text{Tucker decomposition}} \right\|_2^2$$

$$\text{s.t. } \underbrace{\mathbf{W}^\top \mathbf{W} = \mathbf{I}_R}_{\text{orthogonal origin patterns}}$$

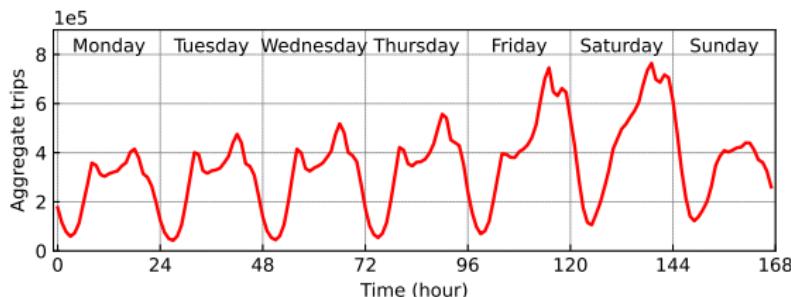
Human Mobility

- Chicago rideshare: 96.6M trips (2019) vs. 57.3M trips (2022)

Pickup trips aggregated over 52 weeks in 2019

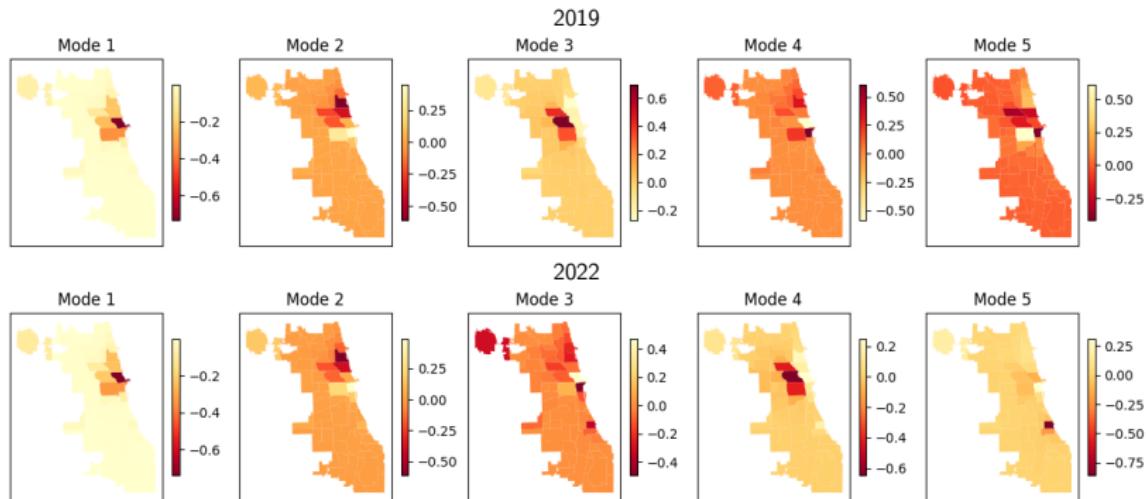


Pickup trips aggregated over 52 weeks in 2022



Human Mobility

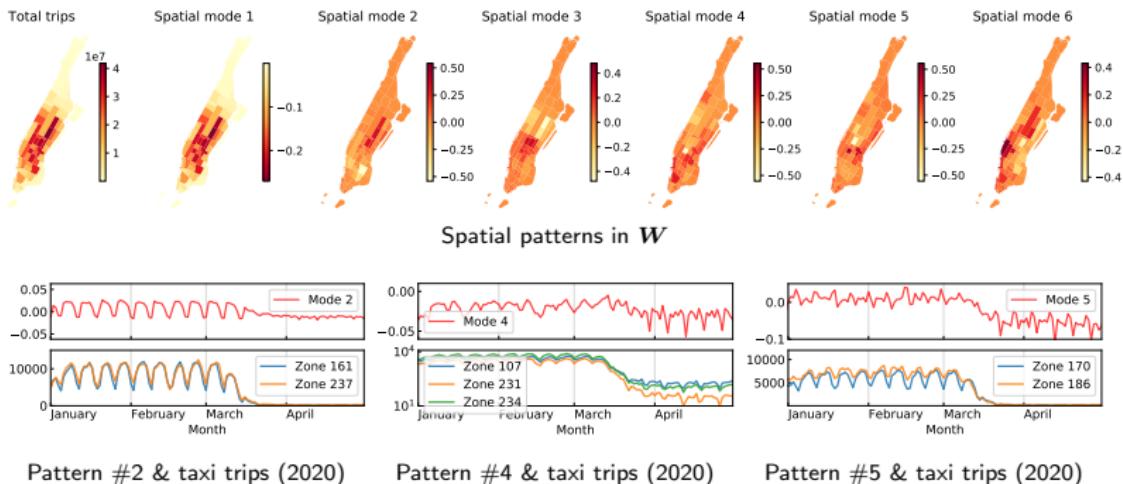
- Rideshare trip data: 77 origins \times 77 destinations \times 168 hours



- Identify the changes in pickup zones before/after COVID-19

Human Mobility

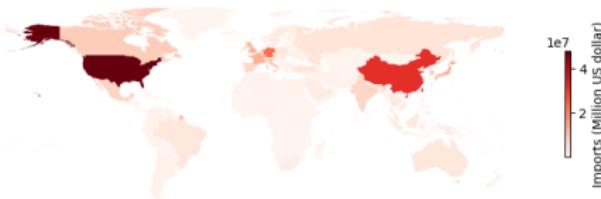
- NYC taxi dataset (pickup)
 - Remarkable decrease of trips due to COVID-19



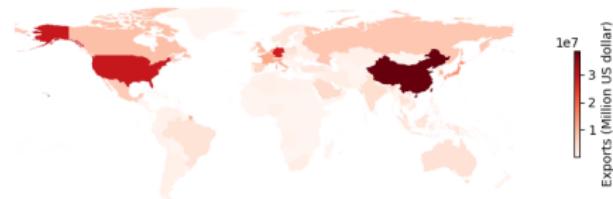
- Identify the changes of temporal patterns due to COVID-19

Beyond Transport: International Trade

- Import/Export merchandise trade values (annual)⁵ (215 countries/regions & period of 2000-2022)
 - Total merchandise trade values
 - Represent import/export trade data as a 215-by-23 matrix



Imports from 2000 to 2022



Exports from 2000 to 2022

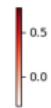
⁵The dataset is available at <https://stats.wto.org>.



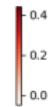
Import pattern 1



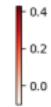
Import pattern 2



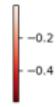
Import pattern 3



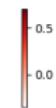
Import pattern 4



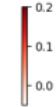
Export pattern 1



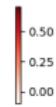
Export pattern 2



Export pattern 3

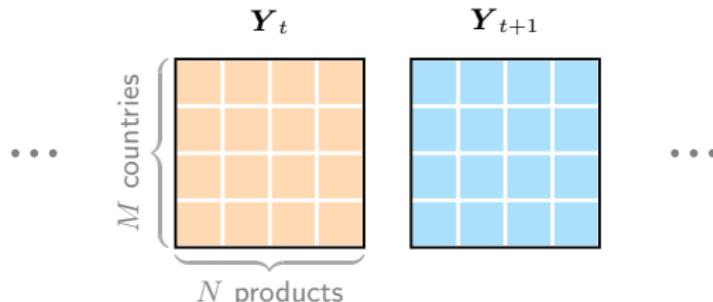


Export pattern 4



Beyond Transport: International Trade

- Three-dimensional trade (Economy, Product, Year)



- On spatiotemporal systems $\mathcal{Y} \in \mathbb{R}^{M \times N \times T}$:

$$\mathbf{y}_{n,t+1} = \underbrace{\mathbf{A}_{n,t} \mathbf{y}_{n,t} + \boldsymbol{\epsilon}_{n,t}}_{\text{time-varying \& product-varying}}$$

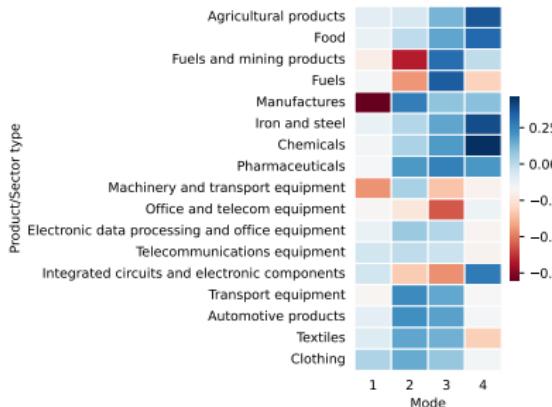
- Optimization problem of DATEF:

$$\min_{\mathcal{G}, \mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{x}} \frac{1}{2} \sum_{n \in [N]} \sum_{t \in [T-1]} \|\mathbf{y}_{n,t+1} - (\mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{U} \times_3 \mathbf{V} \times_4 \mathbf{x}_t^\top) \mathbf{y}_{n,t}\|_2^2$$

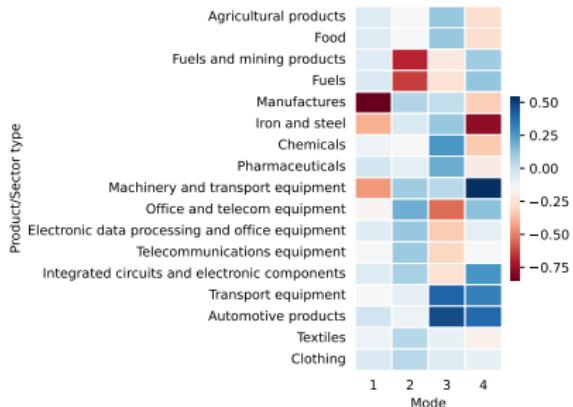
s.t.
$$\underbrace{\mathbf{W}^\top \mathbf{W} = \mathbf{I}_R}_{\text{orthogonal country patterns}}$$

Product Patterns

- On 17 merchandise types



Imports



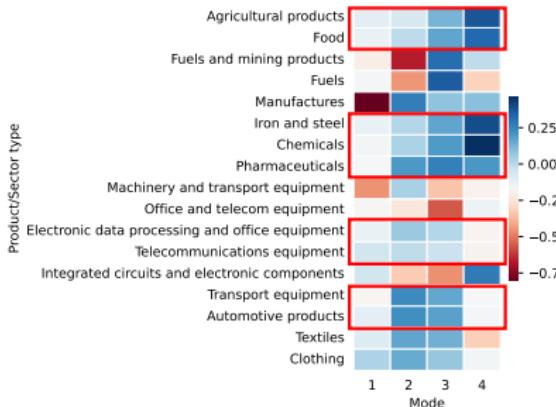
Exports

- Classify import/export merchandise according to product patterns
- Basic principle:

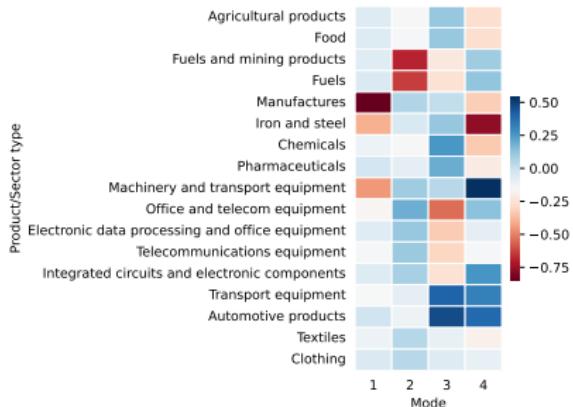
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



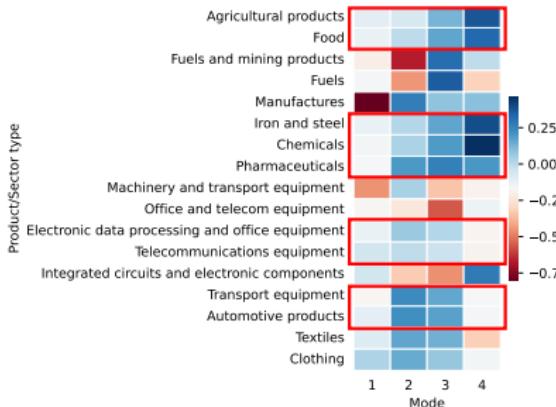
Exports

- Classify import/export merchandise according to product patterns
- Basic principle:

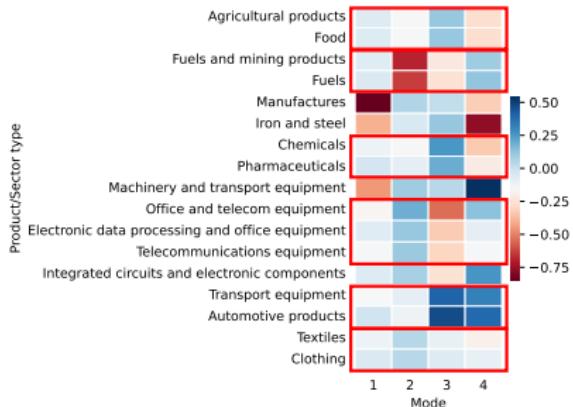
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



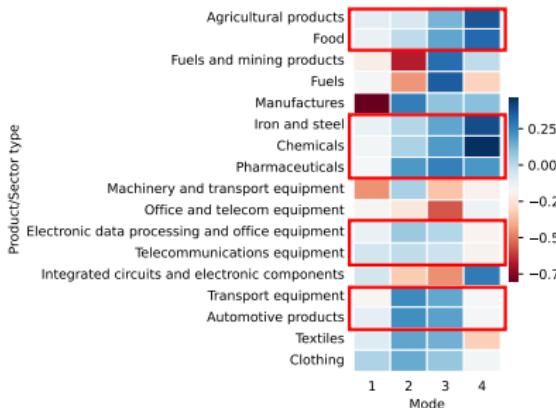
Exports

- Classify import/export merchandise according to product patterns
- Basic principle:

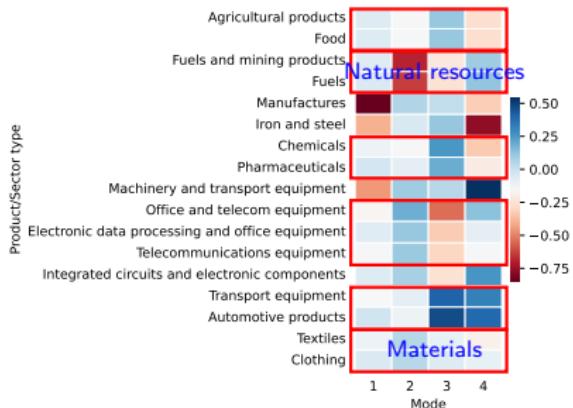
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



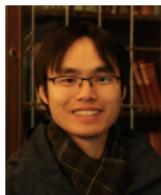
Exports

- Classify import/export merchandise according to product patterns
- Basic principle:

Import: What we buy? (demand) vs. Export: What we sell? (supply)

Quantifying Periodicity of Human Mobility

(Working Ideas on Interpretable ML & Causality & Nonlinear Programming)



Xinyu Chen
MIT DUSP



Dingyi Zhuang
MIT CEE



Yunhan Zheng
MIT SMART



Shenhao Wang
UF Urban AI



Jinhua Zhao
MIT DUSP



Ryan Qi Wang
Northeastern CEE

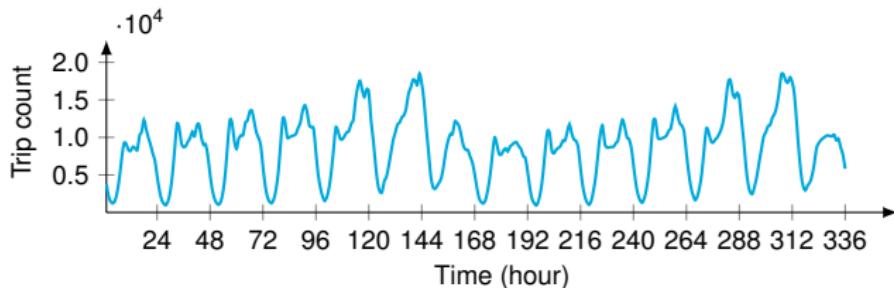


Lijun Ding
UCSD Math



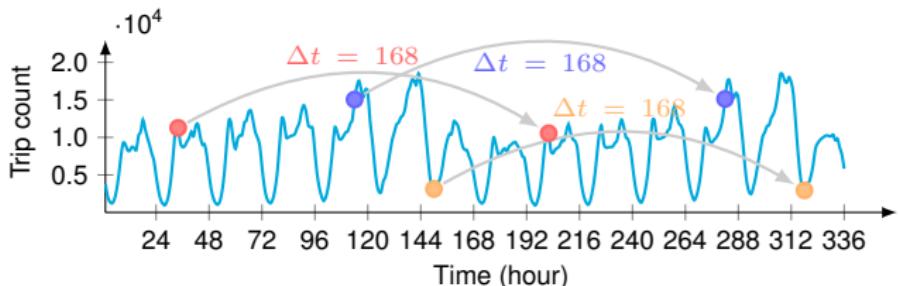
Vassilis Digalakis Jr
HEC Paris

Motivation



Two-week rideshare trip time series in Chicago since April 1, 2024

Motivation



Weekly periodicity of rideshare trip time series

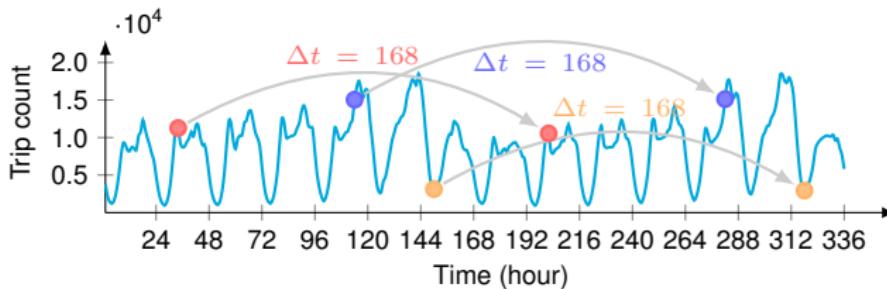
What motivates us most about **periodicity**?

- ❶ **Resilience and stability of systems:** Empirically measure the periodicity and predictability of urban systems.
- ❷ **Optimization of transport systems:** Optimize resources (e.g., public transit, taxi, rideshare, and micromobility) to meet transport demand efficiently.
- ❸ **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

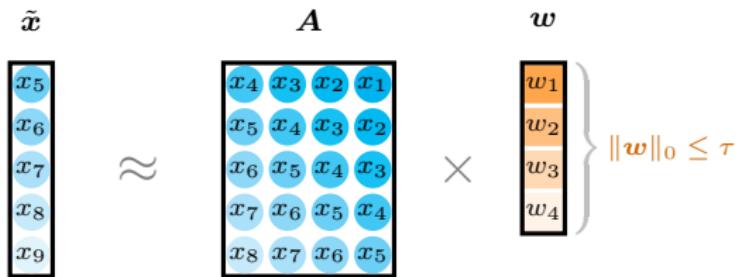
- Sparse coefficient vector \mapsto **Interpretability?**

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

Valorizing Autoregression

- Time series autoregression

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



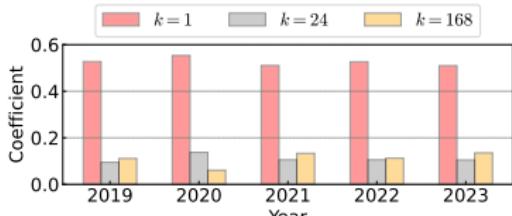
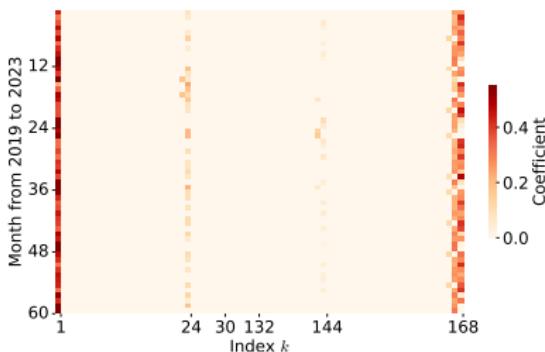
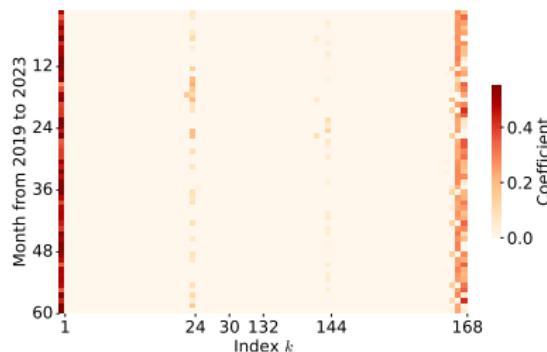
dth-order autoregression on time series $\mathbf{x} = (x_1, x_2, \dots, x_9)^\top$ w/ integer $\tau \in \mathbb{Z}^+$

- Sparse autoregression (Subspace Pursuit Algorithm (Dai & Milenkovic'09))

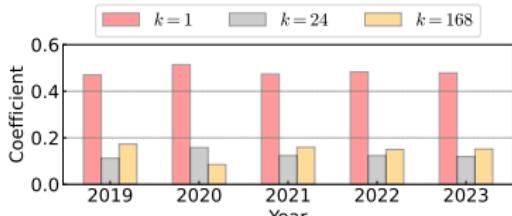
$$\begin{aligned} \min_{\mathbf{w} \geq 0} \quad & \| \tilde{\mathbf{x}} - \mathbf{A}\mathbf{w} \|_2^2 \\ \text{s.t.} \quad & \underbrace{\|\mathbf{w}\|_0 \leq \tau}_{\text{Sparsity w/ } \ell_0\text{-norm}} \end{aligned}$$

Envisioning Human Mobility

- NYC rideshare: Sparse coefficient vectors across different months



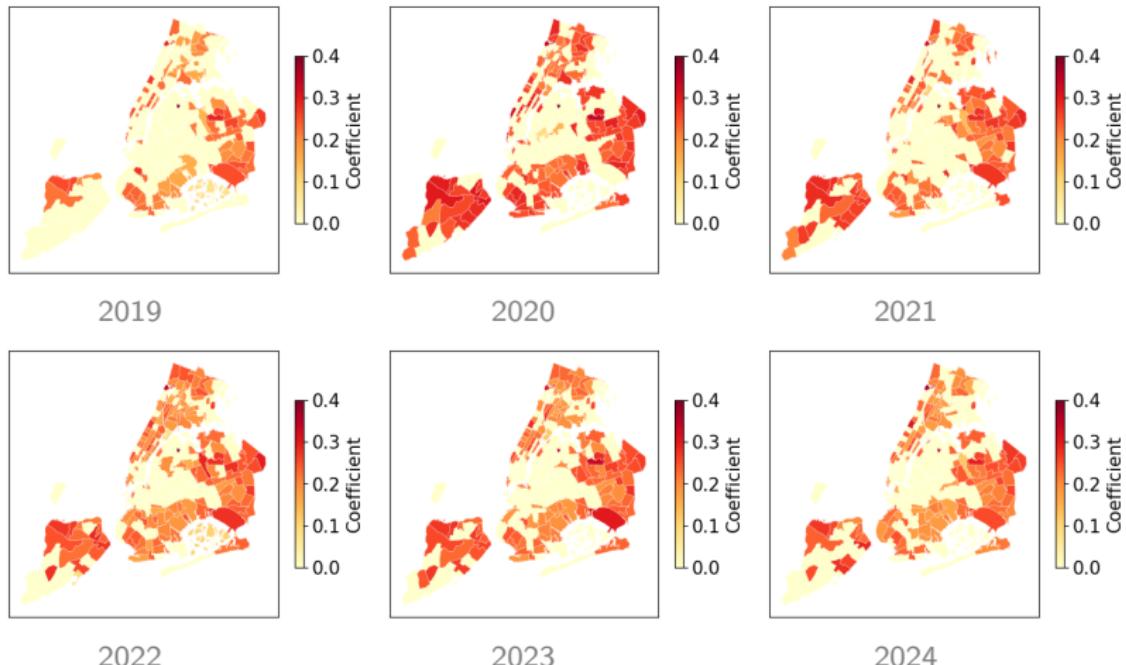
NYC pickup trips



NYC dropoff trips

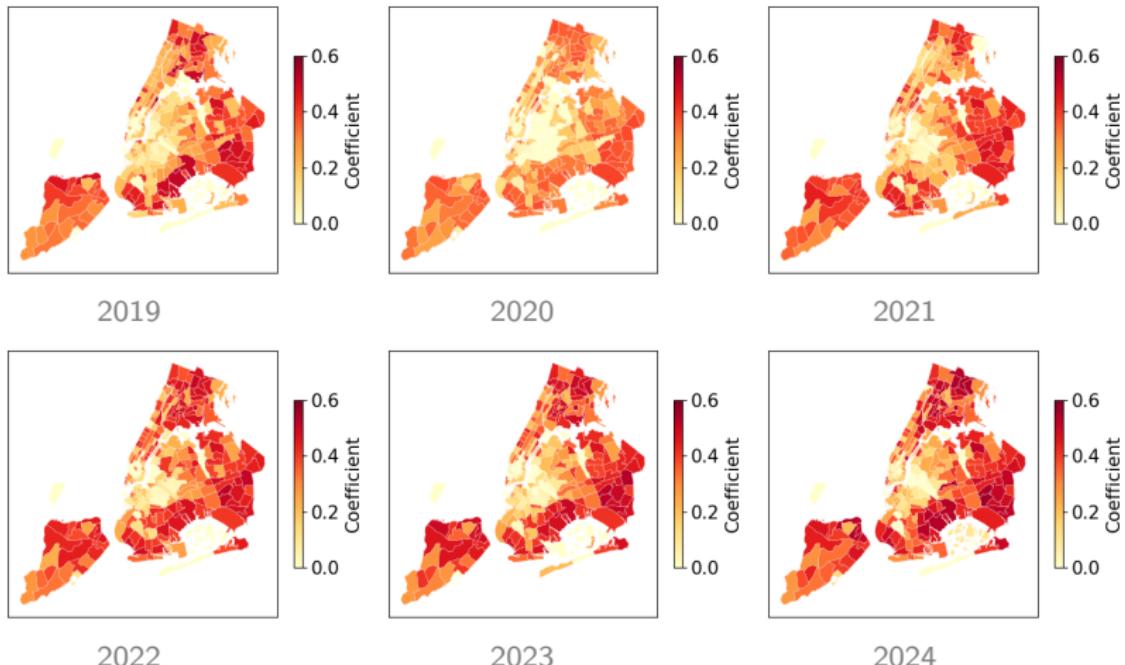
- Stronger daily periodicity vs. weaker weekly periodicity in 2020

Stronger Daily Periodicity in 2020



- High-demand areas are less periodic
- More areas show remarkable daily periodicity in 2020

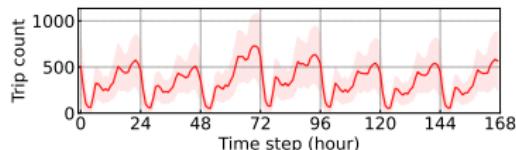
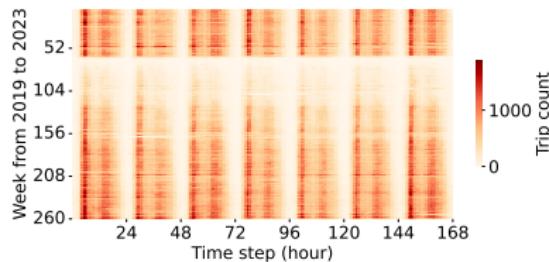
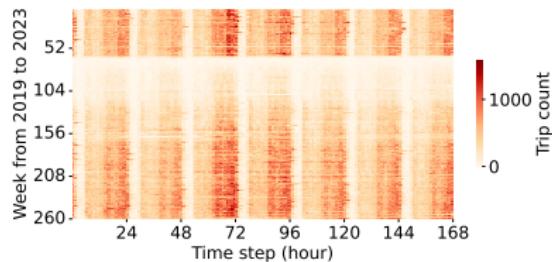
Weaker Weekly Periodicity in 2020



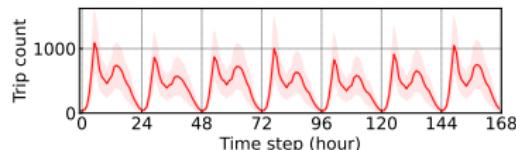
- High-demand areas are less periodic
- Less areas show remarkable weekly periodicity in 2020

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



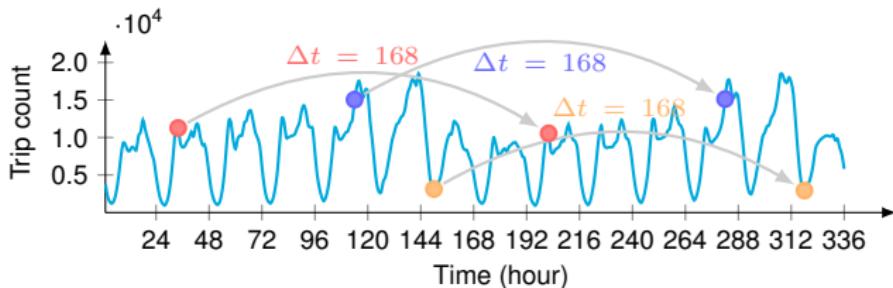
Dropoff trips

- Sparse coefficient vectors:

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

Solution Quality

- Subspace pursuit (SP) sometimes fails



Periodicity of rideshare trip time series

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

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

Time-Varying Transport Systems

- Rideshare trip data $\{\mathbf{x}_\gamma\}_{\gamma \in [\delta]}$ across $\gamma \in [\delta]$ months/years
- **(Ours)** Reformulate interpretable sparse autoregression:

$$\begin{aligned} & \min_{\{\mathbf{w}_\gamma\}_{\gamma \in [\delta]}} \sum_{\gamma \in [\delta]} \|\tilde{\mathbf{x}}_\gamma - \mathbf{A}_\gamma \mathbf{w}_\gamma\|_2^2 \\ \text{s.t. } & \begin{cases} \mathbf{w}_\gamma \geq 0 & (\text{non-negativity}) \\ \|\mathbf{w}_\gamma\|_0 \leq \tau & (\text{sparsity}) \\ \text{supp}(\mathbf{w}_\gamma) = \text{supp}(\mathbf{w}_{\gamma+1}) & (\text{no local difference}) \end{cases} \end{aligned}$$

making these coefficient vectors comparable across δ months/years.

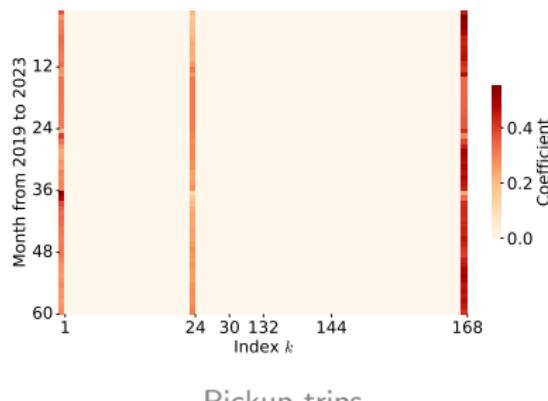
- Constraints w/ binary variables $\beta_\gamma \in \{0, 1\}^d$:

$$\underbrace{0 \leq \mathbf{w}_\gamma \leq \beta_\gamma}_{\text{upper bound } \{0, 1\}} \quad \underbrace{\sum_{k \in [d]} \beta_{\gamma, k} \leq \tau}_{\text{sum of binary var.}} \quad \underbrace{\beta_\gamma - \beta_{\gamma+1} = 0}_{\text{comparability across } \mathbf{w}_\gamma, \forall \gamma}$$

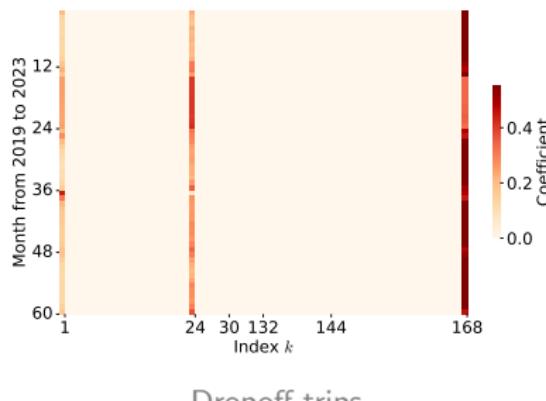
- MIP problem w/ $2d\delta$ decision variables!
- **(Efficiency?)** ML prunes the search space, e.g., $2\tau_0\delta$ decision variables ($\tau < \tau_0 \ll d$) instead.

John F. Kennedy International Airport

- Coefficients $\{w_\gamma\}_{\gamma \in [\delta]}$ at $S = \{\underbrace{1}_{\text{local}}, \underbrace{24}_{\text{daily}}, \underbrace{168}_{\text{weekly}}\}$ across $\delta = 60$ months
 - ① Stronger weekly periodicity of dropoff trips than pickup trips
 - ② Stronger daily periodicity in 2020
 - ③ Weaker weekly periodicity in 2020



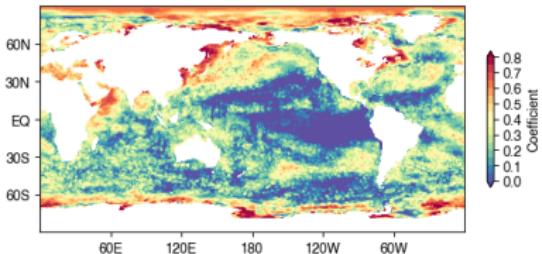
Pickup trips



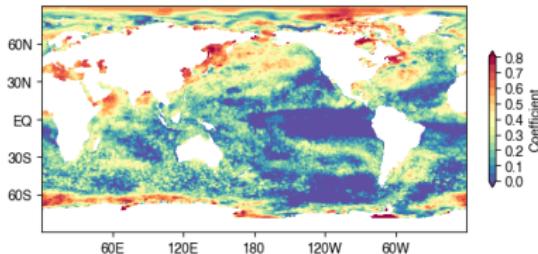
Dropoff trips

- Identify system patterns that evolve over time for human mobility

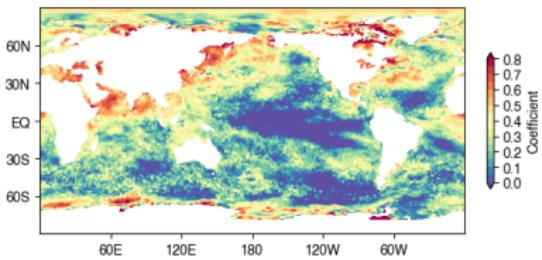
Beyond Transport: Sea Surface Temperature



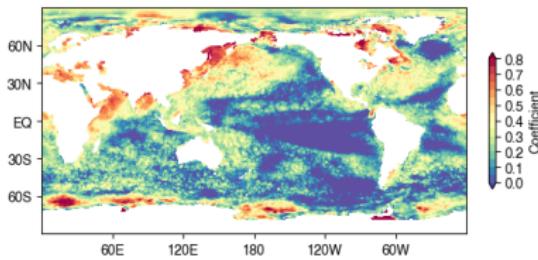
1980s



1990s



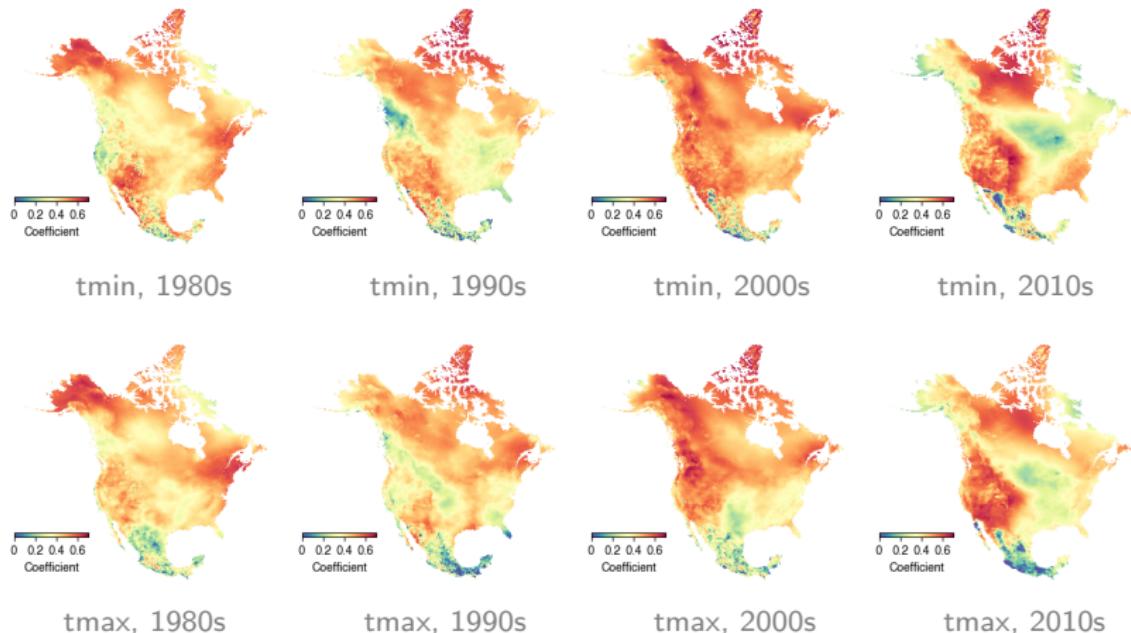
2000s



2010s

- Identify yearly periodicity at $k = 12$ from SST data ($\tau = 4$)
 - ❶ The areas of El Nino events are less seasonal/predictable
 - ❷ Arctic becomes less seasonal/predictable in the past 20 years
- Insights into climate change & global warming & sustainable development

Beyond Transport: North America Temperature



- Identify yearly periodicity at $k = 12$ from temperature data ($\tau = 4$)
 - ① Stronger yearly seasonality in high-latitude areas
 - ② Less seasonal temperature in south areas (e.g., Mexico)
 - ③ Seasonality patterns in 2000s & 2010s are different from 1980s & 1990s



Thanks for your attention!

Any Questions?

About me:

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