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

Tensor Decomposition, Interpretable ML, Mathematical Programming

Xinyu Chen

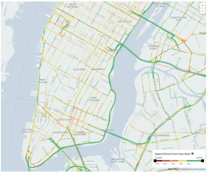
Postdoctoral Associate, MIT

February 7, 2025

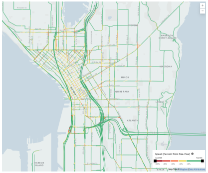
Transport Data

- Transport & mobility application scenarios

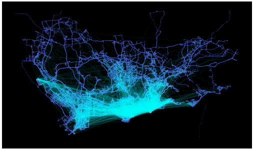
Highway (Portland)



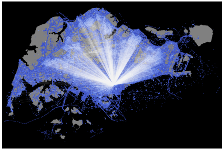
Uber movement (NYC)



Uber movement (Seattle)



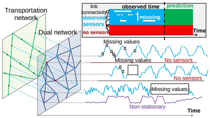
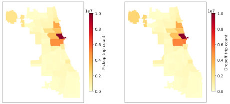
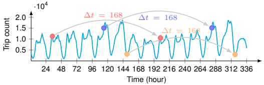
Taxi trajectory (Shenzhen)



Human movement (Singapore)

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

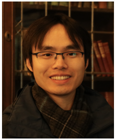
Data-Driven Transportation Analytics

Part I	Part II	Part III
Traffic Imputation	**Pattern Discovery**	**Periodicity Qualification**
		
Chen et al.'24	Chen et al.'24	Working paper
Matrix/tensor completion	Time series autoregression	Time series autoregression
Circulant matrix	Tensor decomposition	Sparse learning
Circular convolution	Conjugate gradient	Greedy methods
Fast Fourier transform	Procrustes problems	Mixed-integer programming

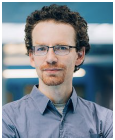
ML + Optimization for Transportation

Spatiotemporal Traffic Data Imputation

(Matrix/Tensor Factorization)



Xinyu Chen
UdeM $\rightarrow$ MIT



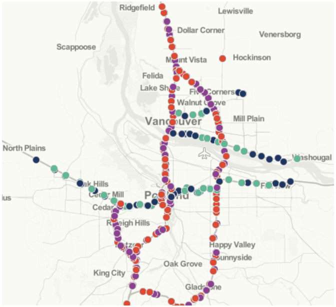
Nicolas Saunier
PolyMtl

Lijun Sun
McGill

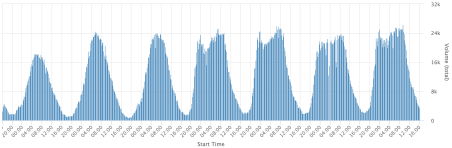
- **Chen** & Sun (2022). "Bayesian Temporal Factorization for Multidimensional Time Series Prediction". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (9): 4659–4673.
- **Chen** et al. (2024). "Laplacian Convolutional Representation for Traffic Time Series Imputation". *IEEE Transactions 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

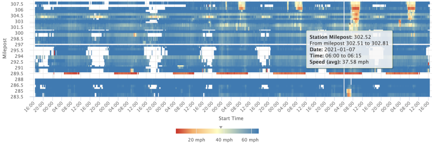
- Portland highway traffic data[1]



Highway network & sensor locations



Traffic volume



Traffic speed

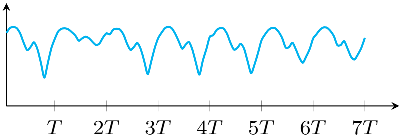
- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps
- Traffic volume/speed shows strong spatial/temporal dependencies
- Missing data are there, how to improve data quality?

[1] https://portal.its.pdx.edu/home

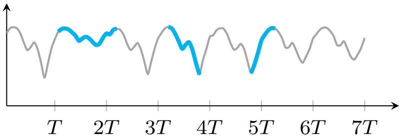
Time Series Imputation

Global/local trends in sparse data?

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

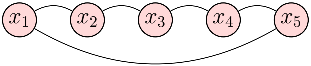


- Local trends (e.g., short-term time series trends):



Local Trend Modeling

- Intuition of Laplacian matrix



Undirected and circulant graph

$$\xrightarrow{\text{Modeling}} \quad \boldsymbol{L}=\begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

(Circulant) Laplacian matrix

- Laplacian kernel:

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

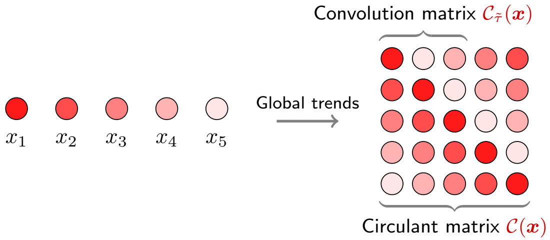
extending to the degree 2τ (i.e., graph connectivity) for $\boldsymbol{x} \in \mathbb{R}^T$.

- Temporal regularization:

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

Global Trend Modeling

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



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

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

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

on data $\boldsymbol{y}$ w/ observed index set Ω.

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

$$\min_{\boldsymbol{x}} \ \|\mathcal{C}_{\tilde{\tau}}(\boldsymbol{x})\|_*$$
$$\text{s.t. } \|\mathcal{P}_\Omega(\boldsymbol{x}-\boldsymbol{y})\|_2 \leq \epsilon$$

on data $\boldsymbol{y}$ w/ observed index set Ω.

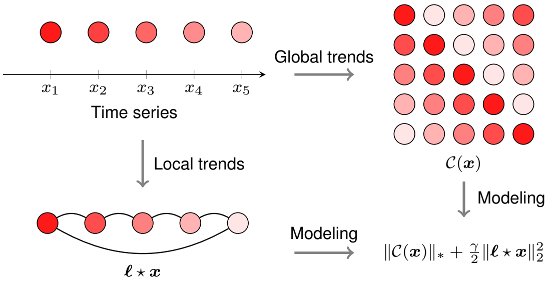
Global + Local Trends?

Laplacian Convolutional Representation (LCR)

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

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

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



Laplacian Convolutional Representation

- Augmented Lagrangian function:[2]

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

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

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

- Reformulate the optimization as ℓ_1-norm minimization:

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

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

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

Laplacian Convolutional Representation

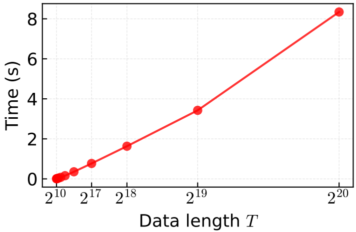
ℓ_1-norm Minimization (Liu & Zhang'23)

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

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

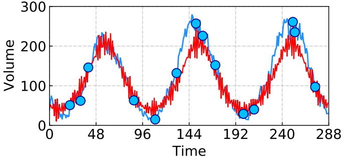
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \underbrace{\max\{0, |\hat{h}_t| - 1/\delta\}}_{\text{shrinkage (e.g., ReLU)}}, t \in [T].$$

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



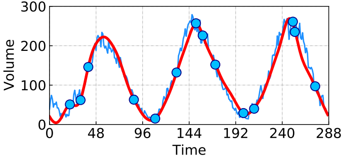
- Traffic speed imputation[3] (95% missing rate)

$\|\mathcal{C}(\boldsymbol{x})\|_*$



$\Downarrow$ Plus **local** time series trends

$\|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2$



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:

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

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

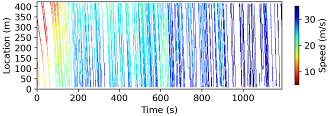
Experiments

Speed field reconstruction in German highways[4]

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

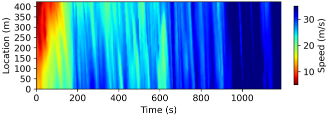
$$\min_{\boldsymbol{X}} \quad \underbrace{\|\mathcal{C}(\boldsymbol{X})\|_*}_{\text{global trend}} + \frac{\gamma}{2} \underbrace{\|(\boldsymbol{\ell}_s \boldsymbol{\ell}^\top) \star \boldsymbol{X}\|_F^2}_{\text{local trend}}$$

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

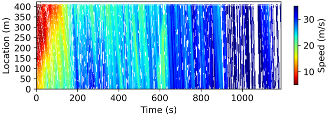


Sparse speed field

LCR-2D $\longrightarrow$



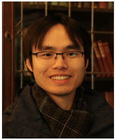
Reconstructed speed field



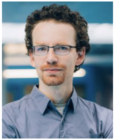
Ground-truth speed field

[4] Highway Drone (HighD) dataset at https://www.highd-dataset.com/

Discovering Dynamic Patterns from Spatiotemporal Systems



Xinyu Chen
UdeM → MIT



Nicolas Saunier
PolyMtl

Lijun Sun
McGill

Dingyi Zhuang
MIT

HanQin Cai
UCF

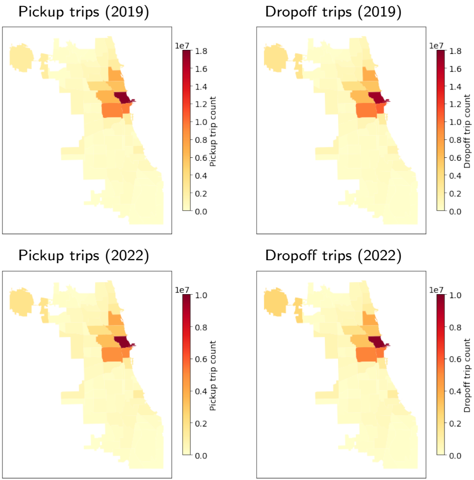
Shenhao Wang
UF

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

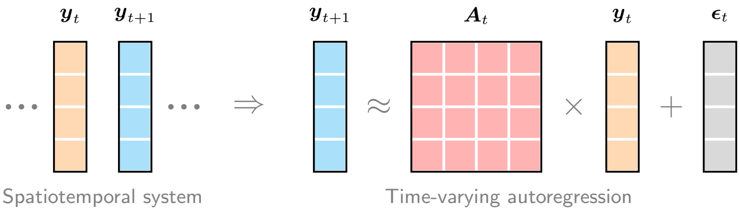
- Chicago rideshare: 96.6M trips (2019) vs. 57.3M trips (2022)



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

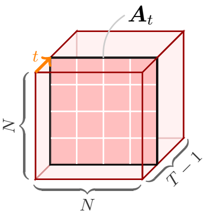
Autoregression

- How to characterize dynamical systems?



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

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



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

1927 — Higher-Order SVD

Frank Lauren Hitchcock

1960s — Tucker Decomposition

Ledyard R. Tucker

1970 — CP Decomposition

J. Douglas Carroll
Jih-Jie Chang
Richard A. Harshman

2009 — Tensor Decompositions and Applications

Tamara G. Kolda

2011 — Tensor-Train Decomposition

Ivan Oseledets

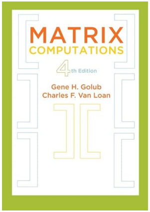
Dynamic Autoregressive Tensor Factorization

- Tensor factorization:

$$\mathcal{A} = \underbrace{\mathcal{G} \times_1 \boldsymbol{W} \times_2 \boldsymbol{V} \times_3 \boldsymbol{X}}_{\text{Tucker decomposition}}$$

$$\Updownarrow$$

$$\boldsymbol{A}_t = \mathcal{G} \times_1 \underbrace{\boldsymbol{W}}_{\text{spatial modes}} \times_2 \boldsymbol{V} \times_3 \underbrace{\boldsymbol{x}_t^\top}_{\text{temporal modes}}$$



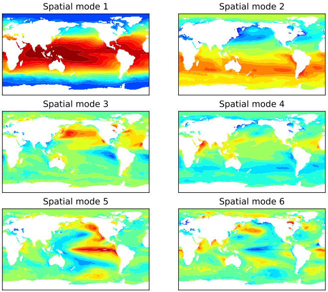
- Optimization problem:

$$\begin{aligned} \min_{\mathcal{G},\boldsymbol{W},\boldsymbol{V},\boldsymbol{X}} \; & \frac{1}{2}\sum_{t\in[T-1]} \left\|\boldsymbol{y}_{t+1} - (\mathcal{G} \times_1 \boldsymbol{W} \times_2 \boldsymbol{V} \times_3 \boldsymbol{x}_t^\top)\boldsymbol{y}_t\right\|_2^2 \\ \text{s.t.} \; & \underbrace{\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_R}_{\text{orthogonal spatial modes}} \end{aligned}$$

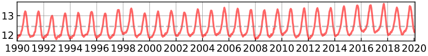
on spatiotemporal data $\boldsymbol{Y} \in \mathbb{R}^{N\times T}$.

Beyond Transport: Sea Surface Temperature

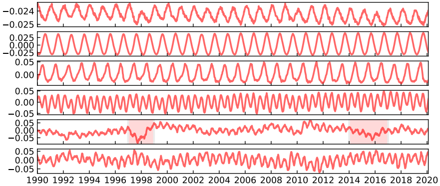
- Sea surface temperature (**SST**) dataset



Spatial modes in $\boldsymbol{W}$



Time series of mean temperature

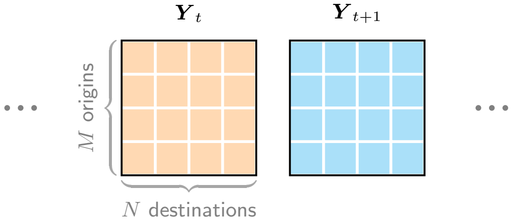


Temporal modes in $\boldsymbol{X}$

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

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

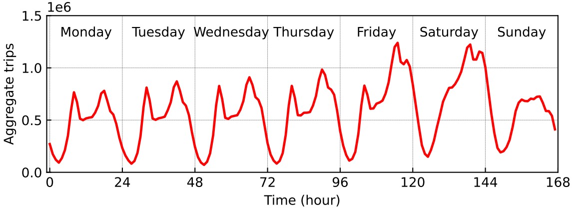
- Optimization problem:

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

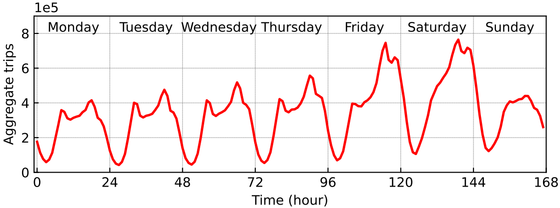
$$\text{s.t.} \quad \underbrace{\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_R}_{\text{orthogonal origin patterns}}$$

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

Pickup trips aggregated over 52 weeks in 2019

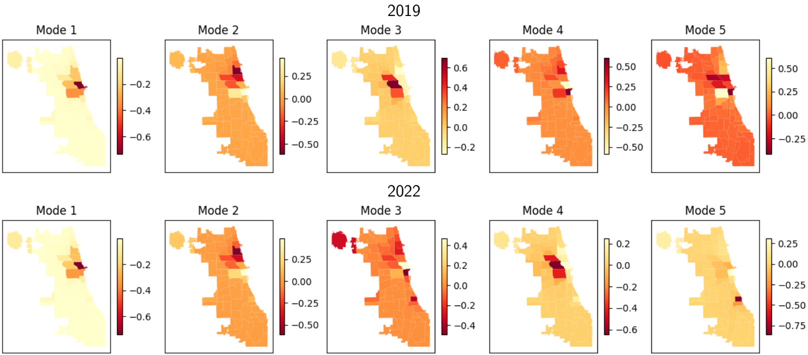


Pickup trips aggregated over 52 weeks in 2022



Human Mobility

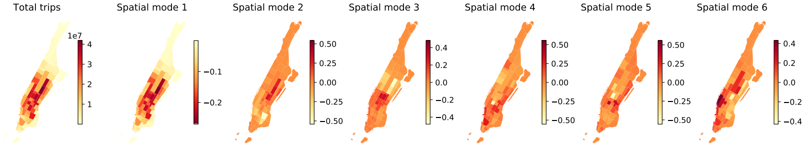
- Rideshare trip data: 77 origins $\times$ 77 destinations $\times$ 168 hours



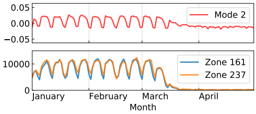
- Identify the changes in pickup zones before/after COVID-19

Human Mobility

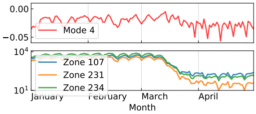
- NYC taxi dataset (pickup)
  - Remarkable decrease of trips due to COVID-19



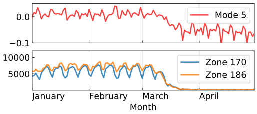
Spatial patterns in $\boldsymbol{W}$



Pattern #2 & taxi trips (2020)



Pattern #4 & taxi trips (2020)

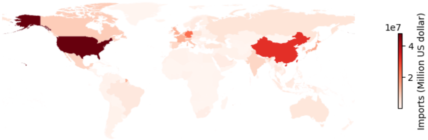


Pattern #5 & taxi trips (2020)

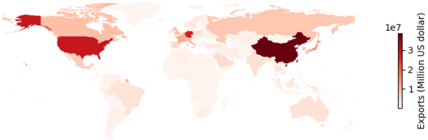
- Identify the changes of temporal patterns due to COVID-19

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

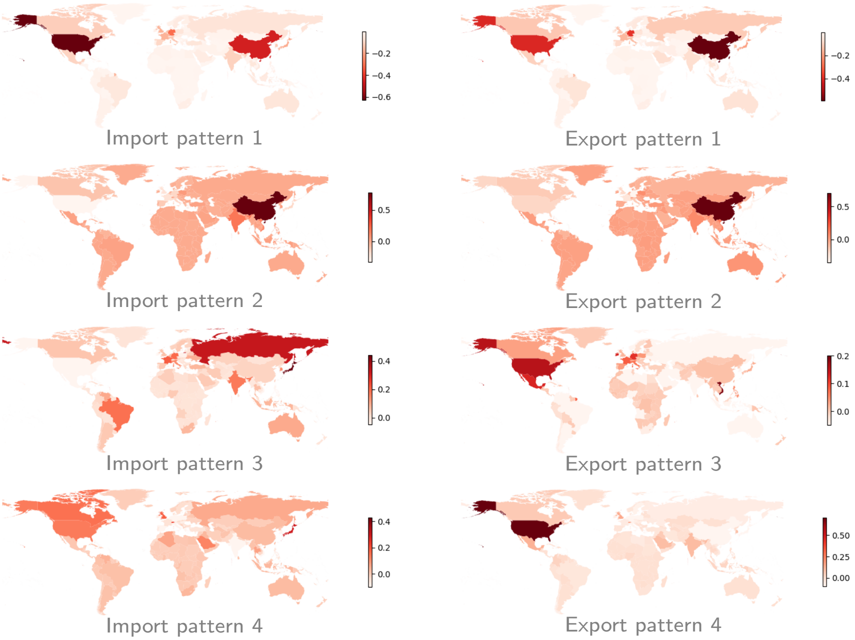


Imports from 2000 to 2022



Exports from 2000 to 2022

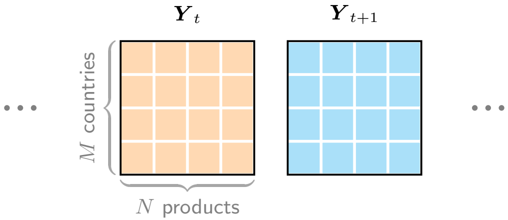
[5] The dataset is available at https://stats.wto.org.



Import pattern 1
−0.2
−0.4
−0.6
Export pattern 1
−0.2
−0.4
Import pattern 2
0.5
0.0
Export pattern 2
0.5
0.0
Import pattern 3
0.4
0.2
0.0
Export pattern 3
0.2
0.1
0.0
Import pattern 4
0.4
0.2
0.0
Export pattern 4
0.50
0.25
0.00

Beyond Transport: International Trade

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



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

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

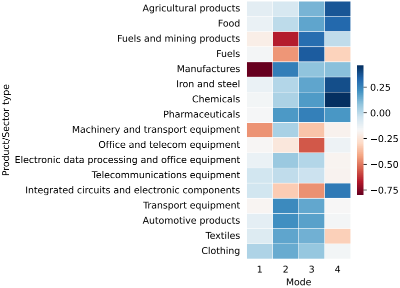
- Optimization problem of DATF:

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

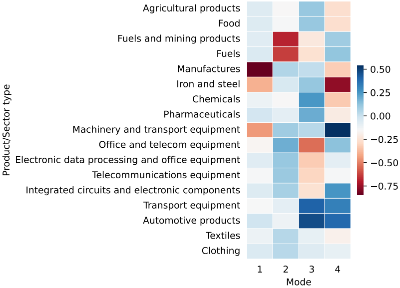
$$\text{s.t.} \quad \underbrace{\boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}_R}_{\text{orthogonal country patterns}}$$

Product Patterns

- On 17 merchandise types



Imports



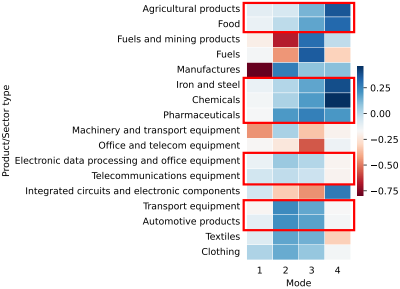
Exports

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

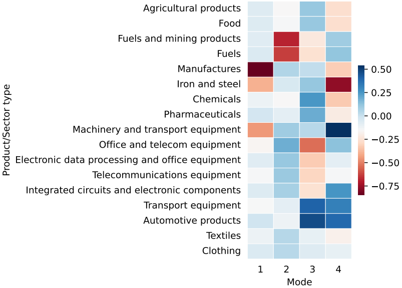
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



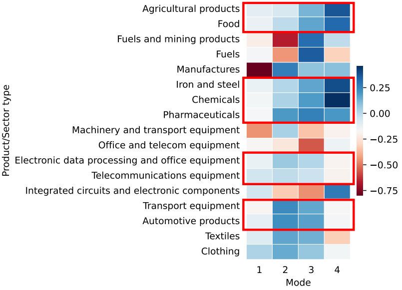
Exports

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

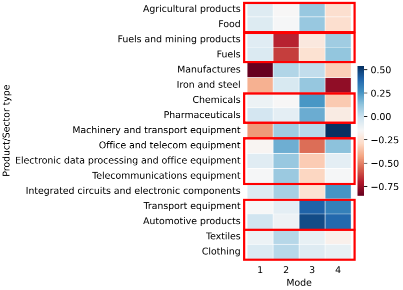
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



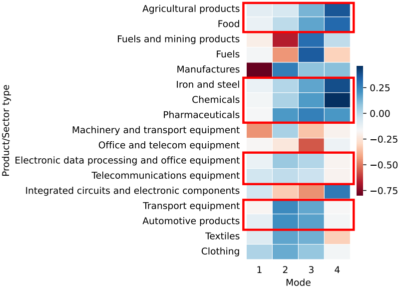
Exports

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

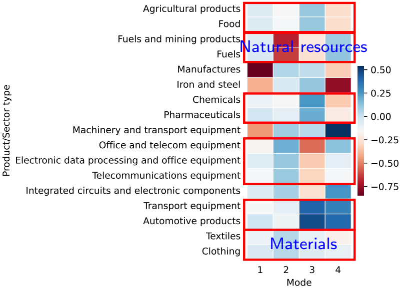
Import: What we buy? (demand) vs. Export: What we sell? (supply)

Product Patterns

- On 17 merchandise types



Imports



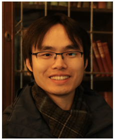
Exports

- 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

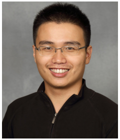
Dingyi Zhuang
MIT CEE

Yunhan Zheng
MIT SMART

Shenhao Wang
UF Urban AI

Jinhua Zhao
MIT DUSP

Ryan Qi Wang
Northeastern CEE

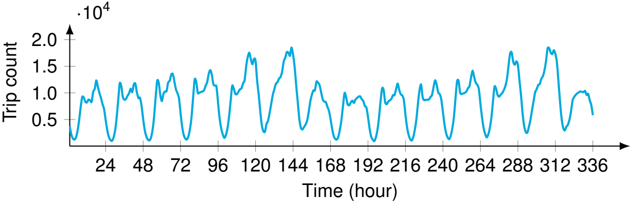


Lijun Ding
UCSD Math

Vassilis Digalakis Jr
HEC Paris

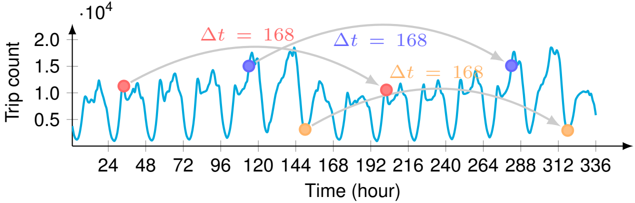
- **Xinyu Chen**, Vassilis Digalakis Jr, Lijun Ding, Jinhua Zhao (2025). "Interpretable Time Series Autoregression". Ready for submission.

Motivation



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

Motivation



Weekly periodicity of rideshare trip time series

What motivate us most about **periodicity**?

❶ **Resilience and stability of systems**: Empirically measure the periodicity and predictability of urban systems.

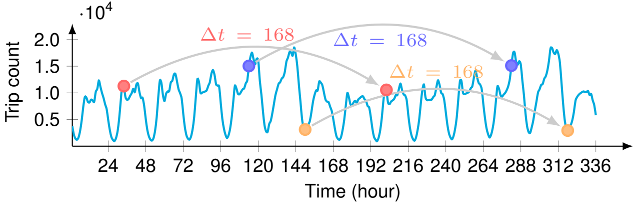
❷ **Optimization of transport systems**: Optimize resources (e.g., public transit, taxi, rideshare, and micromobility) to meet transport demand efficiently.

❸ **Design of sustainable transport & infrastructure**: Implement energy-efficient solutions tailored to peak hours.

Motivation

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

$$\boldsymbol{w} := \arg\min_{\{w_k\}_{k\in[d]}} \sum_{t\in[d+1,T]} \Big(x_t - \sum_{k\in[d]} w_k x_{t-k}\Big)^2$$



Periodicity of rideshare trip time series

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

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

Valorizing Autoregression

- Time series autoregression

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

$$\overset{\tilde{\boldsymbol{x}}}{\begin{bmatrix}x_5\\x_6\\x_7\\x_8\\x_9\end{bmatrix}} \approx \overset{\boldsymbol{A}}{\begin{bmatrix}x_4 & x_3 & x_2 & x_1\\ x_5 & x_4 & x_3 & x_2\\ x_6 & x_5 & x_4 & x_3\\ x_7 & x_6 & x_5 & x_4\\ x_8 & x_7 & x_6 & x_5\end{bmatrix}} \times \overset{\boldsymbol{w}}{\left.\begin{bmatrix}w_1\\w_2\\w_3\\w_4\end{bmatrix}\right\}} \|\boldsymbol{w}\|_0\leq\tau$$

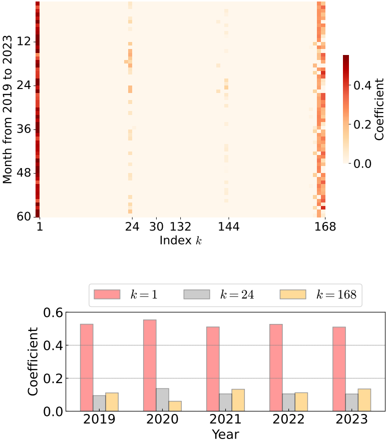
dth-order autoregression on time series $\boldsymbol{x}=(x_1,x_2,\ldots,x_9)^\top$ w/ integer $\tau\in\mathbb{Z}^+$

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

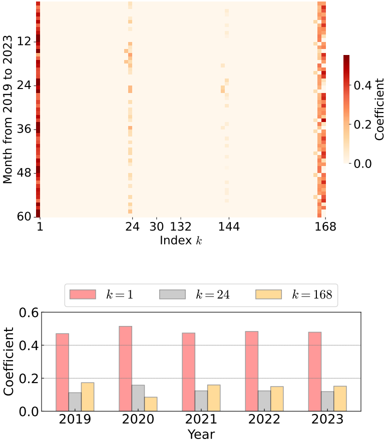
$$\begin{aligned}&\min_{\boldsymbol{w}\geq 0}\ \|\tilde{\boldsymbol{x}}-\boldsymbol{A}\boldsymbol{w}\|_2^2 \\ &\text{s.t.}\quad \underbrace{\|\boldsymbol{w}\|_0\leq\tau}_{\text{Sparsity w/ }\ell_0\text{-norm}}\end{aligned}$$

Envisioning Human Mobility

- NYC rideshare: Sparse coefficient vectors across different months



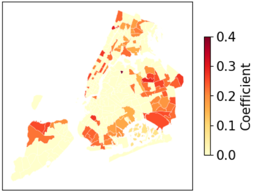
NYC pickup trips



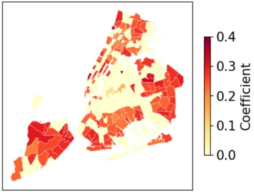
NYC dropoff trips

- Stronger daily periodicity vs. weaker weekly periodicity in 2020

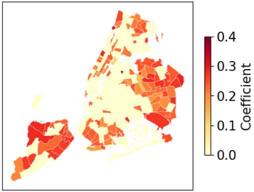
Stronger Daily Periodicity in 2020



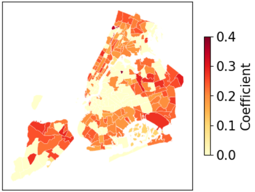
2019



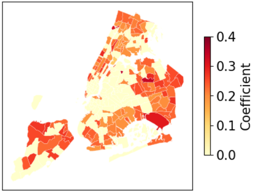
2020



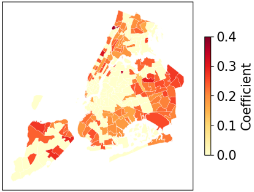
2021



2022



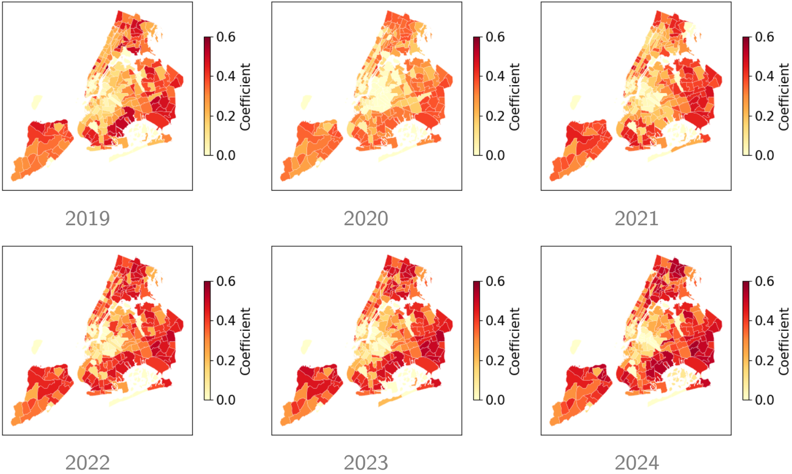
2023



2024

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

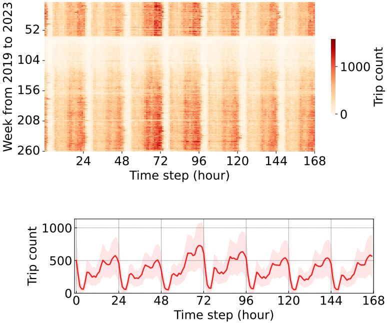
Weaker Weekly Periodicity in 2020



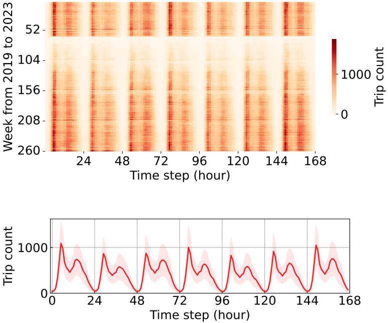
- High-demand areas are less periodic
- Less areas show remarkable weekly periodicity in 2020

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



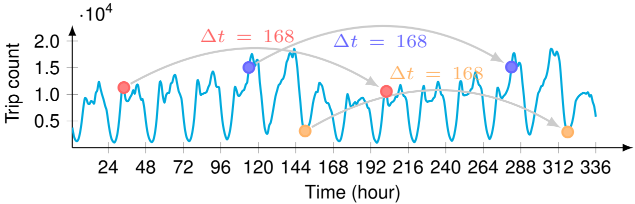
Dropoff trips

- Sparse coefficient vectors:

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

Solution Quality

- Subspace pursuit (SP) sometimes fails



Periodicity of rideshare trip time series

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

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

Time-Varying Transport Systems

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

$$\min_{\{\boldsymbol{w}_\gamma\}_{\gamma\in[\delta]}} \sum_{\gamma\in[\delta]} \|\tilde{\boldsymbol{x}}_\gamma - \boldsymbol{A}_\gamma \boldsymbol{w}_\gamma\|_2^2$$

$$\text{s.t.} \begin{cases} \boldsymbol{w}_\gamma \geq 0 & \text{(non-negativity)} \\ \|\boldsymbol{w}_\gamma\|_0 \leq \tau & \text{(sparsity)} \\ \operatorname{supp}(\boldsymbol{w}_\gamma) = \operatorname{supp}(\boldsymbol{w}_{\gamma+1}) & \text{(no local difference)} \end{cases}$$

making these coefficient vectors comparable across δ months/years.

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

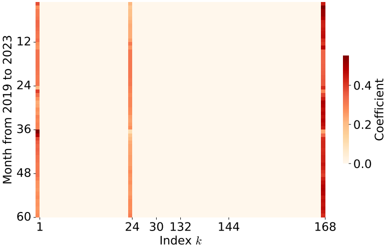
$$\underbrace{0 \leq \boldsymbol{w}_\gamma \leq \boldsymbol{\beta}_\gamma}_{\text{upper bound } \{0,1\}} \qquad \underbrace{\sum_{k\in[d]} \beta_{\gamma,k} \leq \tau}_{\text{sum of binary var.}} \qquad \underbrace{\boldsymbol{\beta}_\gamma - \boldsymbol{\beta}_{\gamma+1} = 0}_{\text{comparability across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

- 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.

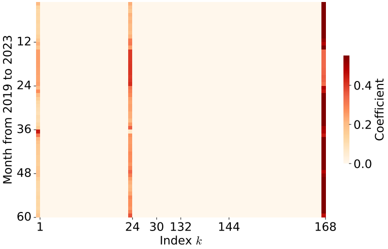
John F. Kennedy International Airport

- Coefficients $\{\boldsymbol{w}_\gamma\}_{\gamma\in[\delta]}$ at $S=\{\underbrace{1}_{\text{local}},\underbrace{24}_{\text{daily}},\underbrace{168}_{\text{weekly}}\}$ across $\delta=60$ months

  ❶ Stronger weekly periodicity of dropoff trips than pickup trips
  ❷ Stronger daily periodicity in 2020
  ❸ Weaker weekly periodicity in 2020



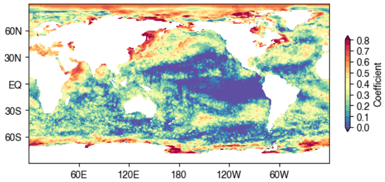
Pickup trips



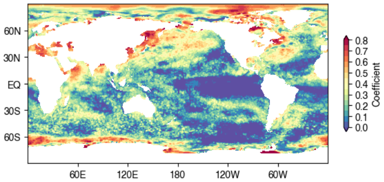
Dropoff trips

- Identify system patterns that evolve over time for human mobility

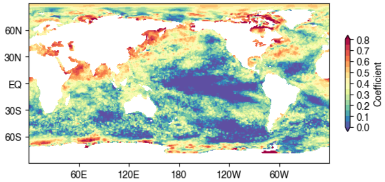
Beyond Transport: Sea Surface Temperature



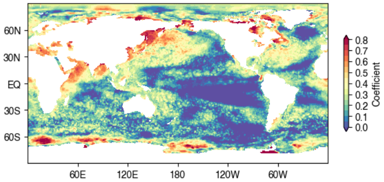
1980s



1990s



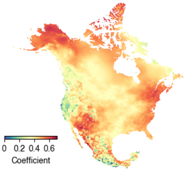
2000s



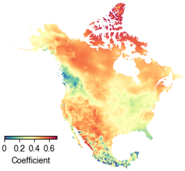
2010s

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

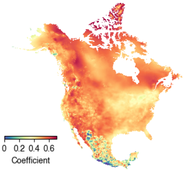
Beyond Transport: North America Temperature



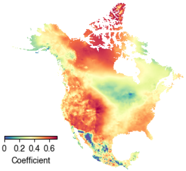
tmin, 1980s



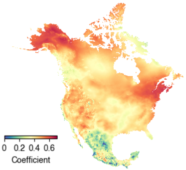
tmin, 1990s



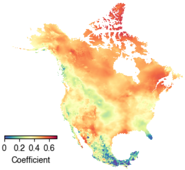
tmin, 2000s



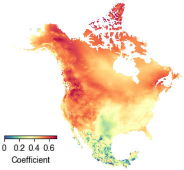
tmin, 2010s



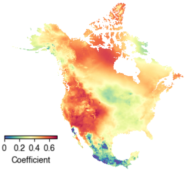
tmax, 1980s



tmax, 1990s



tmax, 2000s



tmax, 2010s

- 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

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