Machine Learning and Optimization for Understanding Spatiotemporal Systems

Time Series Imputation & Periodicity Quantification

Xinyu Chen

Postdoctoral Associate, MIT

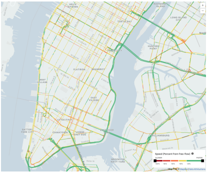
May 22, 2025

Orlando, USA

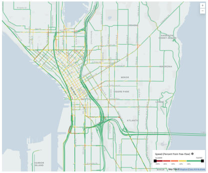
Spatiotemporal Data

- Transport & mobility application scenarios

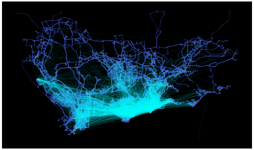
Highway (Portland)



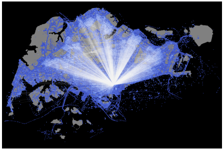
Uber movement (NYC)



Uber movement (Seattle)



Taxi trajectory (Shenzhen)

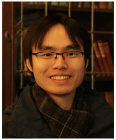


Human movement (Singapore)

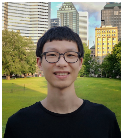
- Challenges: Sparsity, high-dimensionality (network-scale), and multi-dimensionality (complicated data structure), time-varying systems

Spatiotemporal Data Imputation

❒ Convolution ❒ Fast Fourier transform ❒ Optimization w/ ℓ_1-norm
❒ Time series imputation ❒ Speed field reconstruction

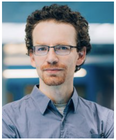


Xinyu Chen
UdeM → MIT



Zhanhong Cheng
McGill → UF

HanQin Cai
UCF



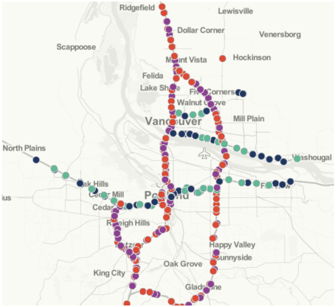
Nicolas Saunier
PolyMtl

Lijun Sun
McGill

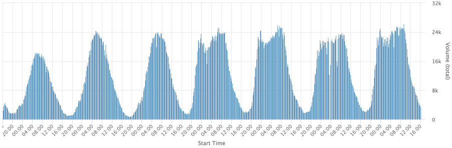
- **Xinyu Chen**, Zhanhong Chen, HanQin Cai, Nicolas Saunier, Lijun Sun (2024). "Laplacian Convolutional Representation for Traffic Time Series Imputation". *IEEE Transactions on Knowledge and Data Engineering*, 36 (11): 6490–6502.
- Blog post: Understanding time series convolution.
  https://spatiotemporal-data.github.io/posts/ts_conv

Motivation

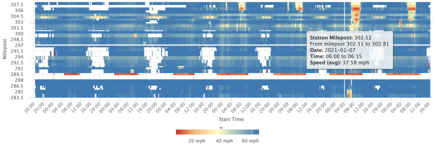
- Portland highway traffic data[1]



Highway network & sensor locations



Traffic volume

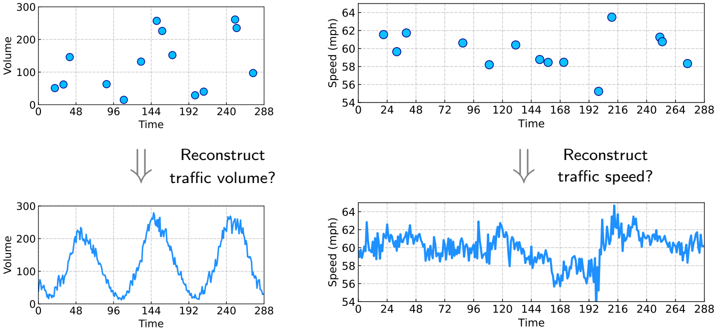


Traffic speed

- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ 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?

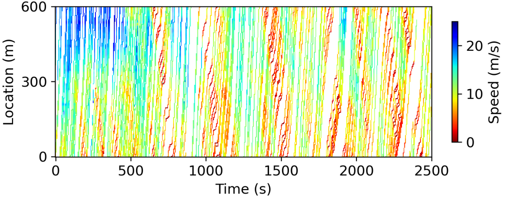
[1] https://portal.its.pdx.edu/home

Motivation

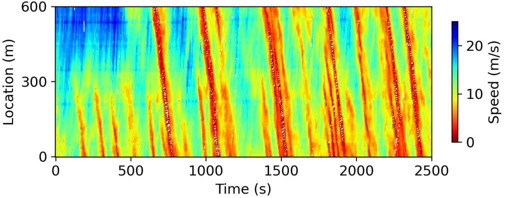


- How to utilize the global trends of traffic time series?
- How to produce local consistency of traffic data?

Motivation



200-by-500 matrix (NGSIM) $\Downarrow$ Reconstruct speed field from 20% sparse trajectories?

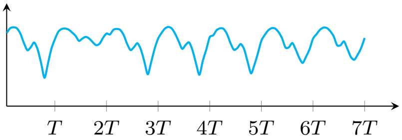


- How to learn from sparse spatiotemporal data?
- How to characterize spatial/temporal local dependencies?

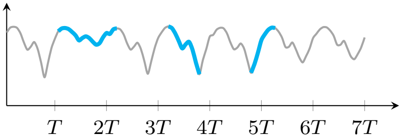
Time Series Imputation

Global/local trends in sparse data?

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

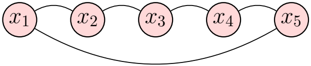


- Local trends (e.g., short-term time series trends):



Local Trend Modeling

- Intuition of Laplacian matrix



Undirected and circulant graph

$\xrightarrow{\text{Modeling}}$

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

(Circulant) Laplacian matrix

- Laplacian kernel:

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

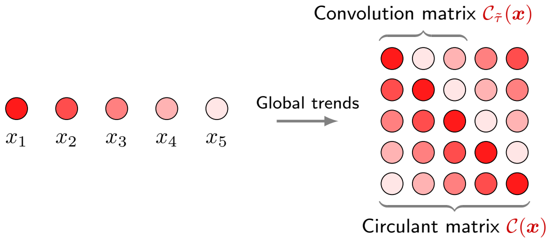
extending to the degree 2τ (i.e., graph connectivity) for $\boldsymbol{x} \in \mathbb{R}^T$.

- 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.

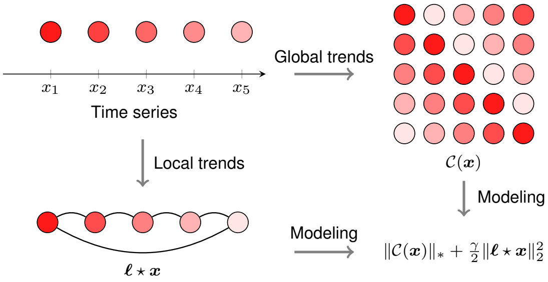
Global + Local Trends?

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

- LCR model:

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

$$\Longrightarrow \quad \begin{aligned} \min_{\boldsymbol{x}} \; & \underbrace{\|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2}_{\text{global + local}} + \underbrace{\frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z} - \boldsymbol{y})\|_2^2}_{\text{regularization}} \\ \text{s.t. } & \boldsymbol{z} = \boldsymbol{x} \end{aligned}$$

"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:

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

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

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

- Reformulate the optimization as ℓ_1-norm minimization:

$$\boldsymbol{x} := \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{w}/\lambda\|_2^2$$

$$\Longrightarrow \hat{\boldsymbol{x}} := \arg\min_{\hat{\boldsymbol{x}}} \ \underbrace{\|\hat{\boldsymbol{x}}\|_1}_{\ell_1\text{-norm}} + \frac{\gamma}{2T}\|\hat{\boldsymbol{\ell}} \circ \hat{\boldsymbol{x}}\|_2^2 + \frac{\lambda}{2T}\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{z}} + \hat{\boldsymbol{w}}/\lambda\|_2^2$$

Laplacian Convolutional Representation

ℓ_1-norm Minimization (Liu & Zhang'23)

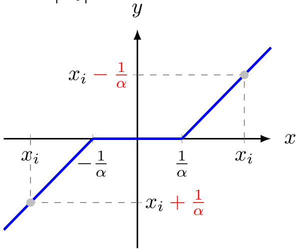
For any $\hat{\boldsymbol{h}} \in \mathbb{C}^T$ and $\delta \in \mathbb{R}$:

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

$$\Downarrow$$

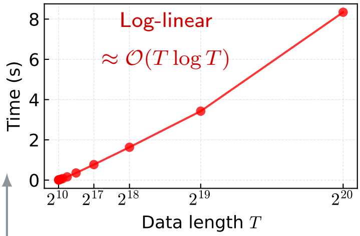
$$\hat{x}_t := \frac{\hat{h}_t}{|\hat{h}_t|} \cdot \underbrace{\max\{0, |\hat{h}_t| - 1/\delta\}}_{\text{shrinkage (e.g., ReLU)}}, t \in [T]$$

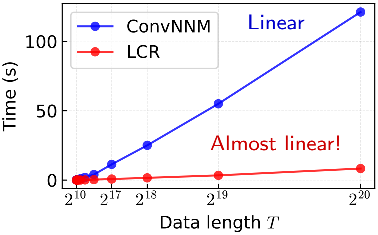
$$y_i = \frac{x_i}{|x_i|} \cdot \max\{|x_i| - 1/\alpha, 0\}$$



Laplacian Convolutional Representation

Time complexity & scalability & efficiency?



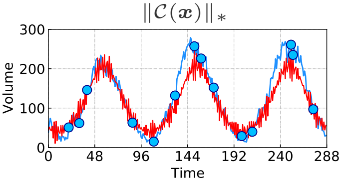


Empirical time complexity

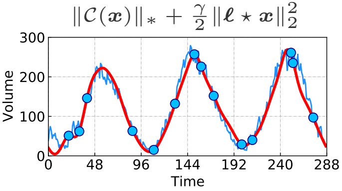
On the synthetic data $\boldsymbol{y} \in \mathbb{R}^T$ with $T \in \{2^{10}, 2^{11}, \dots, 2^{20}\}$

Experiments

- Traffic speed imputation[2] (95% missing rate)



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



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:

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

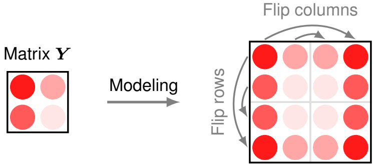
$$\text{s.t.} \quad \|\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq \epsilon$$

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

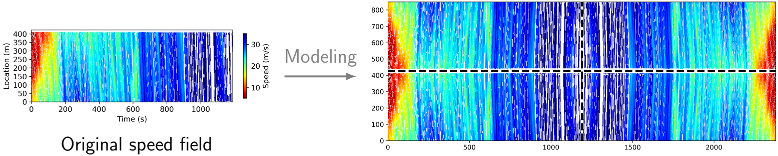
Experiments

Speed field reconstruction[3]

- Flipping operation on a matrix:



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



Original speed field

Flipped speed field

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

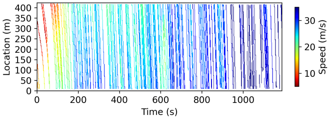
Experiments

Speed field reconstruction in German highways[4]

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

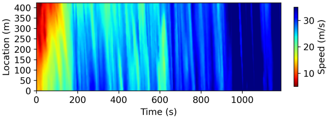
$$\min_{\boldsymbol{X}} \quad \underbrace{\|\mathcal{C}(\boldsymbol{X})\|_*}_{\text{global trend}} + \frac{\gamma}{2} \underbrace{\|(\boldsymbol{\ell}_s \boldsymbol{\ell}^\top) \star \boldsymbol{X}\|_F^2}_{\text{local trend}}$$

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

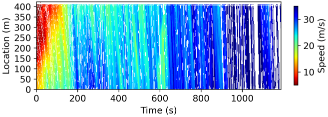


Sparse speed field

LCR-2D



Reconstructed speed field



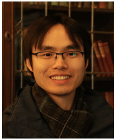
Ground-truth speed field

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

Quantifying Time Series Periodicity

(Ongoing Research)

❒ Interpretable ML ❒ Optimization w/ ℓ_0-norm ❒ Mixed-integer programming
❒ Human mobility regularity ❒ Climate system seasonality



Xinyu Chen
MIT

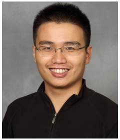
Dingyi Zhuang
MIT

Yunhan Zheng
MIT

Jinhua Zhao
MIT

HanQin Cai
UCF

Ryan Qi Wang
Northeastern

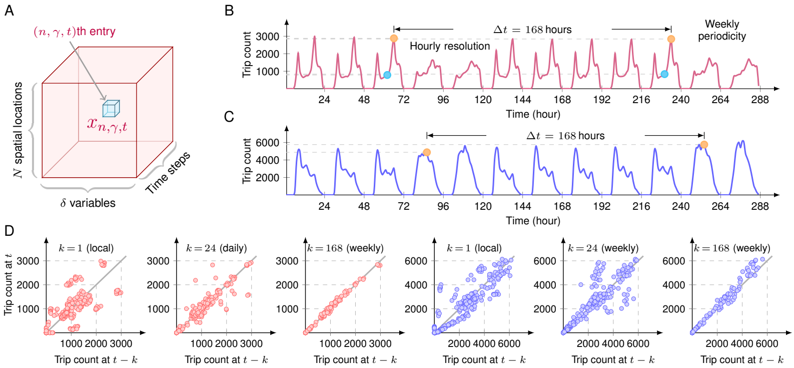


Lijun Ding
UCSD

Vassilis Digalakis Jr
HEC Paris

Motivation

Human mobility data show daily/weekly regularity and periodicity?

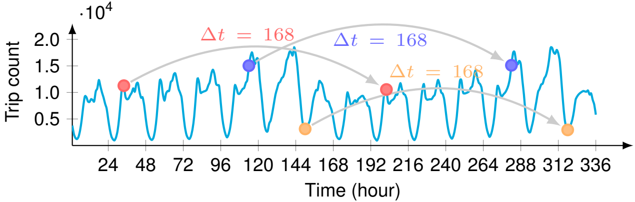


"**Closeness**" to the anti-diagonal $y = x$

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

Motivation

Weekly periodicity of ridesharing trip time series in Chicago



What motivate us most about **periodicity**?

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

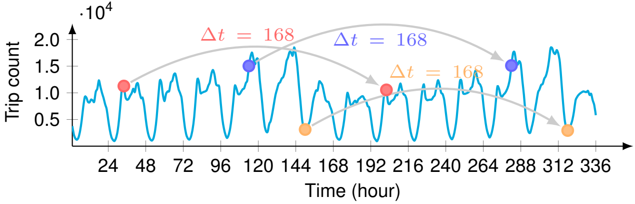
❷ **Optimization of transport systems**: Optimize resources (e.g., public transit, taxi, rideshare, and micromobility) to meet transport demand efficiently.

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

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

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

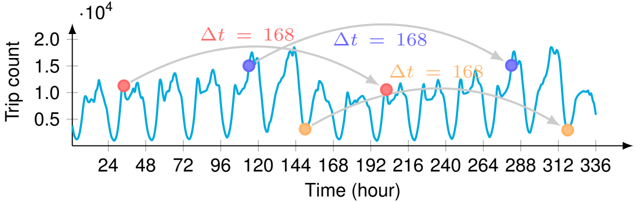
Autoregression on time series $\boldsymbol{x} = (x_1, x_2, \ldots, x_9)^\top$ w/ sparsity $\tau \in \mathbb{Z}^+$

- Sparse autoregression

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

Solution Quality

- Subspace pursuit (SP) sometimes fails



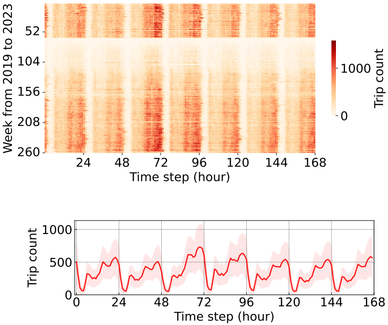
Periodicity of ridesharing trip time series

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

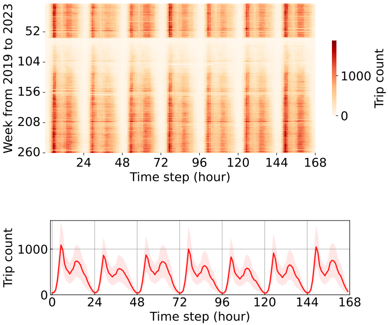
$$\underbrace{\boldsymbol{w} = (\cdots, \underbrace{0.02}_{k=53}, \cdots, \underbrace{0.96}_{k=168})^\top}_{\text{loss func. } = 8.32 \times 10^7 \text{ (SP)}} \quad \text{vs.} \quad \underbrace{\boldsymbol{w} = (\underbrace{0.22}_{k=1}, \cdots, \underbrace{0.77}_{k=168})^\top}_{\text{loss func. } =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

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 $\tau = 3$):

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

Envisioning Human Mobility

Spatially- and Time-Varying 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 $\boldsymbol{\beta}_\gamma \in \{0,1\}^d$:

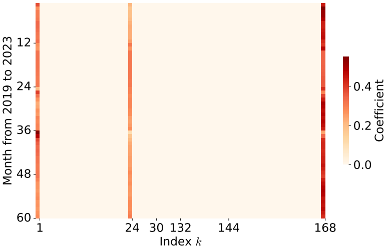
$$\underbrace{0 \leq \boldsymbol{w}_\gamma \leq \boldsymbol{\beta}_\gamma}_{\text{upper bound } \{0,1\}} \qquad \underbrace{\sum_{k\in[d]} \beta_{\gamma,k} \leq \tau}_{\text{sum of binary var.}} \qquad \underbrace{\boldsymbol{\beta}_\gamma - \boldsymbol{\beta}_{\gamma+1} = 0}_{\text{comparability across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

- 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.

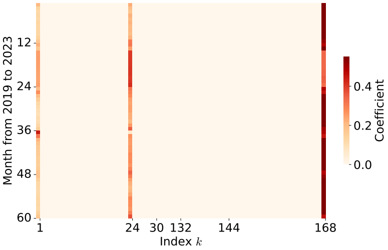
John F. Kennedy International Airport

- 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



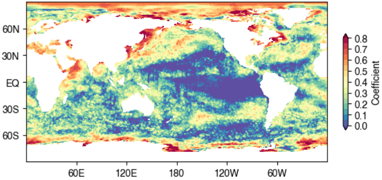
Pickup trips



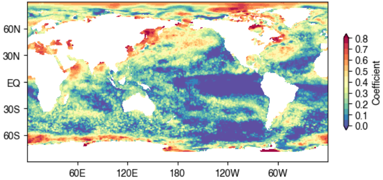
Dropoff trips

- Identify system patterns that evolve over time for human mobility

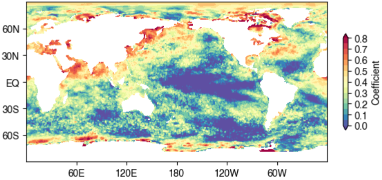
Sea Surface Temperature



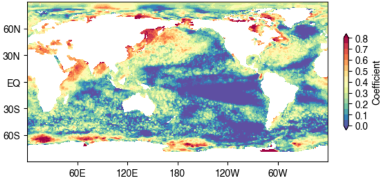
1980s



1990s



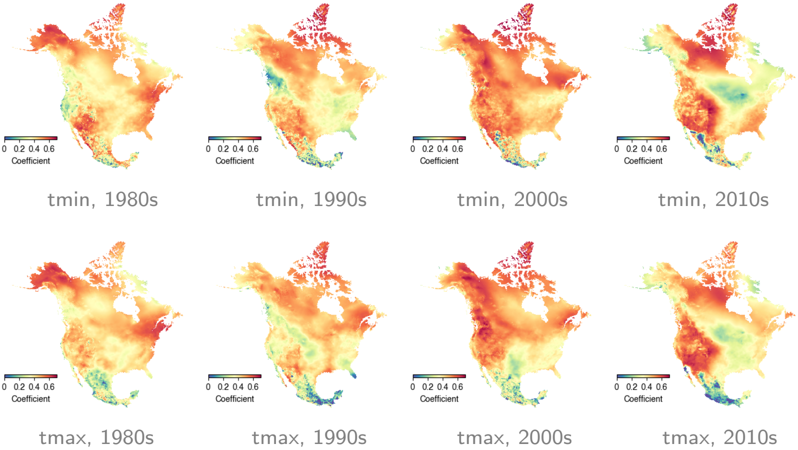
2000s



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

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

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