



MENS
MANUS AND
MACHINA

Quantifying Time Series Periodicity with Interpretable Machine Learning

Urban Human Mobility & Climate Variables

Xinyu Chen

Postdoctoral Associate, MIT

October 13, 2025

Cambridge, USA

Spatiotemporal Data

- Transport & mobility & climate application scenarios



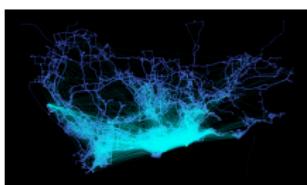
Highway (Portland)



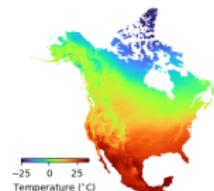
Uber movement (NYC)



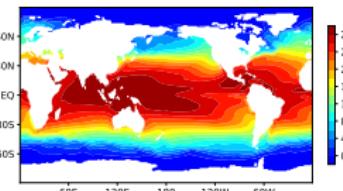
Uber movement (Seattle)



Taxi trajectory (Shenzhen)

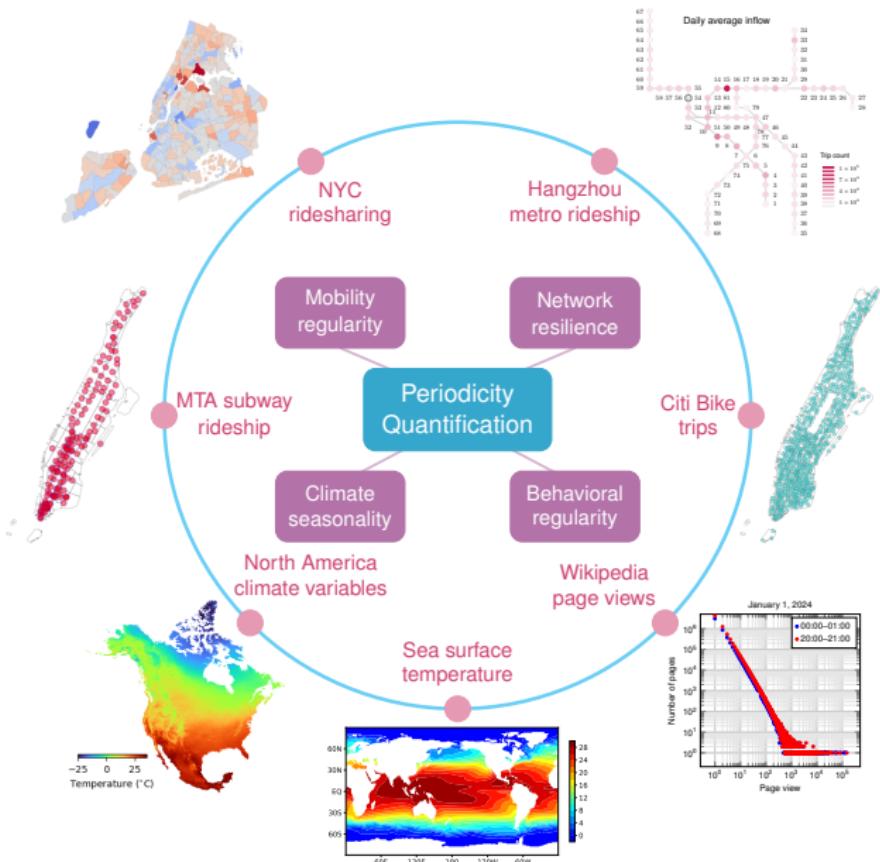


Temperature (NA)



Temperature (sea surface)

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



- Mobility regularity and resilience

Interpretable Time Series Autoregression



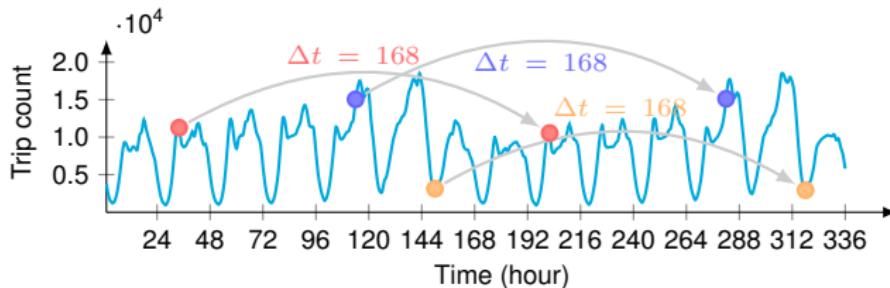
<https://github.com/xinychen/integers>

- Interpretable ML
- ℓ_0 -norm optimization
- Climate system seasonality
- Sparse autoregression
- Mixed-integer optimization
- Human mobility regularity

Valorizing Autoregression

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

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



Periodicity of hourly rideshare trip time series

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

$$\underbrace{\mathbf{w}}_{\text{sparsity } \|\mathbf{w}\|_0 \triangleq 3} = (\underbrace{0.33}_{k=1}, 0, \dots, 0, \underbrace{0.20}_{k=167}, \underbrace{0.46}_{k=168})^\top \in \mathbb{R}^{168}$$

Sparse Autoregression

- Identify the dominant auto-correlations

- $\tau \in \mathbb{Z}^+$: Upper bound of the number of nonzero entries in $w \in \mathbb{R}^d$

$$\tilde{\mathbf{x}} \approx \mathbf{A} \times \mathbf{w}$$
$$\mathbf{w} := \arg \min_{\|\mathbf{w}\|_0 \leq \tau} \sum_{t=d+1}^T \left(\mathbf{x}_t - \sum_{k=1}^d w_k \mathbf{x}_{t-k} \right)^2$$
$$= \arg \min_{\|\mathbf{w}\|_0 \leq \tau} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2$$

- ℓ_0 -norm optimization is NP-hard
- Formulate it as a mixed-integer optimization
 - Introduce binary decision variables $\beta \in \{0, 1\}^d$

$$\begin{array}{ll} \min_w \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 & \min_{\mathbf{w}, \beta} \|\tilde{\mathbf{x}} - \mathbf{A}\mathbf{w}\|_2^2 \\ \text{s.t. } \underbrace{\|\mathbf{w}\|_0 \leq \tau}_{\clubsuit \text{ sparsity of } \mathbf{w}} & \iff \text{s.t. } -\beta \leq \mathbf{w} \leq \beta, \underbrace{\|\beta\|_1 \leq \tau}_{\clubsuit \text{ sparsity of } \beta} \end{array}$$

Sparse Autoregression Done Right

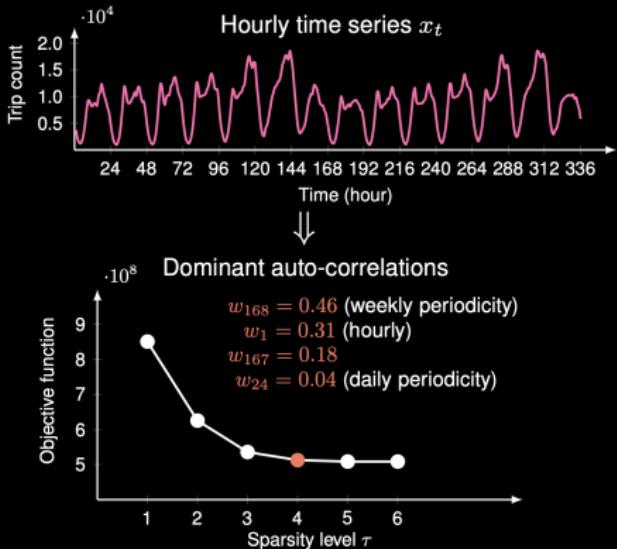
$$\min_{\mathbf{w}, \boldsymbol{\beta}} \underbrace{\sum_{t=d+1}^T \left(x_t - \sum_{k=1}^d w_k x_{t-k} \right)^2}_{\text{Time series autoregression}} \quad \text{s.t. } \underbrace{-\beta_k \leq w_k \leq \beta_k}_{\text{Lower and upper bounds}}, \quad \underbrace{\sum_{k=1}^d \beta_k \leq \tau}_{\clubsuit \text{ Sparsity}} , \quad \underbrace{\beta_k \in \{0, 1\}}_{\text{Binary variable}}$$

- $\mathbf{w} \in \mathbb{R}^d$: Auto-correlations
- $\boldsymbol{\beta} \in \{0, 1\}^d$: Sparsity pattern
- $d = 168$: Autoregression order

```

1 import numpy as np
2 from docplex.mp.model import Model
3
4 def sparse_ar(x, d, tau):
5     model = Model('Sparse Autoregression')
6     T = x.shape[0]
7     w = [model.continuous_var(name = f'w_{k}') for k in range(d)]
8     beta = [model.binary_var(name = f'beta_{k}') for k in range(d)]
9     model.minimize(model.sum((x[t] - model.sum(w[k] * x[t - k - 1]
10                                for k in range(d))) ** 2
11                                for t in range(d, T)))
12     model.add_constraint(model.sum(beta[k] for k in range(d)) <= tau)
13     for k in range(d):
14         model.add_constraint(w[k] <= beta[k])
15         model.add_constraint(w[k] >= -beta[k])
16     solution = model.solve()
17     return np.array(solution.get_values(w))

```



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

Climate System Seasonality Patterns

(arXiv:2506.22895)

- North America temperature/precipitation Sea surface temperature
- Human mobility regularity Climate system seasonality



Xinyu Chen
Vassilis Digalakis Jr
MIT



Lijun Ding
UCSD

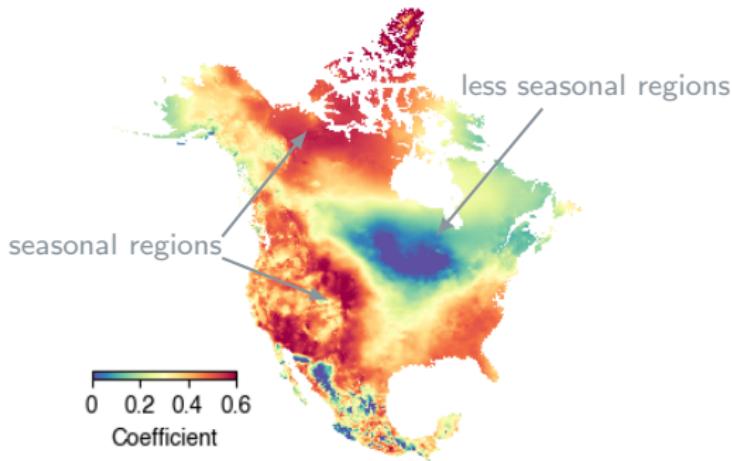


Dingyi Zhuang
MIT



Jinhua Zhao
MIT

Motivation

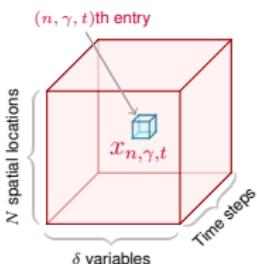


Yearly temperature **seasonality** pattern in 2010s

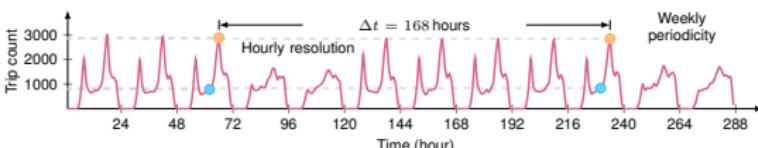
Motivation

Human mobility data show daily/weekly regularity and periodicity?

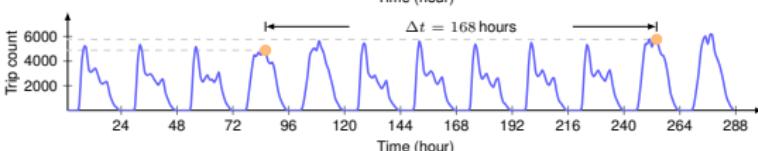
A



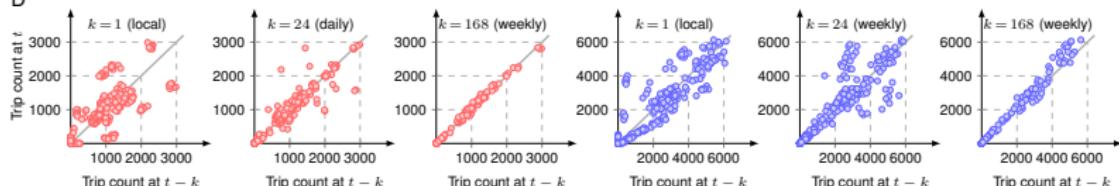
B



C



D

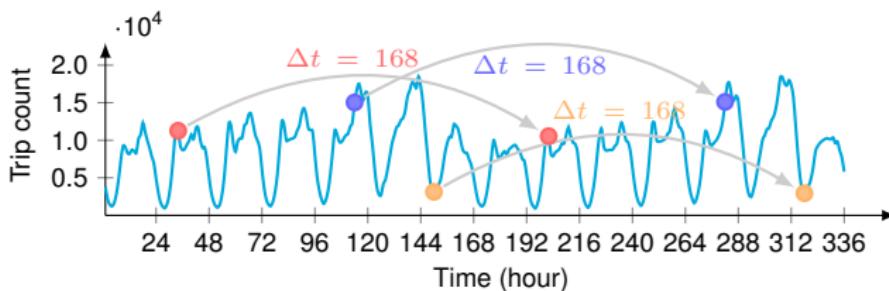


“Closeness” to the
anti-diagonal $y = x$

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

Motivation

Weekly periodicity of ridesharing trip time series in Chicago

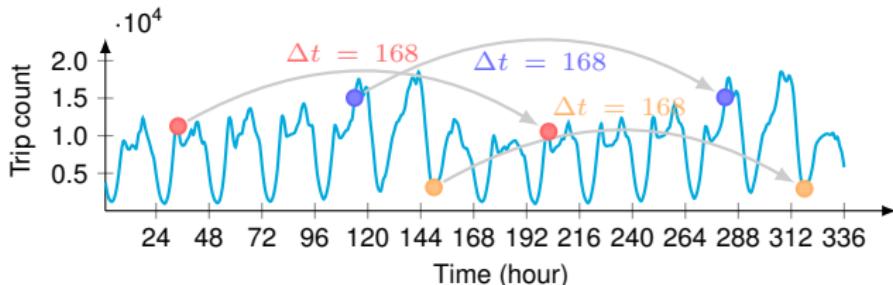


What motivate us most about periodicity?

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

Solution Quality

- Subspace pursuit (SP) sometimes fails



Periodicity of ridesharing trip time series

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

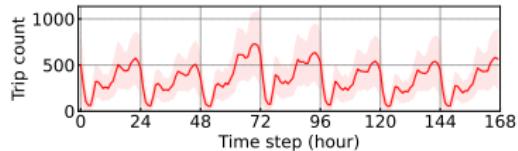
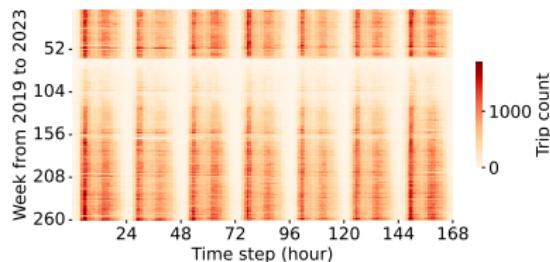
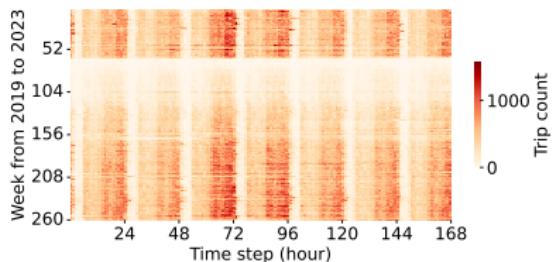
$$\underbrace{\boldsymbol{w} = (\dots, \underbrace{0.02}_{k=53}, \dots, \underbrace{0.96}_{k=168})^\top}_{\text{loss func. } = 8.32 \times 10^7 \text{ (SP)}} \quad \text{vs.} \quad \underbrace{\boldsymbol{w} = (\underbrace{0.22}_{k=1}, \dots, \underbrace{0.77}_{k=168})^\top}_{\text{loss func. } = 6.25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

John F. Kennedy International Airport

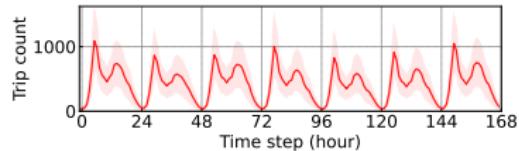
No free lunch

John F. Kennedy International Airport

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



Pickup trips from airport



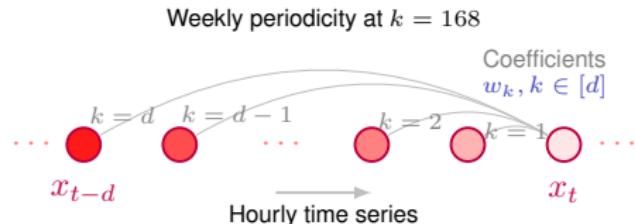
Dropoff trips to airport

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

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

Spatially- and Time-Varying Autoregression

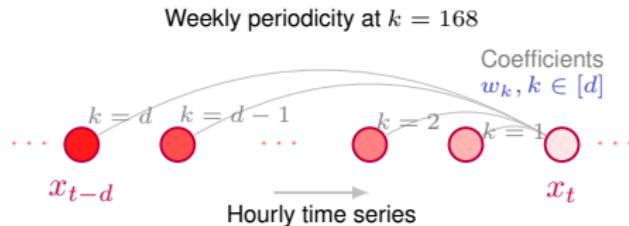
Univariate autoregression



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

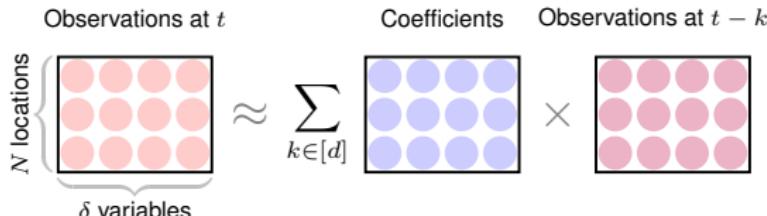
Spatially- and Time-Varying Autoregression

Univariate autoregression



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

Multidimensional autoregression



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

Envisioning Human Mobility

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

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

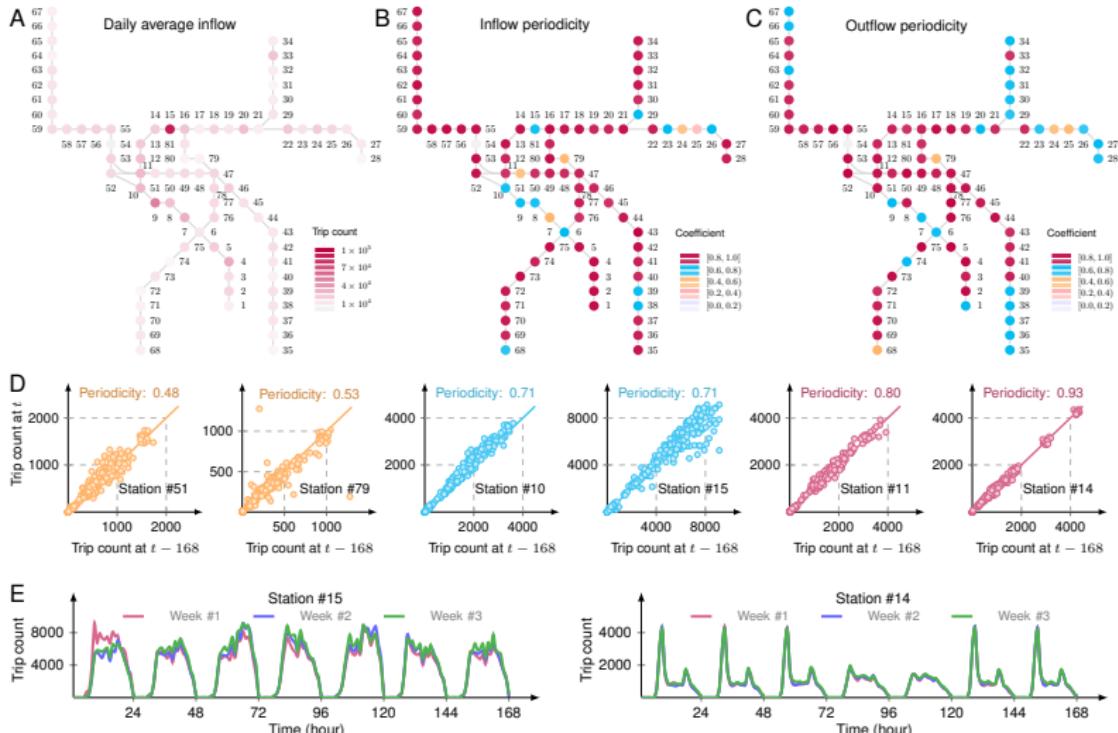
year
spatial location hourly time step

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

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

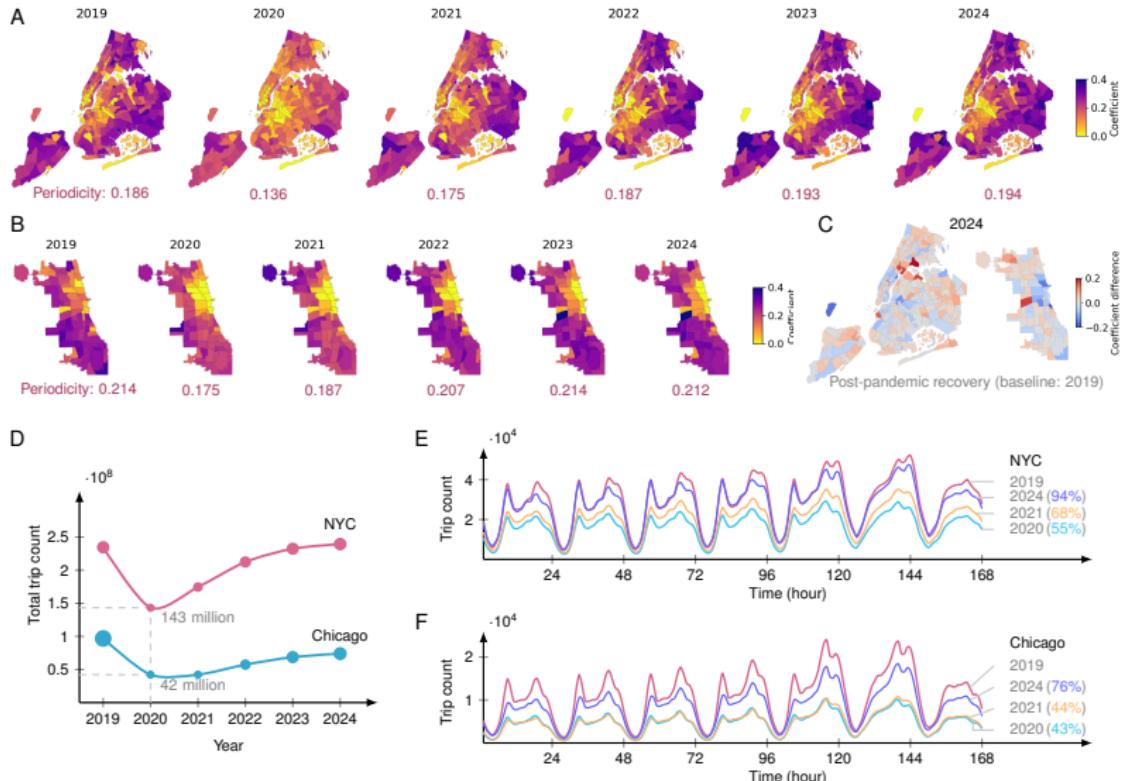
Quantify weekly periodicity by $\{w_{n,\gamma,k}\}$ at index $k = 168$

Envisioning Human Mobility



Hangzhou metro passenger flow in January 2019

Envisioning Human Mobility



Weekly periodicity reveals spatial patterns of ridesharing systems

Understanding Climate Systems

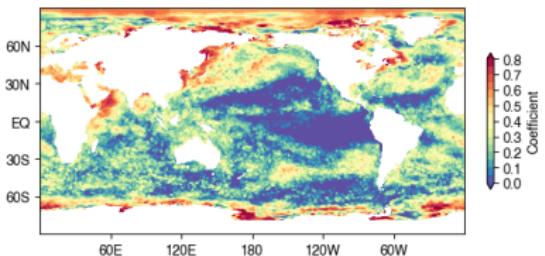
Quantify yearly seasonality by $\{w_{m,n,\gamma,k}\}$ at index $k = 12$

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

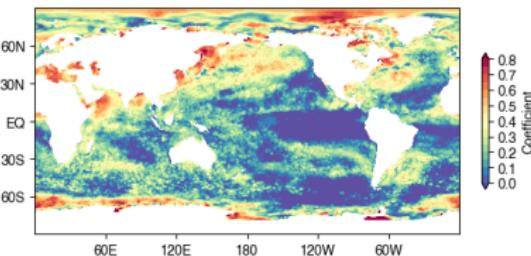
s.t.
$$\begin{cases} \boldsymbol{\beta} \in \{0, 1\}^d & \text{binary decision var.} \\ 0 \leq \boldsymbol{w}_{m,n,\gamma} \leq \boldsymbol{\beta}, \forall m, n, \gamma \\ \|\boldsymbol{\beta}\|_1 \leq \tau \\ \|\boldsymbol{w}_{m,n,\gamma}\|_1 = 1, \forall m, n, \gamma \end{cases}$$

ℓ_1 -normalization

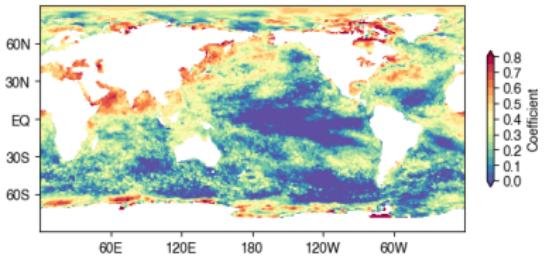
Sea Surface Temperature



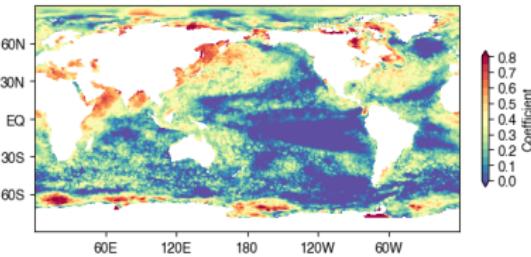
1980s



1990s



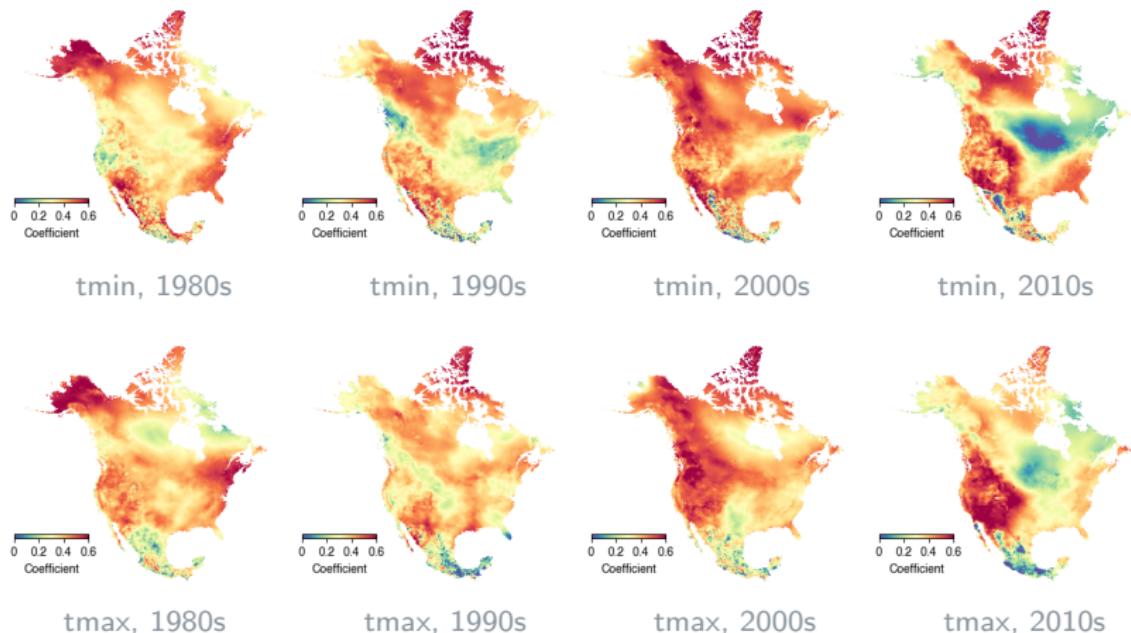
2000s



2010s

- Identify yearly periodicity at $k = 12$ from SST data ($\tau = 3$)
 - ❶ The areas of El Niño events are less seasonal/predictable
 - ❷ Arctic becomes less seasonal/predictable in the past 20 years
- Insights into climate system monitoring

North America Temperature



- Identify yearly periodicity at $k = 12$ from temperature data ($\tau = 3$)
 - 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

Future Work

- Quantifying Behavioral Regularity of Wikipedia



MENS
MANUS AND
MACHINA

Thanks for your attention!

Any Questions?

Slides: https://xinychen.github.io/slides/ml_opt_stsystem.pdf

Essential AR: https://xinychen.github.io/slides/essential_ar.pdf

About me:

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