



# **Future Studies with A Focus on AI for Urban Science**

**Xinyu Chen**

April 19, 2024

# Outline

---

- Land Transport Authority (LTA) proposals<sup>1</sup>
- M3S projects (e.g., spatiotemporal mobility networks)
  - Process urban datasets in Singapore
  - Formulate the problem with machine learning
- DOE projects (TBD)
- Spatiotemporal data modeling<sup>2</sup>
- Current work: Dynamic autoregressive tensor factorization for pattern discovery of spatiotemporal systems (Finished 50%+)
  - Advance the prior art (e.g., dynamic mode decomposition, time-varying autoregression)
  - Use the orthogonal Procrustes rotation to find orthogonal patterns
  - Well-suited to multidimensional dynamical systems (e.g., mobility with dimensions {origin, destination, time})
- Future work: Deep sequence models (TBD)
- Assist other people in the lab and their research
- Present results at INFORMS/TRB meetings

---

<sup>1</sup>[https://www.dropbox.com/home/SMART\\_T4\\_T5\\_HQ](https://www.dropbox.com/home/SMART_T4_T5_HQ)

<sup>2</sup><https://spatiotemporal-data.github.io>

# LTA Proposals

---

A quick look...

- By Yunhan:
  1. Understanding the usage of electric vehicle charging stations and estimating their impacts on local economic vitality
  2. Learning dynamic activity and travel decisions in the post-pandemic era
- By Xinyu:
  3. Identifying travel modes and commuter behavior patterns with machine learning
  4. Anomaly detection and vehicular monitoring on trajectories of autonomous vehicles in future human mobility
- Dingyi:
  5. Analyzing pedestrian jaywalking and vehicle interactions in the age of autonomous vehicles
  6. Uncertainty quantification in spatiotemporal prediction of walk-cycle rides demand

# M3S Project

---

Main tasks:

- Pre-process the urban datasets (e.g., Veraset) in Singapore<sup>3</sup>
- Define the scientific questions in the project
- Formulate the problems with machine learning
- Analyze the results and their impacts

Current ideas: Discovering dynamics of urban human activity with dynamic autoregressive tensor factorization

- (On 2D activity data) Uncover spatial modes/patterns (e.g., POI patterns)
- (On 3D mobility data) Uncover temporal modes/patterns (e.g., long-term changing behavior impacted by special events and policy)

---

<sup>3</sup><https://spatiotemporal-data.github.io/trajectory/veraset/>

# Spatiotemporal Data Modeling

---

## Goals:

- Supporting research for all aspects of spatiotemporal data modeling with machine learning
- Solving many scientific, mathematical, industrial, and engineering problems in:
  - Urban Science
  - Human Mobility Modeling
  - Geospatial Data Analysis
  - Intelligent & Sustainable Urban Systems
  - Optimization & Decision Making

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

# Spatiotemporal Data Modeling

## Plan?

- Coding and computing with data
- Posting scientific questions
- Supporting open-source and reproducible research

### Matching Taxi Trips with Community Areas

There are three basic steps to follow for processing taxi trip data:

- Download taxi trips in 2022 in the .csv format, e.g., `taxi_trips_2022.csv`.
- Use the `pandas` package in Python to process the raw trip data.
- Match trip pickup/dropoff locations with boundaries of the community area.

```
import pandas as pd

data = pd.read_csv('taxi_trips_2022.csv')
data.head()
```

For each taxi trip, one can select some important information:

- **Trip Start Timestamp:** When the trip started, rounded to the nearest 15 minutes.
- **Trip Seconds:** Time of the trip in seconds.
- **Trip Miles:** Distance of the trip in miles.
- **Pickup Community Area:** The Community Area where the trip began. This column will be blank for locations outside Chicago.
- **Dropoff Community Area:** The Community Area where the trip ended. This column will be blank for locations outside Chicago.

```
df = pd.DataFrame()

df['Trip Start Timestamp'] = data['Trip Start Timestamp']
df['Trip Seconds'] = data['Trip Seconds']
df['Trip Miles'] = data['Trip Miles']
df['Pickup Community Area'] = data['Pickup Community Area']
df['Dropoff Community Area'] = data['Dropoff Community Area']

del data
```

Figure 2 shows taxi pickup and dropoff trips (2022) on 77 community areas in the City of Chicago. Note that the average trip duration is **1209.75 seconds** and the average trip distance is **6.16 miles**.

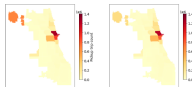


Figure 2. Taxi pickup and dropoff trips (2022) in the City of Chicago, USA. There are 4,763,961 remaining trips after the data processing.

For comparison, Figure 3 shows taxi pickup and dropoff trips (2018) on 77 community areas in the City of Chicago. Note that the average trip duration is **915.62 seconds** and the average trip distance is **3.93 miles**.

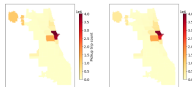


Figure 3. Taxi pickup and dropoff trips (2018) in the City of Chicago, USA. There are 12,484,572 remaining trips after the data processing. See the [data processing codes](#).

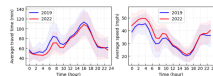


Figure 6. Average travel time and speed from area 32 (e.g., Downtown) to area 76 (e.g., Airport) in both 2019 and 2022.

```
import numpy as np
import matplotlib.pyplot as plt

fig, axs = plt.subplots(2, 1)
ax = fig.add_subplot(1, 2, 1)

# Average travel time in 2019
ax1 = df.groupby(['hour'])['Trip Seconds'].mean().values / 30
ax1 = df.groupby(['hour'])['Trip Seconds'].std().values / 30
plt.plot(ax1, color = 'blue', linewidth = 1.5, label = '2019')
upper = ax1 + 1
lower = ax1 - 1

# Average travel time in 2022
ax2 = df.groupby(['hour'])['Trip Seconds'].mean().values / 30
ax2 = df.groupby(['hour'])['Trip Seconds'].std().values / 30
plt.plot(ax2, color = 'red', linewidth = 1.5, label = '2022')
upper = ax2 + 1
lower = ax2 - 1

ax.set_xlabel('Time (hour)')
ax.set_ylabel('Average travel time (min)')
ax.legend()

fig.savefig('fig6.png', dpi=300)
```

Source: <https://spatiotemporal-data.github.io/Chicago-mobility/taxi-data>

# Current Work

---

Check out the Overleaf project

Past posts:

- Time-varying autoregression:  
[https://spatiotemporal-data.github.io/posts/time\\_varying\\_model/](https://spatiotemporal-data.github.io/posts/time_varying_model/)
- Orthogonal time-varying autoregression:  
<https://spatiotemporal-data.github.io/probs/orth-var/>

Past work:

- X. Chen, C. Zhang, X. Chen, N. Saunier, L. Sun (2024). Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression. IEEE Transactions on Knowledge and Data Engineering. 36 (2): 504-517.

# Future Work

## State-Space Model (SSM)

- **State transition equation:**

$$\underbrace{x_{t+1}}_{\text{state}} = A x_t + B \underbrace{u_t}_{\text{input}} + \underbrace{w_t}_{\mathcal{N}(0, I)}$$

- **Observation equation:**

$$\underbrace{y_t}_{\text{output}} = C x_t + D \underbrace{u_t}_{\text{input}} + \underbrace{v_t}_{\mathcal{N}(0, I)}$$

VS.

## Recurrent Neural Network (RNN)

- **Hidden state update:**

$$h_t = \text{activation}(W_{hh} h_{t-1} + W_{hx} x_t + \underbrace{b_h}_{\text{bias}})$$

- **Output:**

$$\underbrace{y_t}_{\text{output}} = \text{activation}(W_{yh} \underbrace{h_t}_{\text{state}} + \underbrace{b_y}_{\text{bias}})$$

- Long-range spatiotemporal modeling, e.g.,
  - State-space layers ([Smith et al.'22](#))
  - Convolutional SSM ([Smith et al.'23](#))
- Linear-time sequence modeling (e.g., selective SSM, [Gu & Dao'23](#))



## References

---

A short list:

- [Cai et al.'14] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, “Data-driven tight frame construction and image denoising,” *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.
- [Calamai & More'87] P. H. Calamai and J. J. More, “Projected gradient methods for linearly constrained problems,” *Mathematical programming*, vol. 39, no. 1, pp. 93–116, 1987.
- [Golub & Van Loan'13] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.
- [Gu & Dao'23] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces.” *arXiv preprint arXiv:2312.00752*.
- [Smith et al.'22] J. T. H. Smith, A. Warrington, and S. W. Linderman, “Simplified state space layers for sequence modeling.” *arXiv preprint arXiv:2208.04933*. (ICLR'23)
- [Smith et al.'23] J. T. H. Smith, S. De Mello, J. Kautz, S. W. Linderman, and W. Byeon, “Convolutional state space models for long-range spatiotemporal modeling.” *arXiv preprint arXiv:2310.19694*.



# Thanks for your attention!

## Any Questions?

### About me:

- 🏠 Homepage: <https://xinychen.github.io>
- 🐙 GitHub: <https://github.com/xinychen>
- ✉ How to reach me: [chenxy346@gmail.com](mailto:chenxy346@gmail.com)