



**POLYTECHNIQUE  
MONTRÉAL**

UNIVERSITÉ  
D'INGÉNIERIE

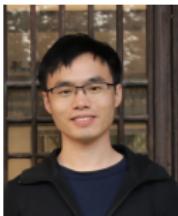


# Spatiotemporal Traffic Data Imputation and Forecasting with Tensor Learning

Ph.D. Research Project

**Xinyu Chen**

May 24, 2022



**Ph.D. candidate**

Xinyu Chen  
Polytechnique Montréal



**Supervisor**

Prof. Nicolas Saunier  
Polytechnique Montréal



**Co-supervisor**

Prof. Lijun Sun  
McGill University

# Outline

- **Motivation**

- Multivariate traffic time series
- Multidimensional traffic time series
- Multiple data behaviors

- **Literature Review**

- Spatiotemporal traffic data imputation
- Spatiotemporal traffic forecasting
- Low-rank tensor learning

- **Objective A**

- Spatiotemporal traffic data imputation

- **Objective B**

- High-dimensional traffic forecasting
- Multidimensional traffic forecasting

- **Objective C**

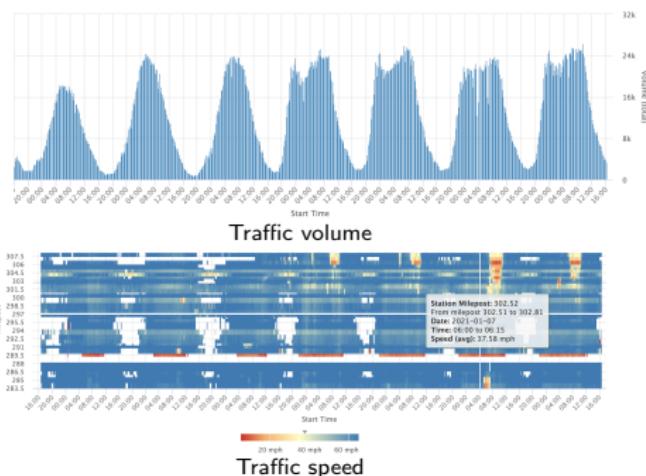
- Multivariate traffic forecasting on sparse data
- Multidimensional traffic forecasting on sparse data

- **Conclusion**

# Multivariate Traffic Time Series

Many spatiotemporal traffic time series data are in the form of **matrix**.

- Example: Portland highway traffic data<sup>1</sup>.



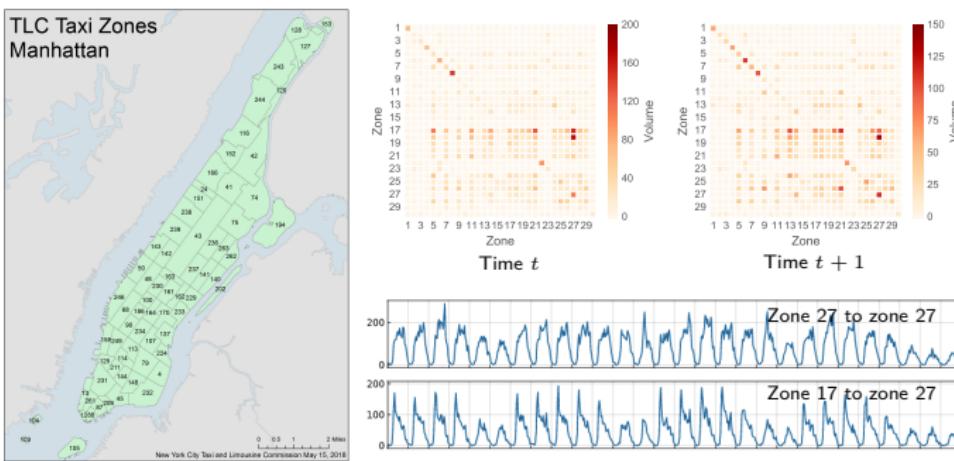
- $X \in \mathbb{R}^{N \times T}$  with  $N$  spatial locations  $\times T$  time steps

<sup>1</sup><https://portal.its.pdx.edu/home>

# Multidimensional Traffic Time Series

Many spatiotemporal traffic time series data are in the form of **tensor**.

- Example: NYC (hourly) taxi flow data<sup>2</sup>.



- $\mathcal{X} \in \mathbb{R}^{M \times N \times T}$  with  $M$  zones  $\times$   $N$  zones  $\times$   $T$  time steps

<sup>2</sup><https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

## Multiple Data Behaviors

---

Spatiotemporal traffic data are time series, but they involve multiple data behaviors.

- Incompleteness & sparsity
- High-dimensionality
- Multidimensionality
- Noises & outliers
- Time-varying behavior
- Nonstationarity
- .....

In addition, spatiotemporal correlations are also very important.

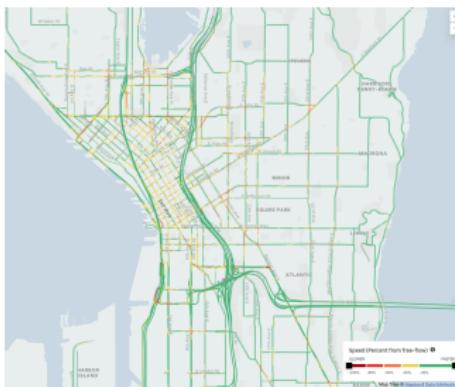
# Multiple Data Behaviors

## Sparsity & high-dimensionality

- Uber (hourly) movement speed data<sup>3</sup>



NYC movement



Seattle movement

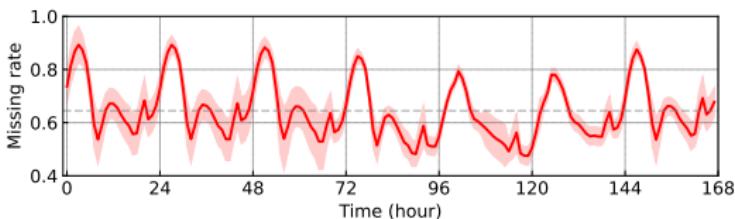
- The average speed on a given road segment for each hour of each day.
- Hourly speeds are computed when road segments have 5+ unique trips.
- **Issue:** insufficient sampling of ridesharing vehicles on the road network.

<sup>3</sup><https://movement.uber.com/>

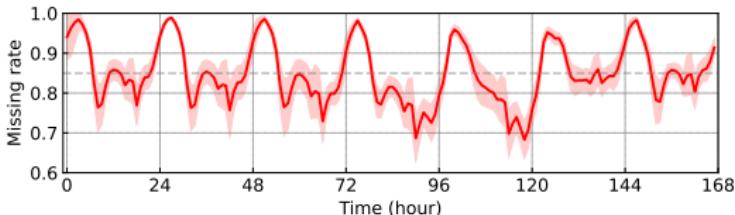
# Multiple Data Behaviors

## Sparsity & high-dimensionality

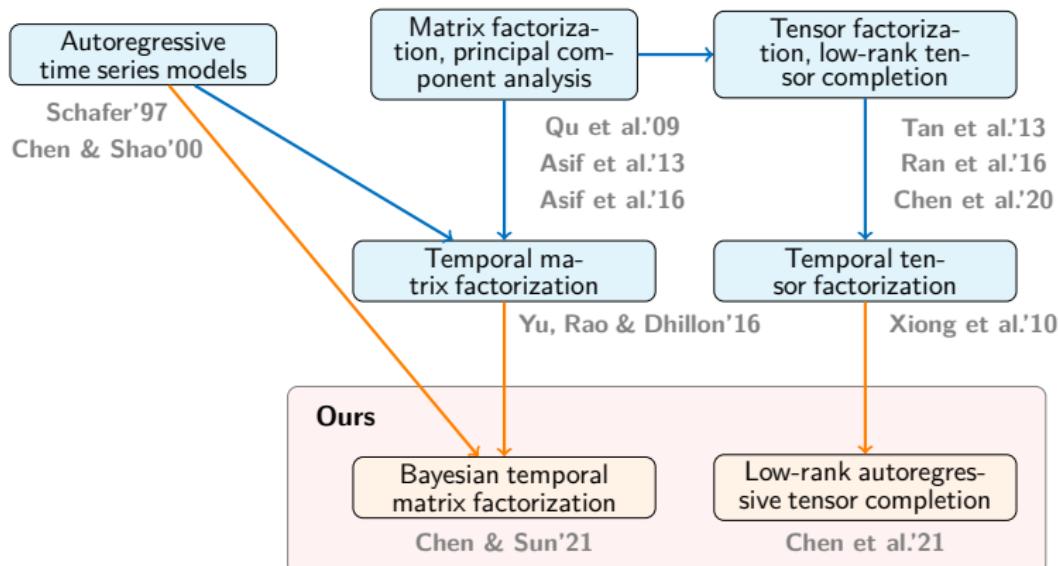
- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Overall missing rate: **64.43%**



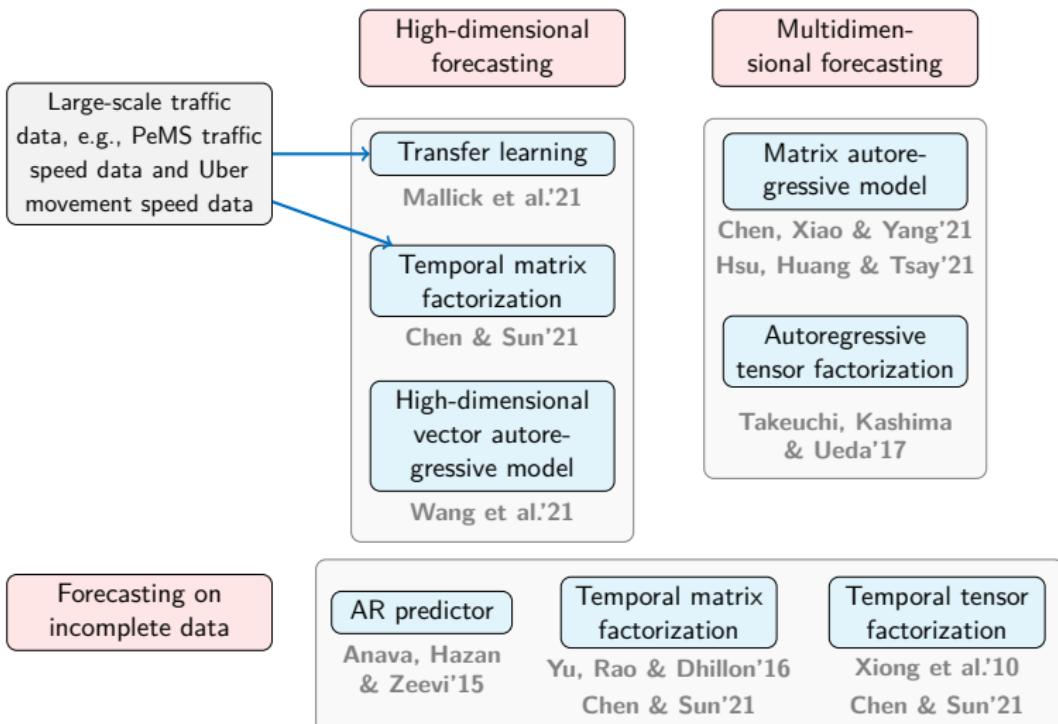
- **Seattle** movement speed data (2019)
  - **63,490** road segments & 8,760 time steps (hours)
  - Overall missing rate: **84.95%**



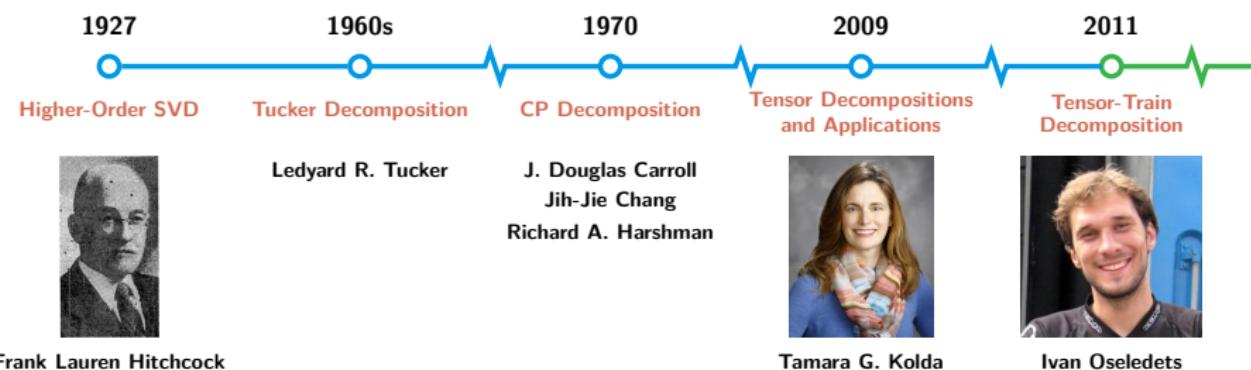
# Spatiotemporal Traffic Data Imputation



# Spatiotemporal Traffic Forecasting



# Low-Rank Tensor Learning



# Low-Rank Tensor Learning

- Low-rank matrix/tensor completion



**Candès & Recht'09:** Convex nuclear norm minimization for matrix completion.

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{s.t. } & \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{Y}) \end{aligned}$$



**Cai, Candès & Shen'10:** Singular value thresholding algorithm.

$$\begin{cases} \mathbf{X}^\ell = \mathcal{D}_\tau(\mathbf{Z}^{\ell-1}) \\ \mathbf{Z}^\ell = \mathbf{Z}^{\ell-1} + \delta_\ell \mathcal{P}_\Omega(\mathbf{Y} - \mathbf{X}^\ell) \end{cases}$$



**Zhang et al.'12:** Nonconvex truncated nuclear norm minimization.



**Liu et al.'13:** Convex nuclear norm minimization for tensor completion.

$$\begin{aligned} \min_{\mathcal{X}} \quad & \|\mathcal{X}\|_* \\ \text{s.t. } & \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Y}) \end{aligned}$$

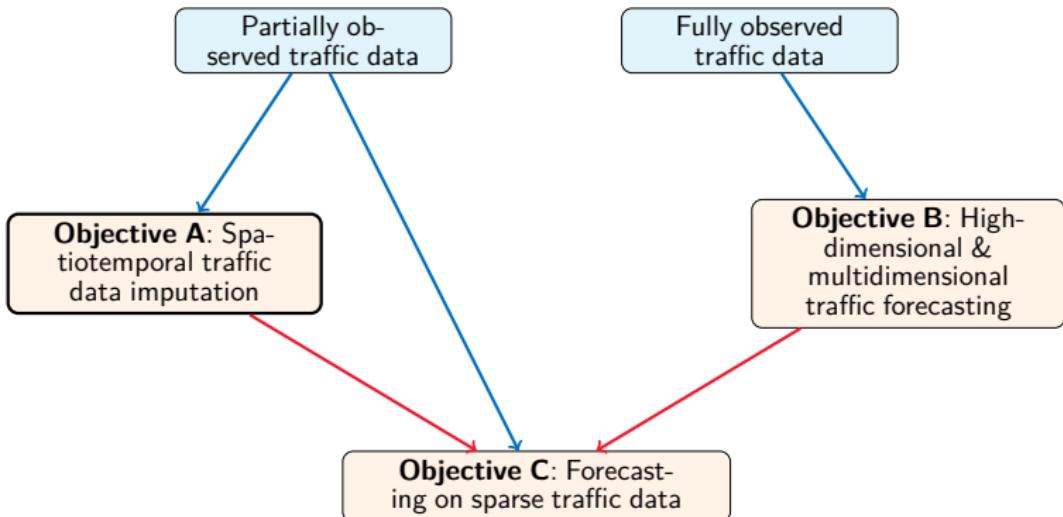


**Lu, Peng & Wei'19:** Tensor nuclear norm induced by linear transform.



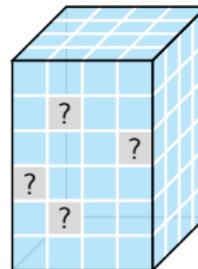
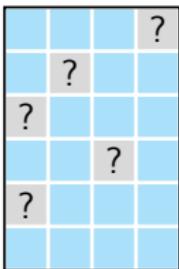
# A Whole Picture of Objectives

We are working on **spatiotemporal traffic data modeling**.



# Spatiotemporal Traffic Data Imputation

- **Objective A:** Given a multivariate time series data like  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  or a multidimensional time series data like  $\mathcal{Y} \in \mathbb{R}^{M \times N \times T}$ , impute the missing values of the data.



[Q]

- How to reconstruct missing values from observed data?
- How to make use of spatiotemporal correlations?
- How to make use of traffic time series dynamics?

# Spatiotemporal Traffic Data Imputation

## Low-rank matrix completion (Candès & Recht'09)

For any partially observed data matrix  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  with observed index set  $\Omega$ , then low-rank matrix completion takes the form of

$$\begin{aligned} & \min_{\mathbf{X}} \|\mathbf{X}\|_* \\ & \text{s.t. } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{Y}). \end{aligned} \tag{1}$$

## Low-rank tensor completion (Liu et al.'13)

For any partially observed data matrix  $\mathcal{Y} \in \mathbb{R}^{M \times N \times T}$  with observed index set  $\Omega$ , then low-rank matrix completion takes the form of

$$\begin{aligned} & \min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_* \\ & \text{s.t. } \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathcal{P}_\Omega(\mathcal{Y}). \end{aligned} \tag{2}$$

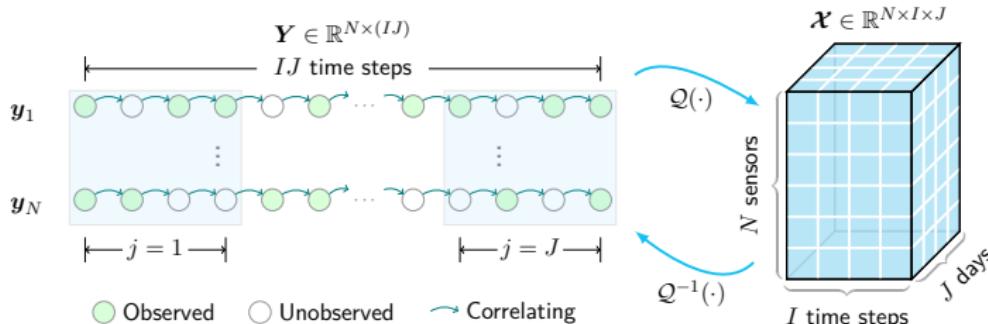
- **Limitation:** Only cover the global consistency.
- **Comment:** For modeling spatiotemporal traffic data, local consistency (e.g., temporal correlations) is also important.

# Spatiotemporal Traffic Data Imputation

## Low-rank autoregressive tensor completion (LATC)

$$\begin{aligned}
 & \min_{\mathcal{X}} \|\mathcal{X}\|_* + \frac{\lambda}{2} \sum_{n=1}^N \sum_{t=d+1}^T (z_{n,t} - \sum_{k=1}^d a_{n,k} z_{n,t-k})^2 \\
 & \text{s.t. } \begin{cases} \mathcal{X} = \mathcal{Q}(\mathbf{Z}), \\ \mathcal{P}_{\Omega}(\mathbf{Z}) = \mathcal{P}_{\Omega}(\mathbf{Y}). \end{cases}
 \end{aligned} \tag{3}$$

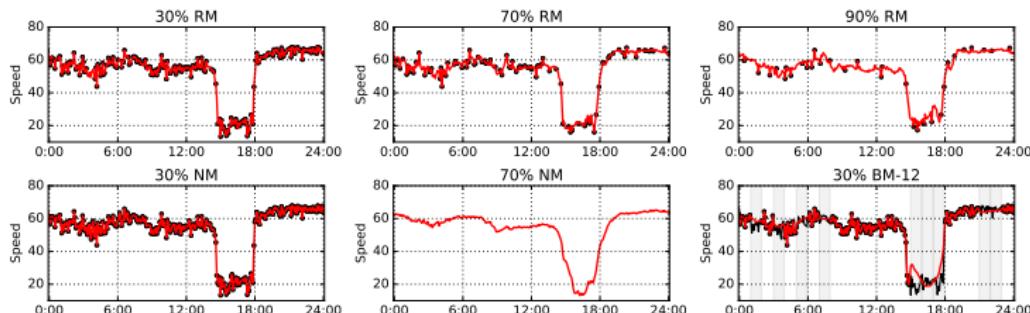
- Advantage:** Global consistency + local consistency.



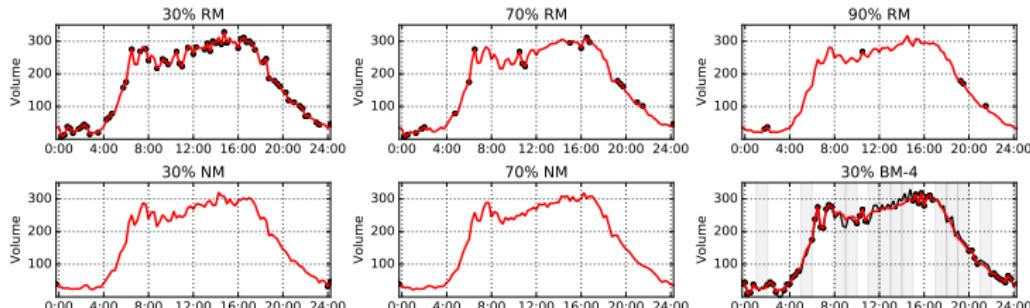
# Spatiotemporal Traffic Data Imputation

## LATC imputation

- Seattle freeway traffic speed data

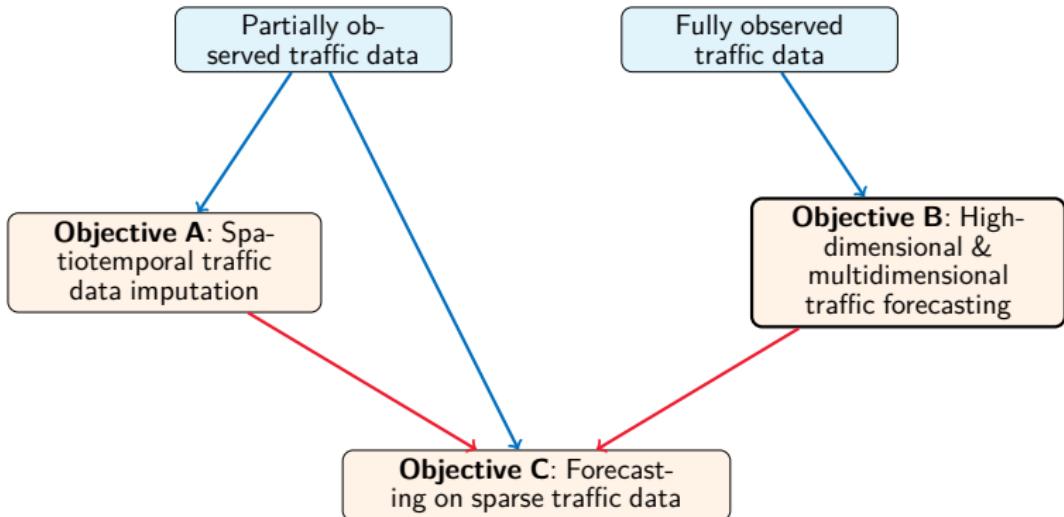


- Portland highway traffic volume data



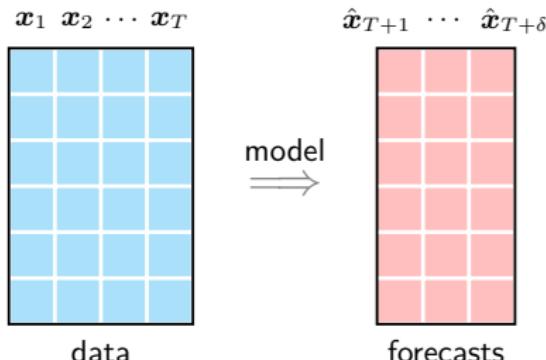
# A Whole Picture of Objectives

We are working on **spatiotemporal traffic data modeling**.



# High-Dimensional Traffic Forecasting

- **Objective B-1:** Given a multivariate traffic time series  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{R}^N$  with  $N \gg T$  ("tall-skinny"), forecast data points  $\hat{\mathbf{x}}_{T+\delta}, \delta \in \mathbb{N}^+$ .

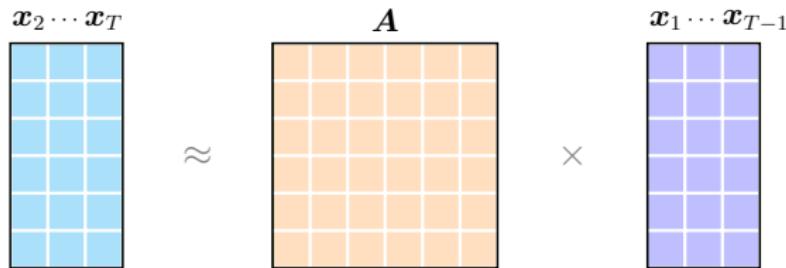


- **Solution:** For time series  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{R}^N$ , the  $d$ th-order vector autoregressive (VAR( $d$ )) model:  $\mathbf{x}_t = \sum_{k=1}^d \mathbf{A}_k \mathbf{x}_{t-k} + \boldsymbol{\epsilon}_t$ .
- **Advantage:** Co-evolution patterns
- **Limitation:** Over-parameterization

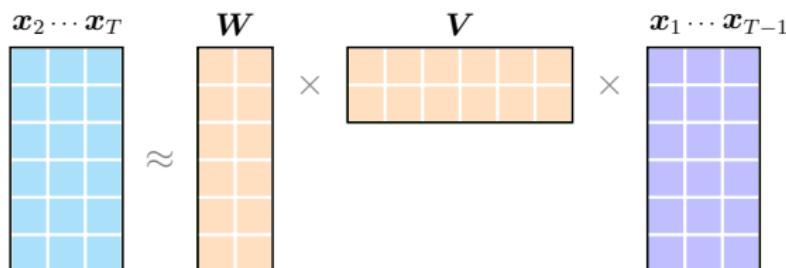
# High-Dimensional Traffic Forecasting

## VAR(1) model

- Over-parameterization in the case of  $N \gg T$ .



- Reduced-rank autoregression:  $A = WV$  with  $W \in \mathbb{R}^{N \times R}, V \in \mathbb{R}^{R \times N}$ .



# High-Dimensional Traffic Forecasting

## VAR( $d$ ) model

- Recall that  $\mathbf{x}_t = \sum_{k=1}^d \mathbf{A}_k \mathbf{x}_{t-k} + \boldsymbol{\epsilon}_t$ .
- Coefficients  $\mathbf{A}_k \in \mathbb{R}^{N \times N}$ ,  $k = 1, \dots, d$  are tensor, e.g.,  $\mathcal{A} \in \mathbb{R}^{N \times N \times d}$ .

## VAR( $d$ ) with Tucker decomposition (Wang et al.'21)

For VAR( $d$ ) on the multivariate time series  $\mathbf{x}_t \in \mathbb{R}^N$ ,  $t = 1, \dots, T$ , the reduced-rank VAR via Tucker decomposition is given by

$$\min_{\mathcal{G}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} \frac{1}{2} \sum_{t=d+1}^T \|\mathbf{x}_t - (\mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{U} \times_3 \mathbf{H})_{(1)} \mathbf{z}_t\|_2^2 \quad (4)$$

where  $\mathbf{z}_t = (\mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-d}^\top)^\top \in \mathbb{R}^{dN}$ . The multilinear rank is  $(R_1, R_2, R_3)$ .

$\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$  is the core tensor, while  $\mathbf{W} \in \mathbb{R}^{N \times R_1}$ ,  $\mathbf{U} \in \mathbb{R}^{N \times R_2}$ , and  $\mathbf{H} \in \mathbb{R}^{d \times R_3}$  are the component matrices.

**Advantage:** High compression rate.

**Limitations:** Nonconvex optimization; the model is failing in nonstationary time series.

# High-Dimensional Traffic Forecasting

## Reduced-rank time-varying VAR

Given traffic data samples  $\{\mathbf{x}_{t-d}, \dots, \mathbf{x}_t\}$  at time  $t$ , let  $\mathbf{y}_t \triangleq \mathbf{x}_t \in \mathbb{R}^N$  and  $\mathbf{z}_t \triangleq (\mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-d}^\top)^\top \in \mathbb{R}^{dN}$ , then the reduced-rank time-varying VAR takes

$$\begin{aligned} \min_{\mathbf{G}_t, \mathbf{W}_t, \mathbf{U}_t, \mathbf{H}_t} \quad & \frac{1}{2} (\|\mathbf{G}_t - \mathbf{G}_{t-1}\|_F^2 + \|\mathbf{W}_t - \mathbf{W}_{t-1}\|_F^2 \\ & + \|\mathbf{U}_t - \mathbf{U}_{t-1}\|_F^2 + \|\mathbf{H}_t - \mathbf{H}_{t-1}\|_F^2) \\ & + \frac{\lambda}{2} \|\mathbf{y}_t - \mathbf{W}_t \mathbf{G}_t (\mathbf{H}_t \otimes \mathbf{U}_t)^\top \mathbf{z}_t\|_2^2 \end{aligned} \quad (5)$$

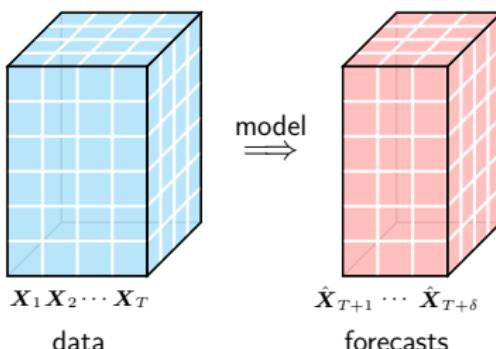
where  $\mathbf{G}_t \in \mathbb{R}^{R \times (dR)}$ ,  $\mathbf{W}_t \in \mathbb{R}^{N \times R}$ ,  $\mathbf{U}_t \in \mathbb{R}^{N \times R}$ ,  $\mathbf{H}_t \in \mathbb{R}^{d \times d}$ .

## Advantages:

- Produce time-varying parameters.
- Overcome nonstationarity.

# Multidimensional Traffic Forecasting

- **Objective B-2:** Given a multidimensional traffic time series  $\mathbf{X}_1, \dots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$ , forecast data points  $\hat{\mathbf{X}}_{T+\delta}, \delta \in \mathbb{N}^+$ .



[Q]

- How to perform forecasting on this kind of data?
- How to preserve the intrinsic tensor representation of data?

# Multidimensional Traffic Forecasting

## Matrix autoregressive model (Chen, Xiao & Yang'21)

Given matrix-variate time series  $\mathbf{X}_t \in \mathbb{R}^{M \times N}$ ,  $t = 1, \dots, T$ , then the  $d$ th-order matrix autoregressive (MAR( $d$ )) model takes the form of

$$\mathbf{X}_t = \sum_{k=1}^d \mathbf{A}_k \mathbf{X}_{t-k} \mathbf{B}_k^\top + \mathbf{E}_t \quad (6)$$

where  $\mathbf{A}_k \in \mathbb{R}^{M \times M}$ ,  $\mathbf{B}_k \in \mathbb{R}^{N \times N}$ ,  $k = 1, \dots, d$  are the coefficient matrices.

### Advantages:

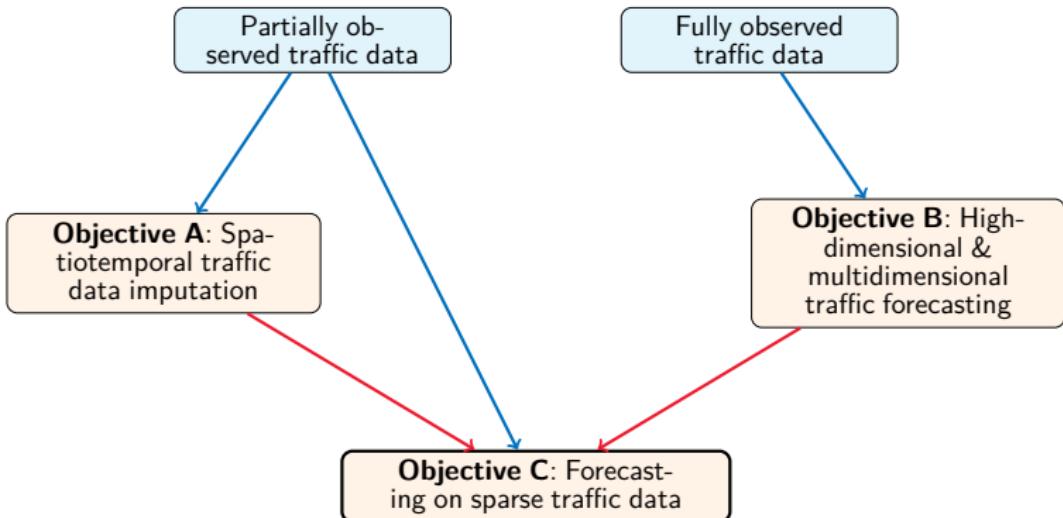
- Preserve the intrinsic tensor representation.
- Reduce parameters in autoregressive models (if  $n = \max\{M, N\}$ ), e.g.,

$$\mathcal{O}(n^4) \text{ in VAR(1)} \quad \text{vs.} \quad \mathcal{O}(n^2) \text{ in MAR(1)}$$

**Limitation:** Failing in nonstationary time series.

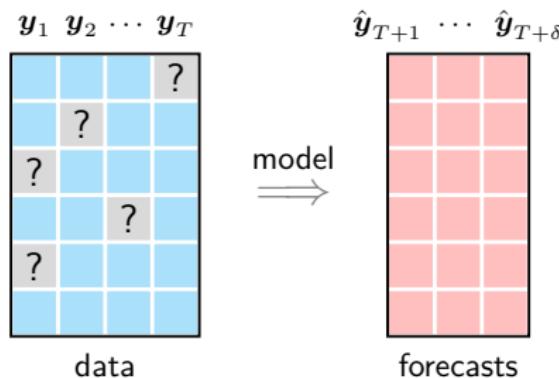
# A Whole Picture of Objectives

We are working on **spatiotemporal traffic data modeling**.



# Multivariate Traffic Forecasting on Sparse Data

- **Objective C-1:** Given a partially observed data  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  consisting of time series  $y_1, \dots, y_T \in \mathbb{R}^N$ , forecast data points  $y_{T+\delta}, \delta \in \mathbb{N}^+$ .



[Q]

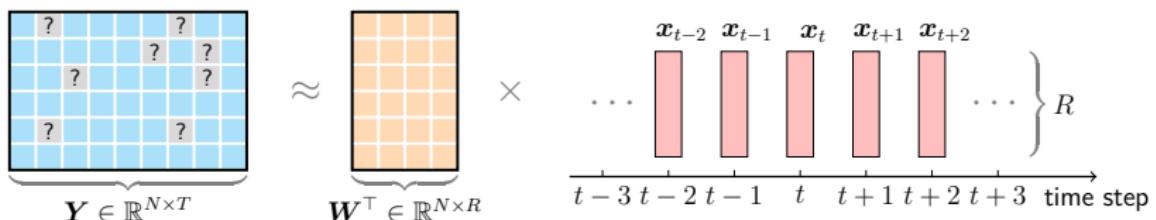
- How to learn from *high-dimensional* and *sparse* data?
- How to model *nonstationarity* in time series?
- How to perform forecasting on these time series?

# Multivariate Traffic Forecasting on Sparse Data

## Temporal matrix factorization (Yu et al.'16; Chen & Sun'21)

Given any partially observed time series data  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  with observed index set  $\Omega$ , then temporal matrix factorization assumes a  $d$ th-order vector autoregressive (VAR) process on the temporal factor matrix:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{X}, \{\mathbf{A}_k\}_{k=1}^d} & \frac{1}{2} \|\mathcal{P}_\Omega(\mathbf{Y} - \mathbf{W}^\top \mathbf{X})\|_F^2 + \frac{\rho}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{X}\|_F^2) \\ & + \frac{\lambda}{2} \sum_{t=d+1}^T \|\mathbf{x}_t - \sum_{k=1}^d \mathbf{A}_k \mathbf{x}_{t-k}\|_2^2 \end{aligned} \quad (7)$$



⚠ VAR is usually built on stationary time series (temporal factors).

# Multivariate Traffic Forecasting on Sparse Data

## Nonstationary temporal matrix factorization (NoTMF)

Given any partially observed time series data  $\mathbf{Y} \in \mathbb{R}^{N \times T}$  with observed index set  $\Omega$ , then we assume a season- $m$  differencing on the latent temporal factors:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{X}, \{\mathbf{A}_k\}_{k=1}^d} & \frac{1}{2} \|\mathcal{P}_\Omega(\mathbf{Y} - \mathbf{W}^\top \mathbf{X})\|_F^2 + \frac{\rho}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{X}\|_F^2) \\ & + \frac{\lambda}{2} \sum_{t=d+m+1}^T \|(\mathbf{x}_t - \mathbf{x}_{t-m}) - \sum_{k=1}^d \mathbf{A}_k (\mathbf{x}_{t-k} - \mathbf{x}_{t-m-k})\|_2^2 \end{aligned} \quad (8)$$

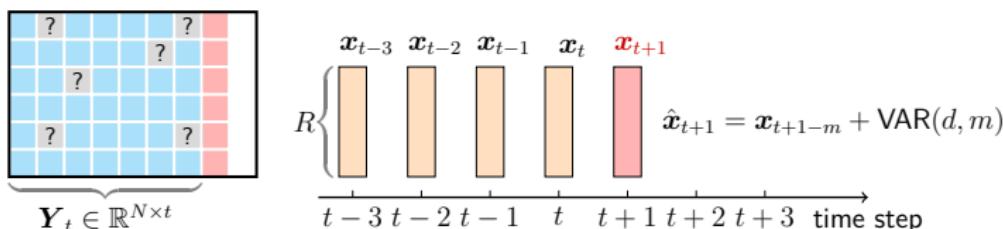
- First-order differencing  $\mathbf{x}'_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ .
  - Second-order differencing  $\mathbf{x}''_t = (\mathbf{x}_t - \mathbf{x}_{t-1}) - (\mathbf{x}_{t-1} - \mathbf{x}_{t-2})$ .
  - Twice-differenced series  $\mathbf{x}'''_t = (\mathbf{x}_t - \mathbf{x}_{t-m}) - (\mathbf{x}_{t-1} - \mathbf{x}_{t-m-1})$ .
- 😊 Stationarizing a time series with differencing can improve the prediction.<sup>4</sup>

<sup>4</sup> Stationarity and differencing: <https://otexts.com/fpp2/stationarity.html>

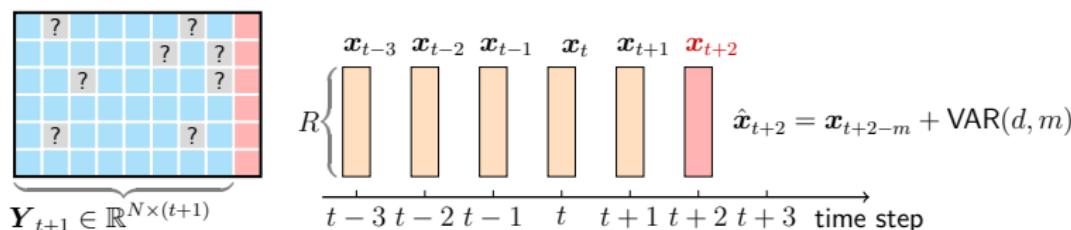
# Multivariate Traffic Forecasting on Sparse Data

NoTMF forecasting on streaming data?

- NoTMF: Use  $\mathbf{Y}_t$  to estimate  $\{\mathbf{W}, \mathbf{X}, \mathbf{A}\}$ .



- Online forecasting (Gultekin & Paisley'18): Fix  $\mathbf{W}$  and use  $\mathbf{Y}_{t+1}$  to update  $\{\mathbf{X}, \mathbf{A}\}$ .

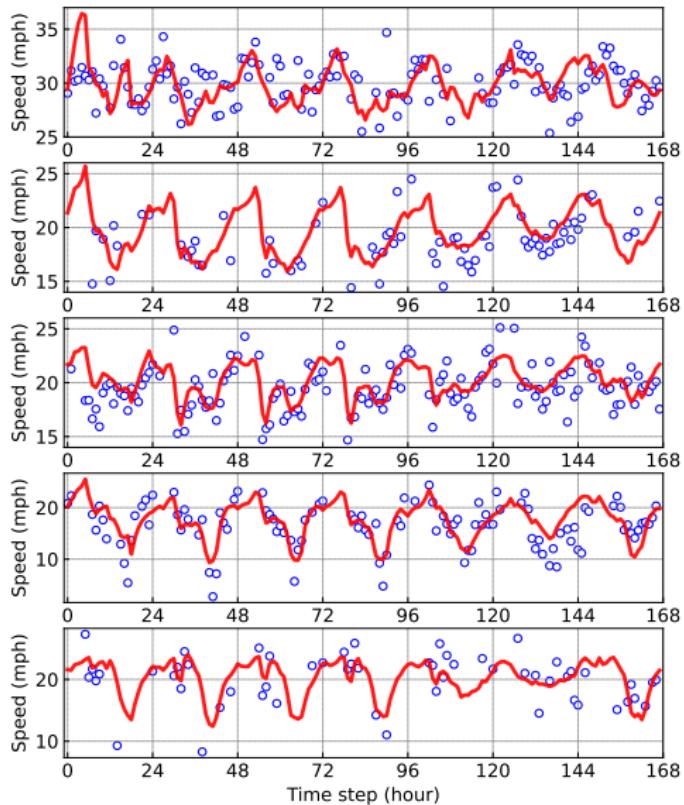


# Multivariate Traffic Forecasting on Sparse Data

**Forecasting performance on NYC Uber movement speed data (MAPE/RMSE):**

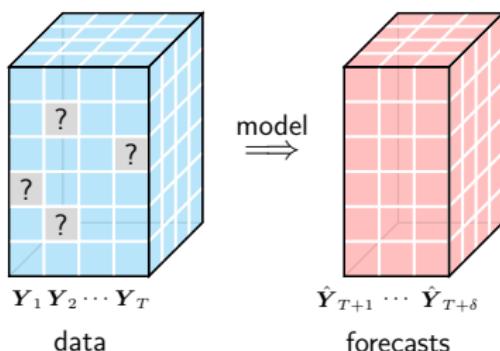
- NoTMF outperforms other models.
- Seasonal differencing in NoTMF is important.

$\delta$	$d$	NoTMF ( $m = 24$ )	NoTMF ( $m = 168$ )	NoTMF-twice ( $m = 168$ )	TMF	TRMF	BTMF
1	1	13.63/2.88	13.53/2.86	<b>13.45/2.85</b>	13.74/2.90	14.50/3.12	14.94/3.13
	2	<b>13.47/2.84</b>	<b>13.41/2.84</b>	13.42/2.84	13.53/2.85	14.14/3.05	15.70/3.41
	3	13.46/2.84	<b>13.39/2.83</b>	13.43/2.84	<b>13.47/2.83</b>	13.87/2.96	15.80/3.34
	6	<b>13.41/2.83</b>	<b>13.39/2.83</b>	13.41/2.83	13.40/ <b>2.83</b>	14.00/2.98	15.45/3.27
2	1	13.91/2.96	13.76/2.94	<b>13.70/2.92</b>	14.24/3.00	15.85/3.43	15.33/3.21
	2	<b>13.77/2.92</b>	<b>13.63/2.89</b>	13.72/2.92	13.87/2.91	15.04/3.31	15.87/3.32
	3	13.72/2.91	<b>13.61/2.89</b>	13.73/2.92	<b>13.81/2.89</b>	15.25/3.36	15.69/3.33
	6	13.59/2.87	<b>13.57/2.88</b>	13.68/2.91	<b>13.63/2.86</b>	14.92/3.24	15.91/3.39
3	1	14.30/3.05	14.06/3.02	<b>14.02/3.00</b>	14.81/3.12	17.52/3.83	15.86/3.32
	2	14.01/2.98	<b>13.84/2.94</b>	13.96/2.98	14.26/2.98	17.32/4.00	16.30/3.40
	3	13.95/2.97	<b>13.79/2.93</b>	13.98/2.98	14.04/2.94	16.91/3.71	16.56/3.49
	6	<b>13.78/2.92</b>	<b>13.73/2.92</b>	13.91/2.96	<b>13.94/2.92</b>	16.72/3.65	15.49/3.27
6	1	<b>14.61/3.11</b>	14.67/3.20	14.98/3.32	15.41/3.21	21.20/4.70	15.99/3.32
	2	<b>14.30/3.03</b>	14.33/3.09	14.90/3.28	14.85/3.07	20.87/5.01	16.04/3.33
	3	<b>14.26/3.03</b>	14.28/3.09	14.86/3.26	<b>14.57/3.01</b>	20.08/4.65	15.67/3.28
	6	<b>14.06/2.97</b>	14.16/3.06	14.80/3.23	14.47/3.00	20.40/4.35	16.38/3.50



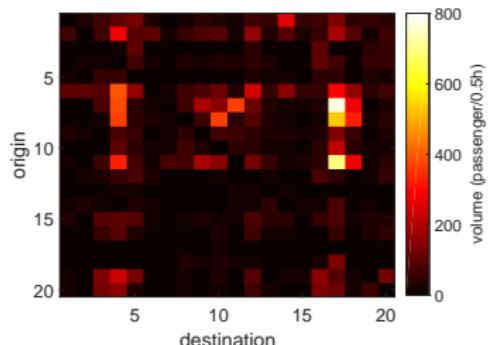
# Multidimensional Traffic Forecasting on Sparse Data

- **Objective C-2:** Given a partially observed data  $\mathcal{Y} \in \mathbb{R}^{M \times N \times T}$  consisting of time series  $\mathbf{Y}_1, \dots, \mathbf{Y}_T \in \mathbb{R}^{M \times N}$ , forecast data points  $\hat{\mathbf{Y}}_{T+\delta}, \delta \in \mathbb{N}^+$ .

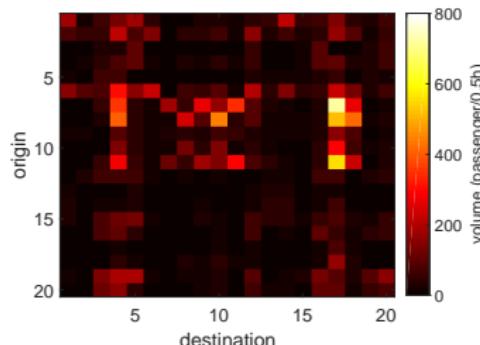


- **Solution:** Temporal tensor factorization, e.g., tensor factorization + VAR (on latent temporal factors).

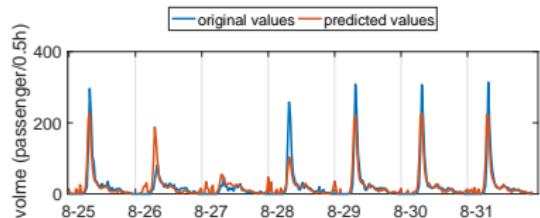
# Multidimensional Traffic Forecasting on Sparse Data



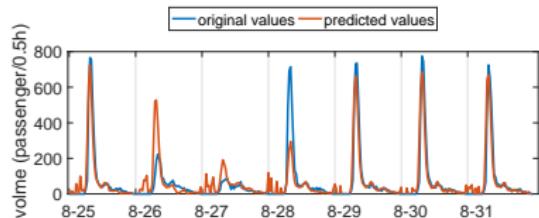
Original OD volume matrix.



Predicted OD volume matrix.



Volume of the #(6,17) OD pair.



Volume of the #(7,17) OD pair.

# Conclusion

---

- Contributions

- *Objective A: Spatiotemporal traffic data imputation.* Develop a low-rank temporal modeling framework and improve the imputation accuracy, efficiency, and scalability.
- *Objective B: High-dimensional and multidimensional forecasting.* Fast and accurate forecasting approach for high-dimensional and large-scale data; tensor representation based autoregressive model for multidimensional data.
- *Objective C: Forecasting on sparse data.* Low-rank temporal modeling framework for traffic time series forecasting in the presence of missing values.

- Significance

- Improve traffic data quality.
- Support data-driven intelligent transportation applications.

# Research Work during Ph.D. Research

- Publications
  - [J1] X. Chen, M. Lei, N. Saunier, L. Sun (2021). Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*. (Early access)
  - [J2] X. Chen, Y. Chen, N. Saunier, L. Sun (2021). Scalable low-rank tensor learning for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies*, 129: 103226.
  - [C1] X. Chen, M. Lei, N. Saunier, L. Sun (2021). Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *The 7th SIGKDD Workshop on Mining and Learning from Time Series (MiLeTS)* at KDD 2021.
- Preprint (under review)
  - [P1] X. Chen, C. Zhang, X.L. Zhao, N. Saunier, L. Sun (2022). Nonstationary temporal matrix factorization for multivariate time series forecasting. arXiv preprint arXiv:2203.10651.
- Open-source projects
  - **transdim**: Machine learning for spatiotemporal traffic data imputation and forecasting. (780 stars & 240 forks on GitHub)  
<https://github.com/xinchen/transdim>
  - **tracebase**: Multivariate time series forecasting on high-dimensional and sparse Uber movement speed data. (20 stars & 5 forks on GitHub)  
<https://github.com/xinchen/tracebase>



POLYTECHNIQUE  
MONTRÉAL

UNIVERSITÉ  
D'INGÉNIERIE



# Thanks for your attention!

## Any Questions?

### About me:

-  Homepage: <https://xinychen.github.io>
-  GitHub: <https://github.com/xinychen> (2.4K+ stars)
-  Blog: <https://medium.com/@xinyu.chen> (30K+ views)
-  How to reach me: [chenxy346@gmail.com](mailto:chenxy346@gmail.com)