

Recovering large-scale incomplete traffic speed via tensor completion

Xinyu Chen, Master's student

Zhaocheng He, professor

Jiawei Wang, Master's student

Yuhuan Lu, Master's student

Research Center of Intelligent Transportation System

Sun Yat-Sen University

Background

- Traffic data can be collected from a number of **mobile** and **stationary** sensors nowadays.
- However, the problem of **missing data** is inevitable due to **communication malfunctions** and **transmission distortions**.
- **Example**
 - Traffic **volume** matrix with **five sensors** and **four 15-minute time windows**.

$$X = \begin{bmatrix} ? & 99 & 449 & 517 \\ ? & ? & 412 & ? \\ 192 & ? & 697 & 687 \\ 185 & ? & 699 & 657 \\ 164 & 68 & ? & ? \end{bmatrix} \in \mathbb{R}^{5 \times 4}$$

- **Question**
 - How to estimate **unknown** $x_{11}, x_{21}, x_{22}, \dots, x_{53}, x_{54}$ according to the **observations** $x_{12} = 99, x_{13} = 449, x_{14} = 517, \dots, x_{51} = 164, x_{52} = 68$?



Matrix Decomposition

- **Model**

- For (i, j) -th entry of $X \in \mathbb{R}^{m \times n}$, if we assume

$$x_{ij} \approx \sum_{k=1}^r u_{ik} v_{jk}$$

u_{ik} is the k -th factor related to the i -th sensor
 v_{jk} is the k -th factor related to the j -th time window

then the **factor matrices** $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ can be learned by solving

$$J = \frac{1}{2} \|S * (X - UV^T)\|_F^2$$

S is an **indicator** for **observed entries**

- **With regularization**

$$J = \frac{1}{2} \|S * (X - UV^T)\|_F^2 + \frac{\lambda}{2} (\|U\|_F^2 + \|V\|_F^2)$$

- $*$: the *Hadamard product* of matrices (or tensors) with same size.
- $\|\cdot\|_F^2$: the sum of squared entries.
- λ : the parameter of regularization term.

Matrix Decomposition

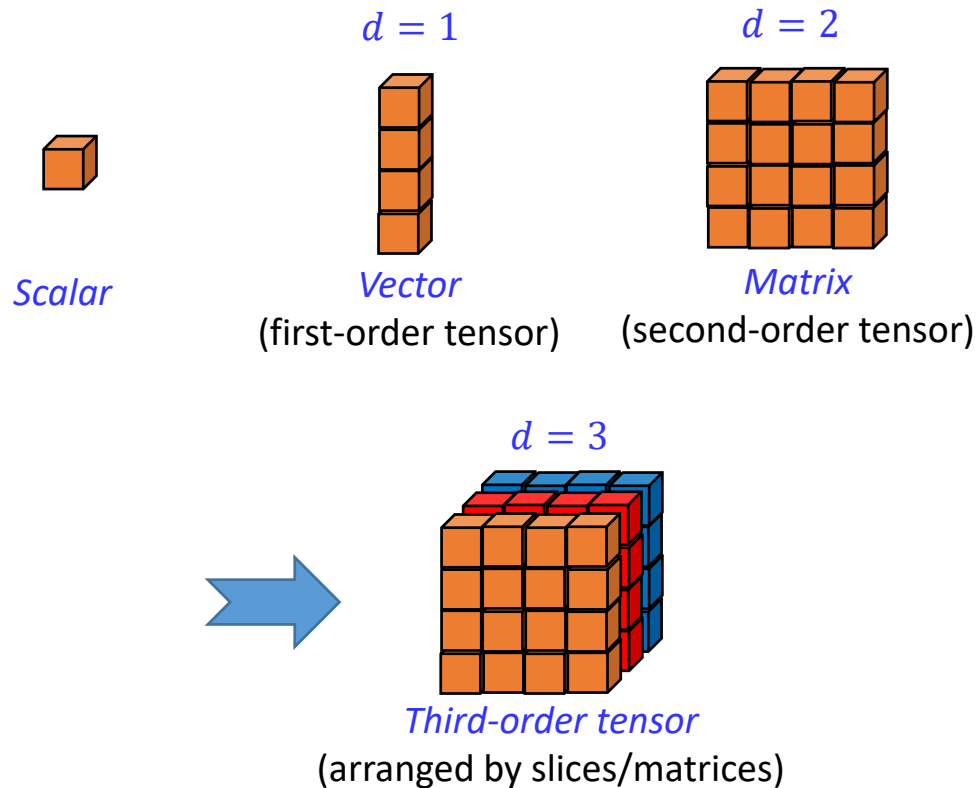
- A simple example

$$\begin{aligned}
 & \begin{bmatrix} 15 & 10 \\ 7 & 12 \\ 12 & 20 \\ 9 & 21 \\ 8 & 18 \end{bmatrix} \times \begin{bmatrix} 4 \\ 7 \end{bmatrix} \begin{matrix} 5 & 11 & 20 \\ 1 & 28 & 23 \end{matrix} = \begin{bmatrix} ? & 85 & 445 & 530 \\ ? & ? & 413 & ? \\ 188 & ? & 692 & 700 \\ 183 & ? & 687 & 663 \\ 158 & 58 & ? & ? \end{bmatrix} \\
 & \begin{bmatrix} 15 & 10 \\ 7 & 12 \\ 12 & 20 \\ 9 & 21 \\ 8 & 18 \end{bmatrix} \times \begin{bmatrix} 4 & 5 \\ 7 & 1 \end{bmatrix} \begin{matrix} 11 & 20 \\ 28 & 23 \end{matrix} = \begin{bmatrix} ? & 85 & 445 & 530 \\ ? & ? & 413 & ? \\ 188 & ? & 692 & 700 \\ 183 & ? & 687 & 663 \\ 158 & 58 & ? & ? \end{bmatrix} \\
 & \Rightarrow \begin{bmatrix} 15 & 10 \\ 7 & 12 \\ 12 & 20 \\ 9 & 21 \\ 8 & 18 \end{bmatrix} \times \begin{bmatrix} 4 & 5 \\ 7 & 1 \end{bmatrix} \begin{matrix} 11 & 20 \\ 28 & 23 \end{matrix} = \begin{bmatrix} ? & 85 & 445 & 530 \\ ? & ? & 413 & ? \\ 188 & 80 & 692 & 700 \\ 183 & ? & 687 & 663 \\ 158 & 58 & ? & ? \end{bmatrix}
 \end{aligned}$$



What is tensor?

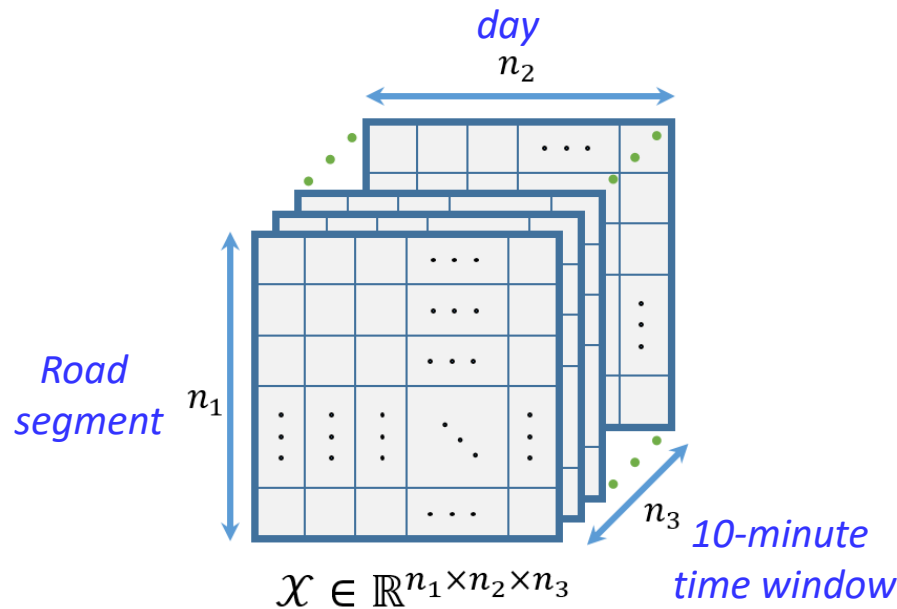
- Intuition:** $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$



What is tensor?

- **Traffic speed dataset**

- 214 road segments
- 61 days (from Aug. 1 to Sep. 30, 2016)
- 144 10-minute time windows
- Million scale & city-wide: $214 \times 61 \times 144 \approx 1.88 \times 10^6$

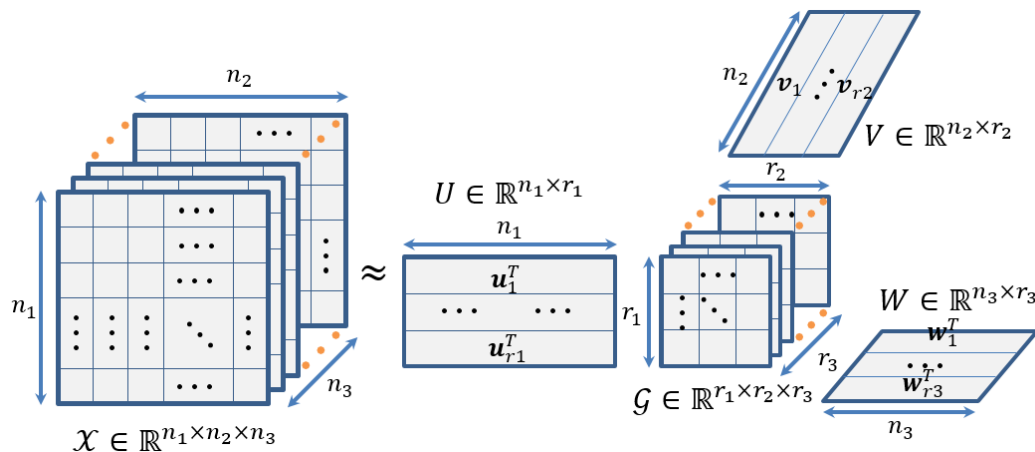


Tensor Decomposition

- **Tucker Decomposition (Tukcer, 1966)**

- Decomposes a given tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into a **core tensor** $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and **factor matrices** $U \in \mathbb{R}^{n_1 \times r_1}$, $V \in \mathbb{R}^{n_2 \times r_2}$ and $W \in \mathbb{R}^{n_3 \times r_3}$ in a sequence.

$$\mathcal{X} \approx \mathcal{G} \times_1 U \times_2 V \times_3 W$$



- **Non-constraint optimization** *Decomposition based approximation*

$J = \frac{1}{2} \|\mathcal{S} * (\mathcal{X} - \mathcal{G} \times_1 U \times_2 V \times_3 W)\|_F^2 + \frac{\lambda}{2} (\|\mathcal{G}\|_F^2 + \|U\|_F^2 + \|V\|_F^2 + \|W\|_F^2)$

an **indicator** for **observed entries**



- \times_q : the *tensor-matrix multiplication* or *modal-q product* between tensor and matrix.
- λ : the parameter of regularization term.

Tensor Decomposition

- **Non-constraint optimization**

$$J = \frac{1}{2} \|\mathcal{S} * (\mathcal{X} - \mathcal{G} \times_1 U \times_2 V \times_3 W)\|_F^2 + \frac{\lambda}{2} (\|\mathcal{G}\|_F^2 + \|U\|_F^2 + \|V\|_F^2 + \|W\|_F^2)$$

an *indicator* for *observed entries*

- **Gradient Descent Method (GDM)**

$$U \leftarrow (1 - \alpha\lambda)U + \alpha(\mathcal{S} * \mathcal{E})_{(1)}(W \otimes V)\mathcal{G}_{(1)}^T$$

$$V \leftarrow (1 - \alpha\lambda)V + \alpha(\mathcal{S} * \mathcal{E})_{(2)}(W \otimes U)\mathcal{G}_{(2)}^T$$

$$W \leftarrow (1 - \alpha\lambda)W + \alpha(\mathcal{S} * \mathcal{E})_{(3)}(V \otimes U)\mathcal{G}_{(3)}^T$$

$$\mathcal{G} \leftarrow (1 - \alpha\lambda)\mathcal{G} + \alpha \cdot \mathcal{E} \times_1 U^T \times_2 V^T \times_3 W^T$$

- **Challenge: non-convex optimization!**

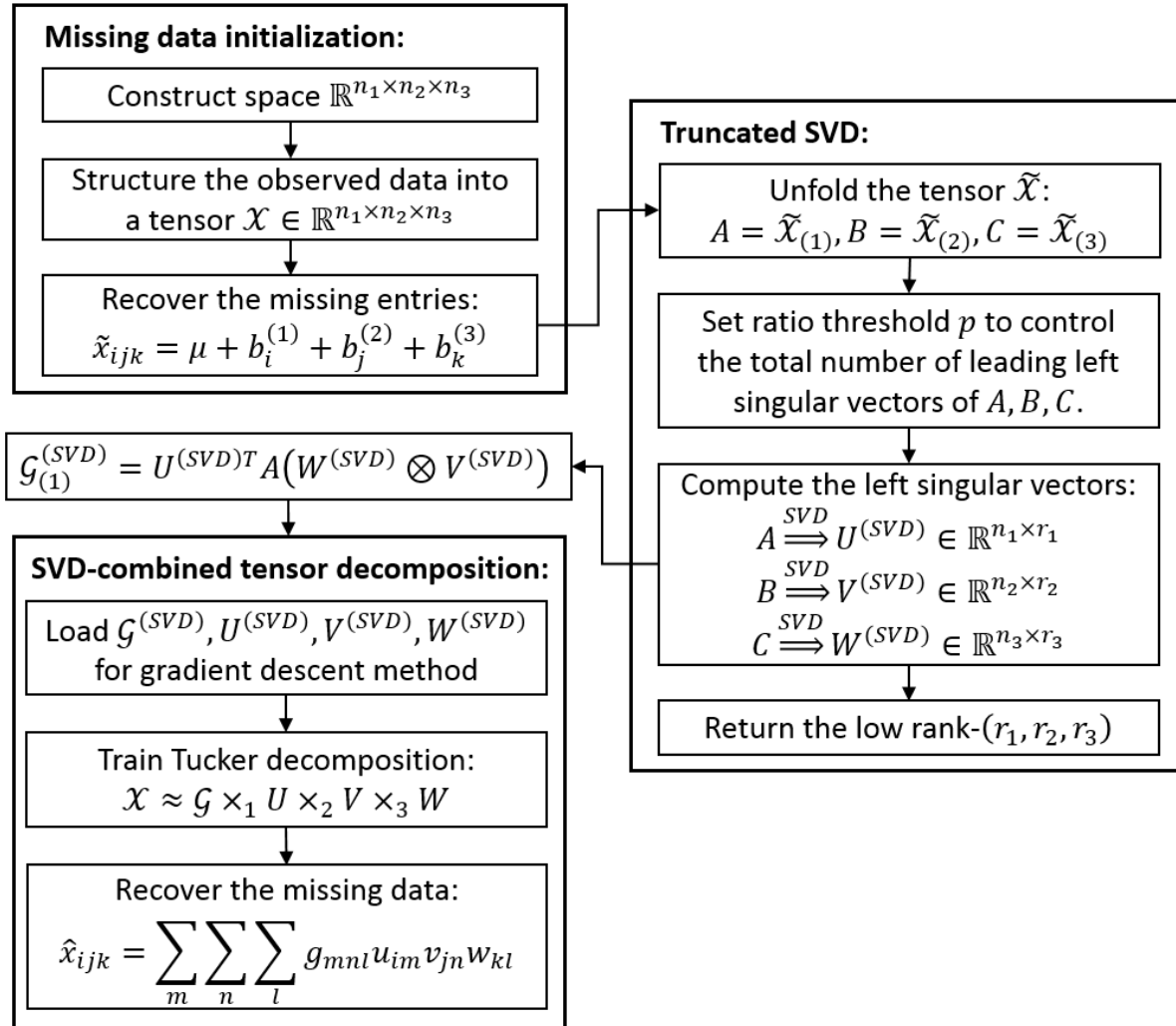
- For any (i, j, k) -th entry, the estimation is formulated by

$$x_{ijk} \approx \sum_{m=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} u_{im} v_{jn} w_{kl}$$



- \otimes : Kronecker product.
- α : learning rate of gradient descent method.
- $\mathcal{X}_{(q)}$: mode- q unfolding of tensor \mathcal{X} .

Modeling framework



HaLRTC

- **High accuracy low rank tensor completion (HaLRTC, Liu et al., 2013)**
 - **Low rank matrix completion** (*the minimum rank solution can be recovered by solving a convex problem, i.e., the minimization of the trace norm*)

$$\begin{array}{ccc} \text{Non-convex} & & \text{Convex} \\ \min_{\hat{X}} \text{rank}(\hat{X}) & \Leftrightarrow & \min_{\hat{X}} \|\hat{X}\|_* \\ \text{s. t. } S * \hat{X} = S * X & & \text{s. t. } S * \hat{X} = S * X \end{array}$$

Trace norm: the sum of singular values

- **Low rank tensor completion**

$$\begin{array}{l} \min_{\hat{\mathcal{X}}, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3} \frac{1}{3} \left(\|\mathcal{B}_{1(1)}\|_* + \|\mathcal{B}_{2(2)}\|_* + \|\mathcal{B}_{3(3)}\|_* \right) \\ \text{s. t. } \begin{cases} \mathcal{S} * \hat{\mathcal{X}} = \mathcal{S} * \mathcal{X} \\ \hat{\mathcal{X}} = \mathcal{B}_q, q = 1, 2, 3 \end{cases} \end{array}$$

- *The rank is a powerful tool to capture some types of global information.*

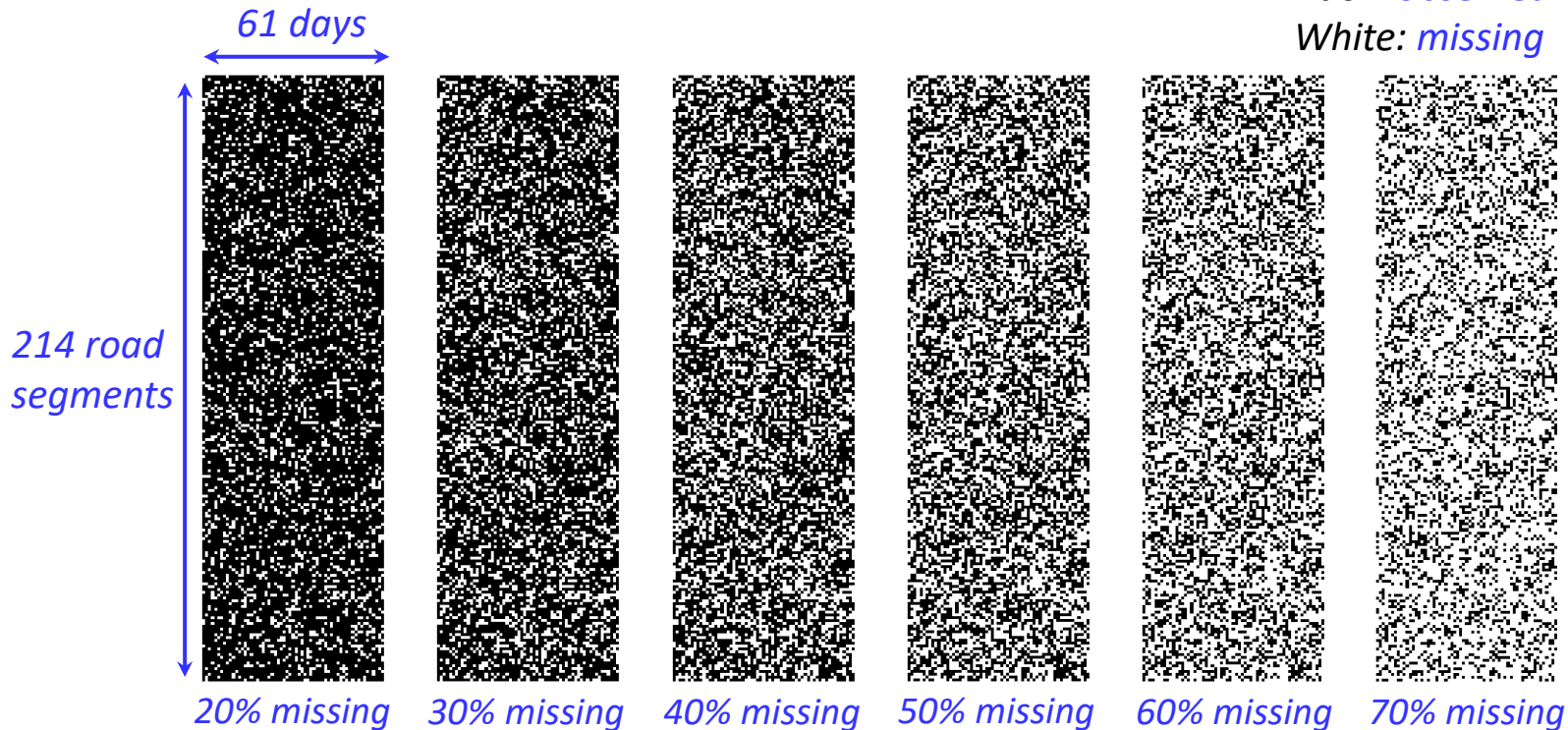


Empirical studies

- **Fiber-like missing**

- $S \in \mathbb{R}^{214 \times 61}$ is a binary matrix where $s_{ij} = \text{round}(a_{ij} + 0.5 - \theta)$ and $a_{ij} \sim \text{Uniform}(0,1)$.
- $\mathcal{X} \leftarrow \mathcal{X} * S$ is a sparse tensor where $S(i, j, :) = S(i, j)$.

Black: observed
White: missing



中山大學
SUN YAT-SEN UNIVERSITY

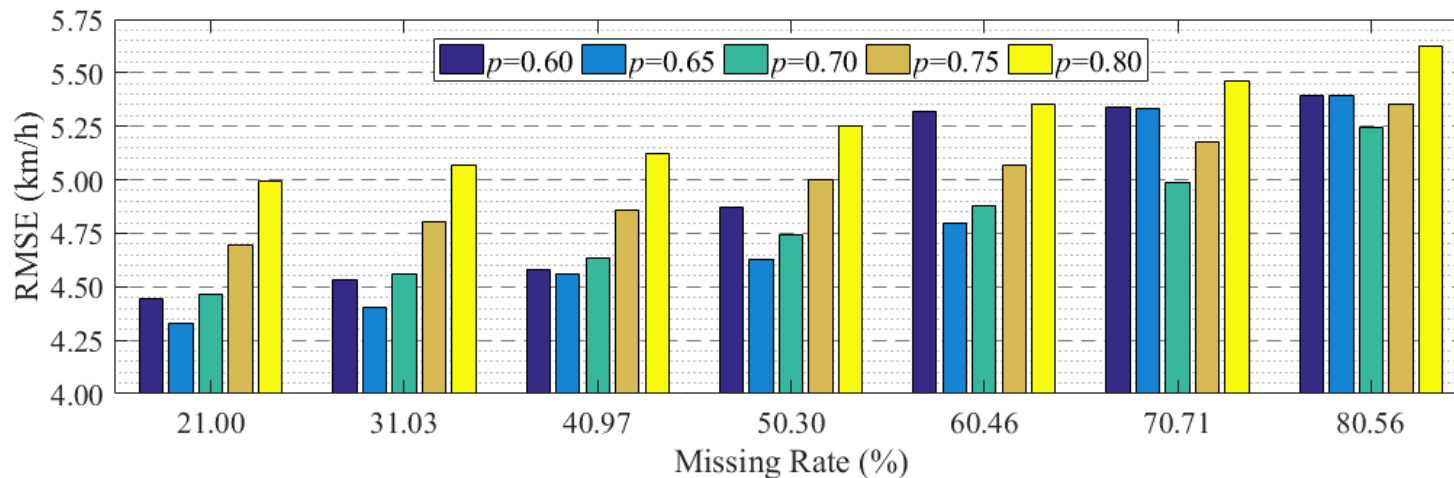
- θ is a variable to condition the missing, for example, if $\theta = 0.7$, then $s_{ij} = \text{round}(a_{ij} - 0.2)$ and approximate 70% entries of S are 0.

Empirical studies

- **Determination of core tensor size**

- In the framework, p is the *ratio threshold of singular values* for each unfolding.
- The *number of leading singular vectors* for unfolding $\tilde{\mathcal{X}}_{(q)}$ is r_q , $q = 1, 2, 3$.
- The *core tensor size* is $r_1 \times r_2 \times r_3$.

Sensitivity test over final results

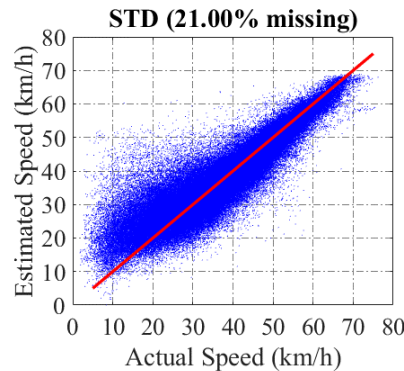


- While missing rate *-ranging from 20% to 60%*, the best threshold is $p = 0.65$.
- For *heavier missing*, the best threshold is $p = 0.70$.

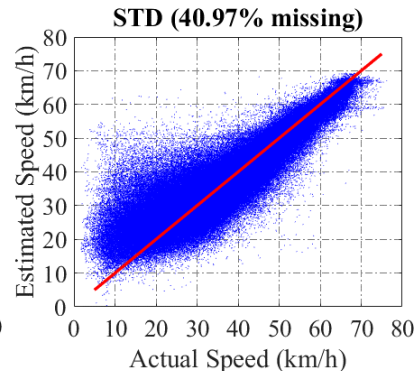


Empirical studies

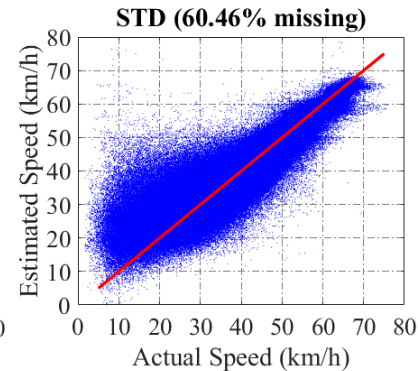
- Overall performance** (missing rate is 20%, 40% and 60%)



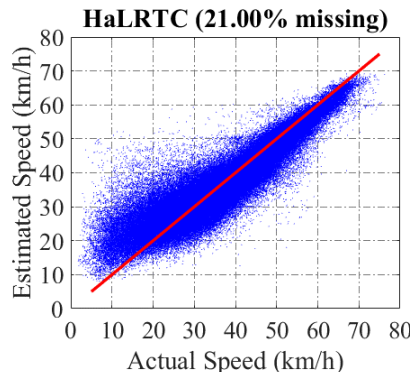
$RMSE = 4.33\text{km/h}$
 $MRE = 7.29\%$



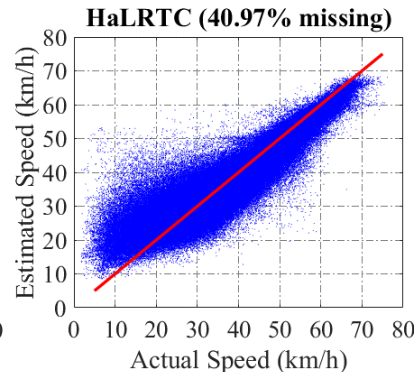
$RMSE = 4.56\text{km/h}$
 $MRE = 7.66\%$



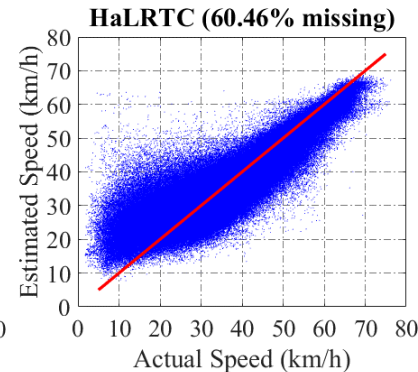
$RMSE = 4.80\text{km/h}$
 $MRE = 8.12\%$



$RMSE = 4.21\text{km/h}$
 $MRE = 7.58\%$



$RMSE = 4.38\text{km/h}$
 $MRE = 8.05\%$

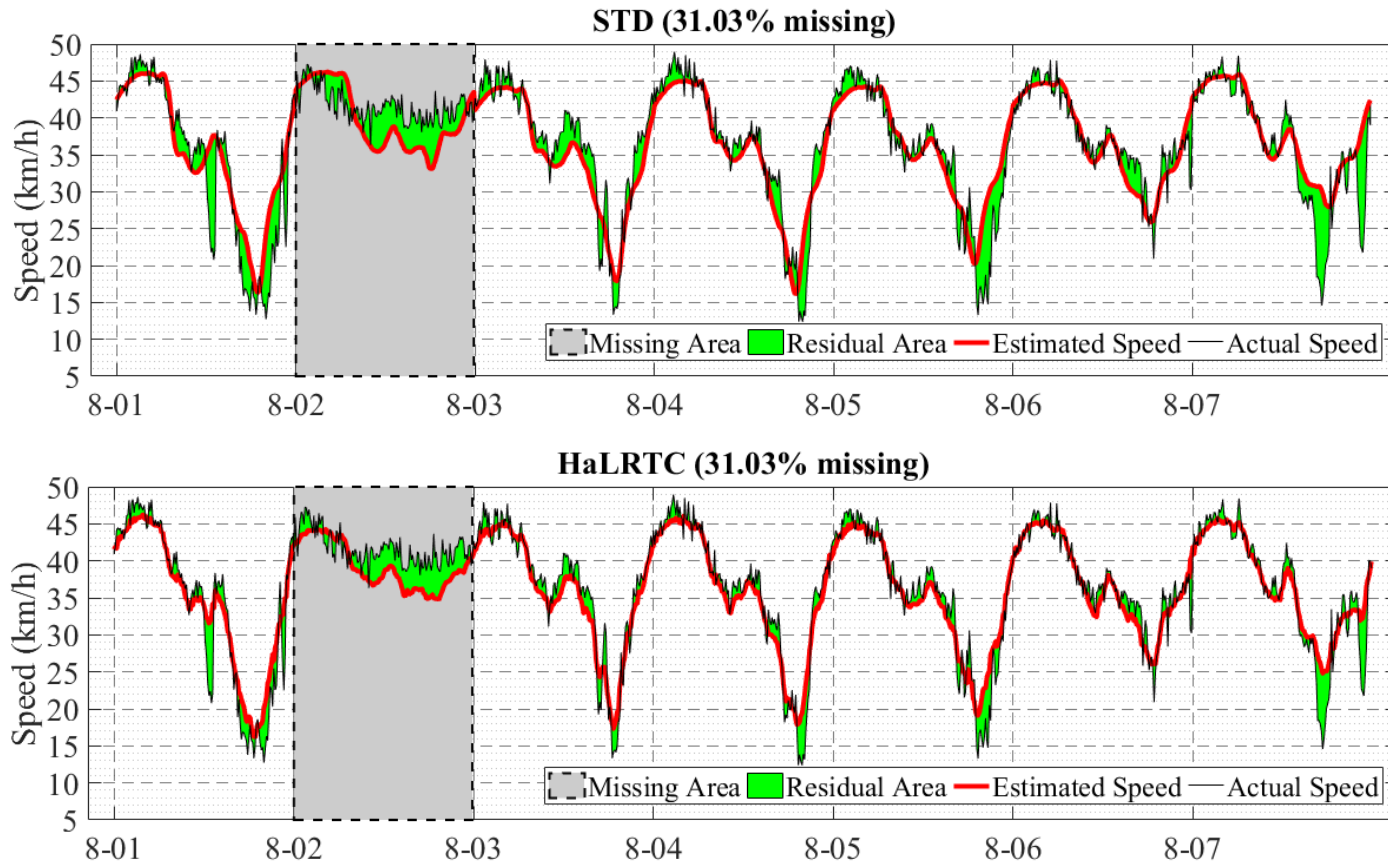


$RMSE = 4.69\text{km/h}$
 $MRE = 8.90\%$



Empirical studies

- **Time-series analysis** (traffic speed of a road segment under *the 30% missing*)

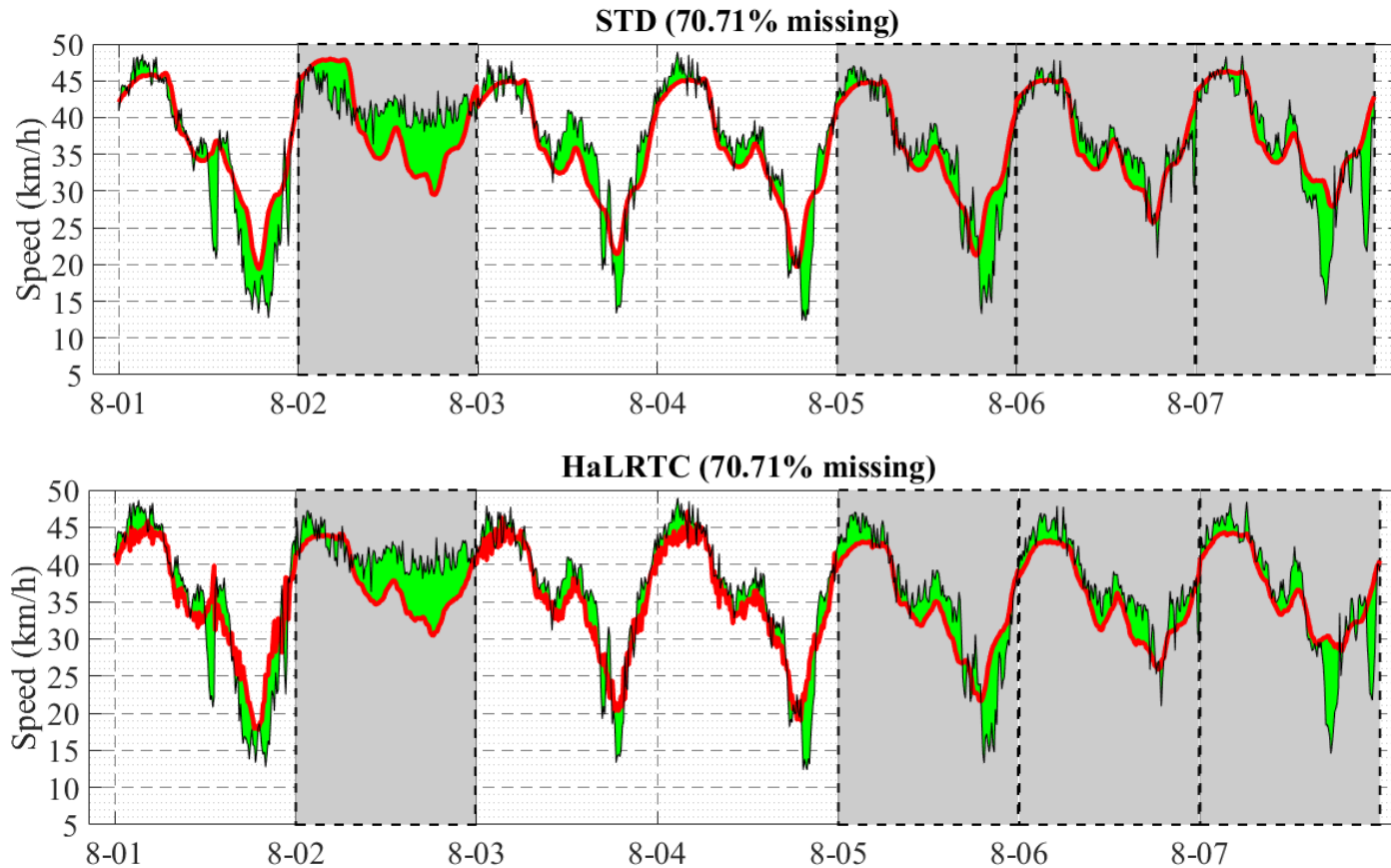


The observations of Aug. 2, 2016 are completely lost



Empirical studies

- **Time-series analysis** (traffic speed of a road segment under *the 70% missing*)

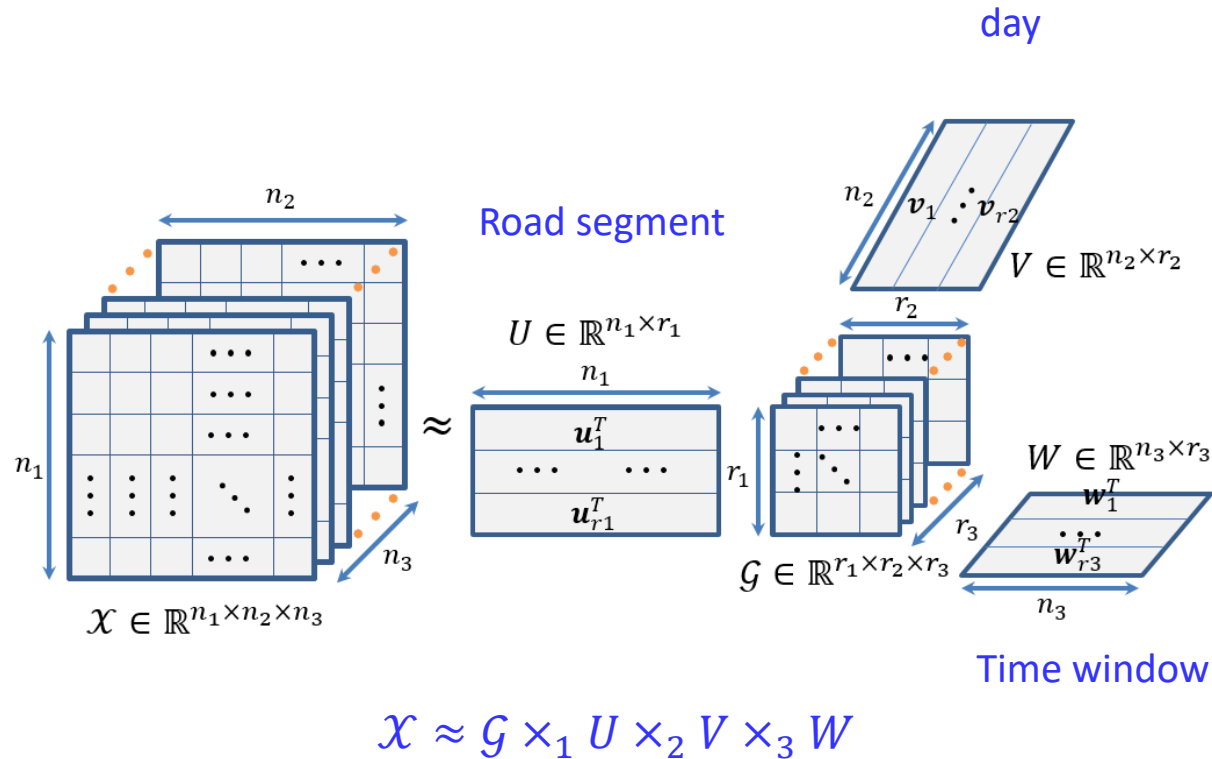


The observations of four days are completely lost



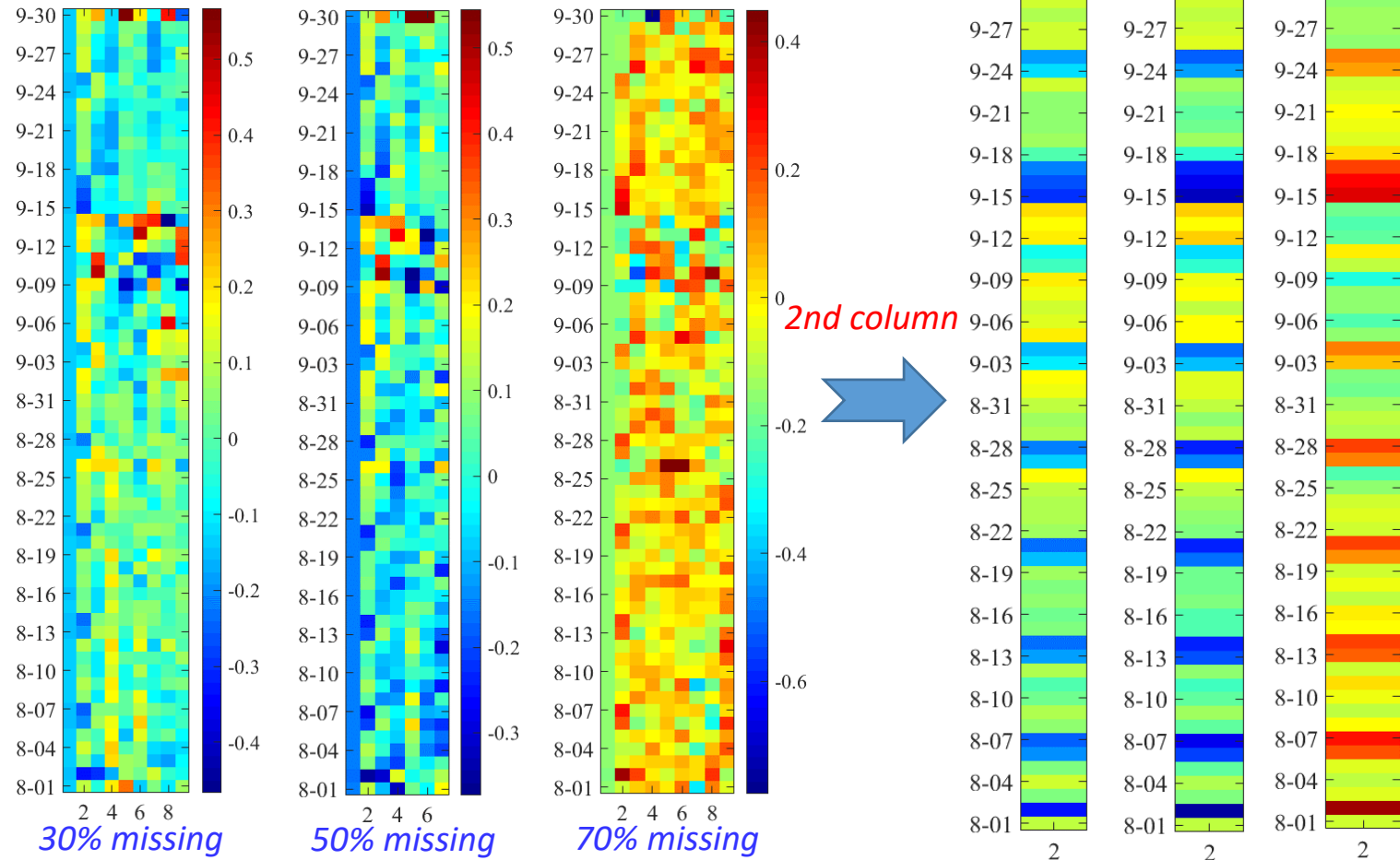
Empirical studies

- **Interpretability** (*Tucker decomposition*)



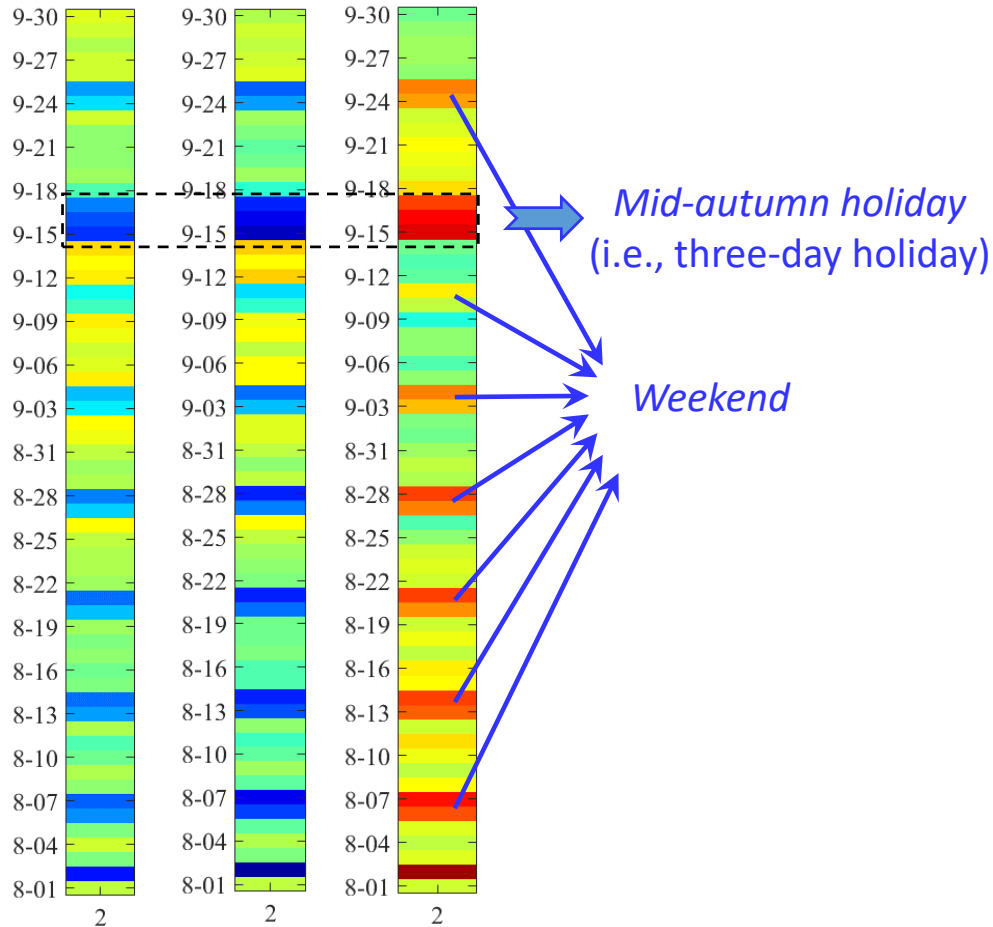
Empirical studies

- Interpretation** (factor matrix $V \in \mathbb{R}^{n_2 \times r_2}$)



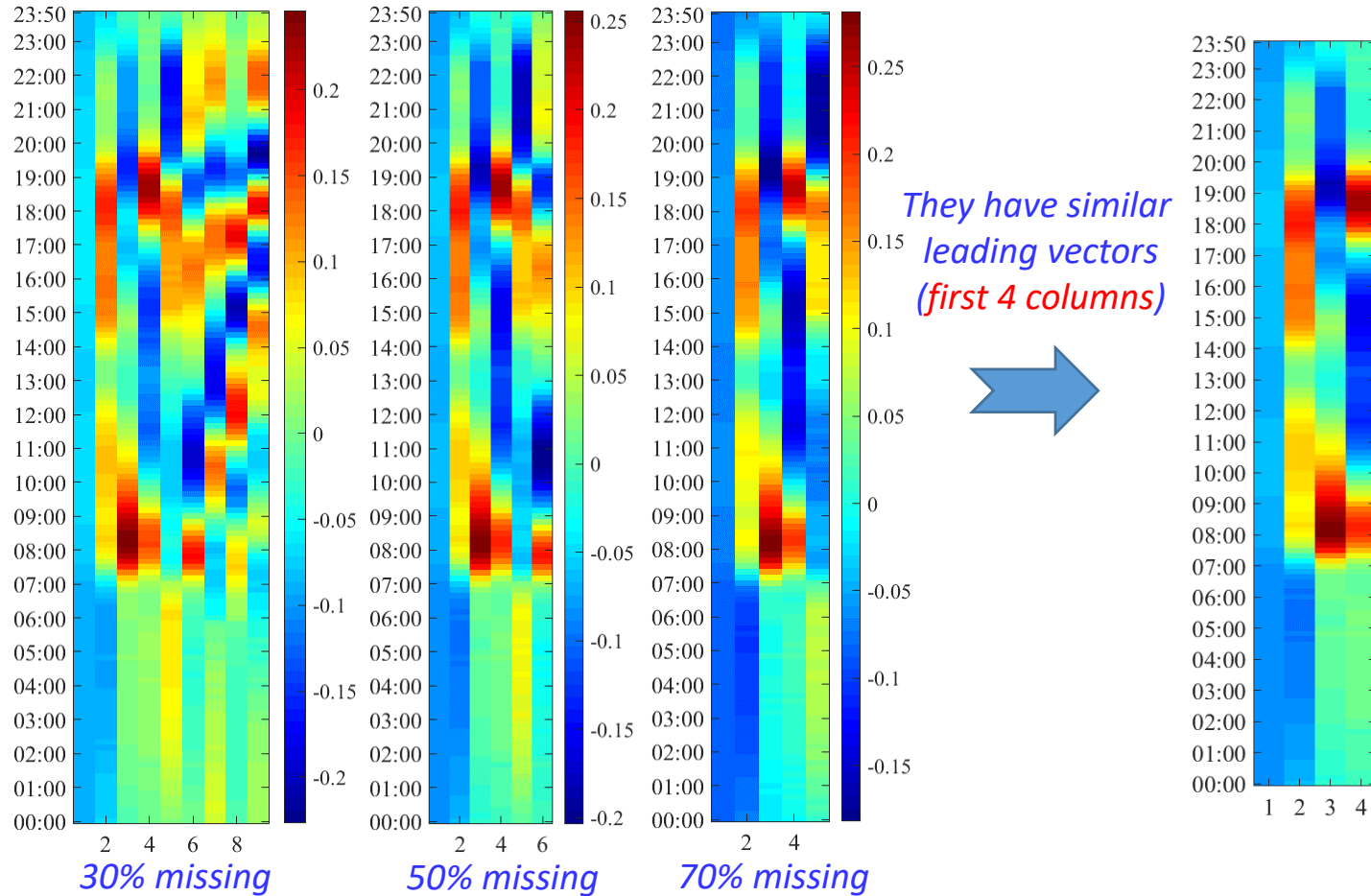
Empirical studies

- Interpretation** (factor matrix $V \in \mathbb{R}^{n_2 \times r_2}$)



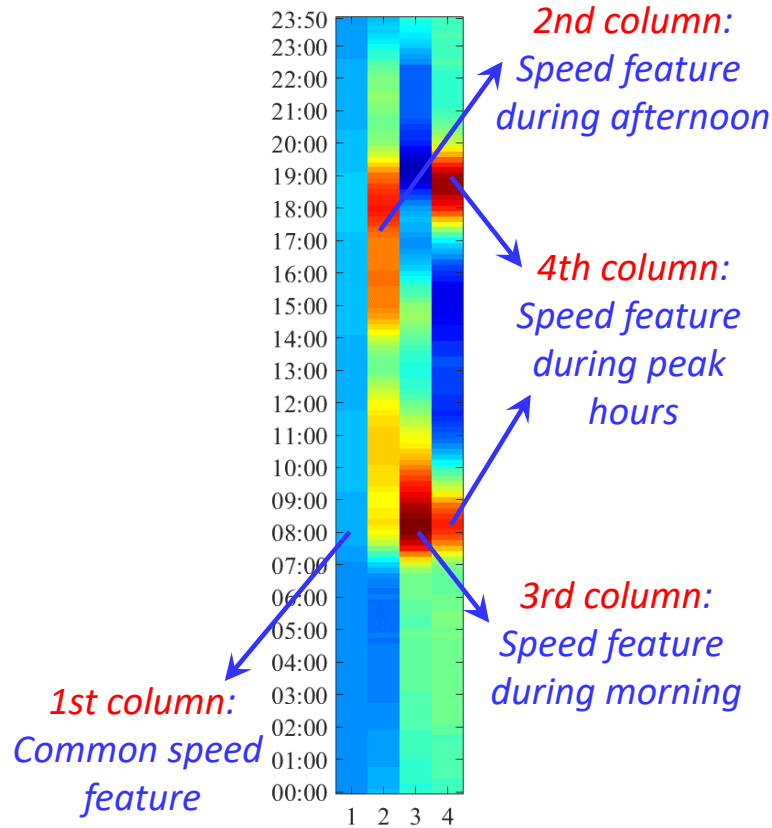
Empirical studies

- Interpretation** (factor matrix $W \in \mathbb{R}^{n_3 \times r_3}$)



Empirical studies

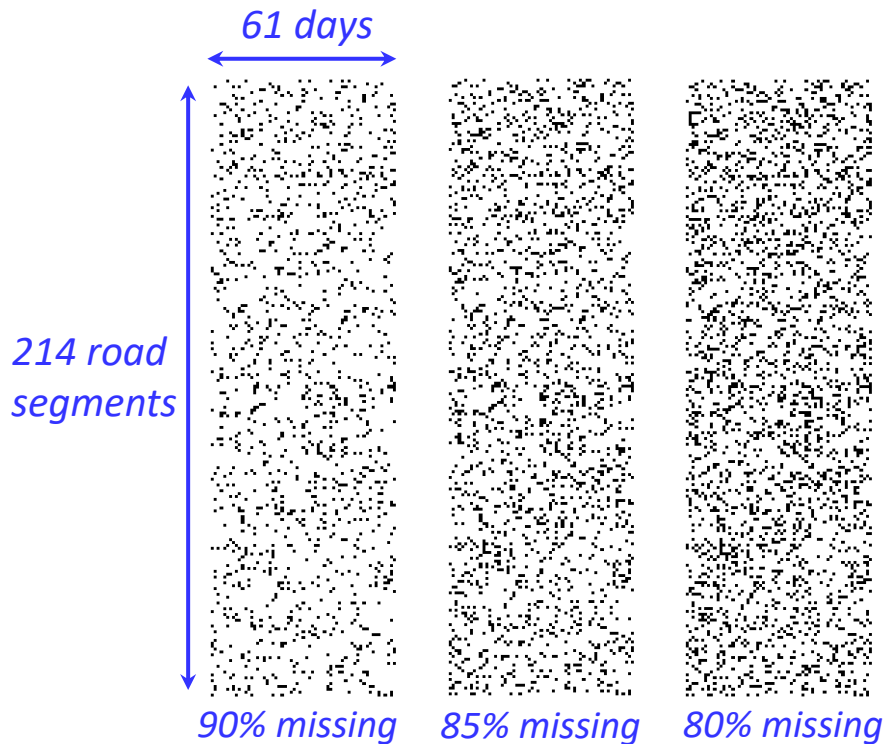
- **Interpretation** (factor matrix $W \in \mathbb{R}^{n_3 \times r_3}$)



- The result also indicate that **7:00~9:00** is *morning peak hours*, and **17:30~19:30** is *afternoon peak hours*.

Empirical studies

- **Deal with heavy missing** (*missing rates are 80%, 85% and 90%*)
 - *Two requirements*: during 61 days, **observations of each road segment cannot be completely lost**; for each day, **at least one speed of road segment is observed**.
 - $S \in \mathbb{R}^{214 \times 61}$: the columns and rows must have *at least one non-zero entry*.



The STD outperforms HaLRTC!

<i>missing rate</i>	<i>Measures</i>	<i>STD</i>	<i>HaLRTC</i>
80%	RMSE	5.25	5.56
	MRE	9.04	11.71
85%	RMSE	5.45	6.25
	MRE	9.46	13.95
90%	RMSE	5.58	10.00
	MRE	9.70	26.37



Appendix

- **STD Algorithm** *Decomposition based approximation*

Algorithm: SVD-combined Tensor Decomposition (STD)

1. **Input:** incomplete tensor \mathcal{X} , binary tensor \mathcal{S}
 2. Set the learning rate α , the regularization parameter λ , and ε
 3. $\mathcal{G} = \mathcal{G}^{(SVD)}, U = U^{(SVD)}, V = V^{(SVD)}, W = W^{(SVD)}$
 4. $\mathcal{X}^{(0)} = \mathcal{G} \times_1 U \times_2 V \times_3 W$
 5. $\mathcal{E} = \mathcal{S} * (\mathcal{X} - \mathcal{G} \times_1 U \times_2 V \times_3 W)$
 6. $U^+ = (1 - \alpha\lambda)U + \alpha(\mathcal{S} * \mathcal{E})_{(1)}(W \otimes V)\mathcal{G}_{(1)}^T$
 7. $V^+ = (1 - \alpha\lambda)V + \alpha(\mathcal{S} * \mathcal{E})_{(2)}(W \otimes U)\mathcal{G}_{(2)}^T$
 8. $W^+ = (1 - \alpha\lambda)W + \alpha(\mathcal{S} * \mathcal{E})_{(3)}(V \otimes U)\mathcal{G}_{(3)}^T$
 9. $\mathcal{G}^+ = (1 - \alpha\lambda)\mathcal{G} + \alpha \cdot \mathcal{E} \times_1 U^T \times_2 V^T \times_3 W^T$
 10. Update $U \leftarrow U^+, V \leftarrow V^+, W \leftarrow W^+, \mathcal{G} \leftarrow \mathcal{G}^+$
 11. $\mathcal{X}^{(1)} = \mathcal{G} \times_1 U \times_2 V \times_3 W$
 12. Check the convergence condition, $\|\mathcal{X}^{(1)} - \mathcal{X}^{(0)}\|_F^2 < \varepsilon$
 13. **while** not (convergence) **do**
 14. $\mathcal{X}^{(0)} = \mathcal{X}^{(1)}$, execute step 5-12
 15. **Return** $\mathcal{G}, U, V, W, \hat{\mathcal{X}} = \mathcal{X}^{(1)}$
-



Appendix

- **HaLRTC Algorithm** *Low-rank approximation*

Algorithm: High Accuracy Low Rank Tensor Completion (**HaLRTC**)

1. **Input:** $\hat{\mathcal{X}}$ with $\mathcal{S} * \hat{\mathcal{X}} = \mathcal{S} * \mathcal{X}$ and $(1 - \mathcal{S}) * \hat{\mathcal{X}} = 0$, ρ , K
 2. Set $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as additive tensors with all entries are 0
 3. **for** $k = 0$ to K **do**
 4. $\mathcal{B}_q = \text{fold}_q \left\{ D_{\frac{1}{3\rho}} \left(\hat{\mathcal{X}}_{(q)} + \frac{1}{\rho} \mathcal{Y}_{q(q)} \right) \right\}, q = 1, 2, 3$
 5. $\hat{\mathcal{X}} = (1 - \mathcal{S}) * \left[\frac{1}{3} \sum_{q=1}^3 \left(\mathcal{B}_q - \frac{1}{\rho} \mathcal{Y}_q \right) \right] + \mathcal{S} * \mathcal{X}$
 6. $\mathcal{Y}_q = \mathcal{Y}_q - \rho(\mathcal{B}_q - \hat{\mathcal{X}}), q = 1, 2, 3$
 7. **Output:** $\hat{\mathcal{X}}$
-



Reference

- Acar, E., Dunlavy, D.M., Kolda, T.G., Mørup, M., 2011. Scalable tensor factorizations for incomplete data. *Chenomentr. Intell. Lab. Syst.* 106(1), 41-56.
- Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M., Telgarsky, M., 2014. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* 15(1): 2773-2832.
- Asif, M.T., Kannan, S., Dauwels, J., Jaillet, P., 2013. Data compression techniques for urban traffic data. In: Computational Intelligence in Vehicles and Transportation Systems (CIVTS). *2013 IEEE Symposium. IEEE*, pp. 44-49.
- Asif, M.T., Srinivasan, K., Mitrovic, N., Dauwels, J., Jaillet, P., 2015. Near-lossless compression for large traffic networks. *IEEE Trans. Intell. Transp. Syst.* 16 (4), 1817-1826.
- Asif, M.T., Mitrovic, N., Dauwels, J., Jaillet, P., 2016. Matrix and tensor based methods for missing data estimation in large traffic networks. *IEEE Trans. Intell. Transp. Syst.* 17 (7), 1816-1825.
- Chen, B., Li, Z., Zhang, S., 2015. On optimal low rank Tucker approximation for tensors: the case for an adjustable core size. *J. Glob. Optim.* 62, 811-832.
- Carroll, J.D., Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika* 35, 283-319.
- Duan, Y., Lv, Y., Liu, Y.-L., Wang, F.-Y., 2016. An efficient realization of deep learning for traffic data imputation. *Transp. Res. Part C: Emerg. Technol.* 72, 168-181.
- Farhan, J., 2015. Overview of missing physical commodity trade data and its imputation using data augmentation. *Transp. Res. Part C: Emerg. Technol.* 54, 1-14.
- Golub, G.H., Van Loan, C.F., 2013. Matrix computations. Johns Hopkins Univ. Press, Baltimore, MD.
- Han, Y., Moutarde, F., 2016. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. *Int. J. ITS Res.* 14, 36-49.
- Hitchcock, F.L., 1927a. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.*, 6, 164-189.
- Hichcock, F.L., 1927b. Multiple invariants and generalized rank of a p-way matrix or tensor, *J. Math. Phys.*, 7, 39-79.



Reference

- Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. *SIAM Rev.* 51(3), 455-500.
- Koren, Y., Bell, R.M., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. *IEEE Comput.* 42 (8), 30-37.
- Lathauwer, L.D., Moor, B.D., Vandewalle, J., 2000. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* 20(4): 1253-1278.
- Li, L., Li, Y., Li, Z., 2013. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. *Transp. Res. Part C: Emerg. Technol.* 34, 108-120.
- Li, L., Su, X., Zhang, Y., Lin, Y., Li, Z., 2015. Trend modeling for traffic time series analysis: An integrated study. *IEEE Trans. Intell. Transp. Syst.* 16 (6), 3430-3439.
- Li, X., Li, M., Gong, Y.-J., Zhang, X.-L., Yin, J., 2016. T-DesP: Destination prediction based on big trajectory data. *IEEE Trans. Intell. Transp. Syst.* 17 (8), 2344-2354.
- Li, Y., Li, Z., Li, L., 2014. Missing traffic data: comparison of imputation methods. *IET Intel. Transp. Syst.* 8 (1), 51-57.
- Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to information retrieval. Cambridge University Press, Cambridge, UK.
- Paterek, A., 2007. Improving regularized singular value decomposition for collaborative filtering. In: *Proceeding of KDD Cup and Workshop*, pp. 5-8.
- Qu, L., Li, L., Zhang, Y., Hu, J., 2009. PPCA-based missing data imputation for traffic flow volume: A systematical approach. *IEEE Trans. Intell. Transp. Syst.* 10 (3), 512-522.
- Ran, B., Tan, H., Wu, Y., Jin, P.J., 2016. Tensor based missing traffic data completion with spatial-temporal correlation. *Physica A* 446, 54-63.
- Robinson, S., Narayanan, B., Toh, N., Pereira, F., 2014. Methods for pre-processing smartcard data to improve data quality. *Transp. Res. Part C: Emerg. Technol.* 49, 43-58.
- Schifanella, C., Candan, K.S., Sapino, M.L., 2014. Multiresolution tensor decompositions with mode hierarchies. *ACM Trans. Knowl. Discovery from Data* 8(2), 10.



Reference

- Soriguera, F., 2016. Highway travel time estimation with data fusion. Springer, Berlin.
- Sun, L., Axhausen, K.W., 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transp. Res. Part B: Methodol.* 91, 511-524.
- Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., Li, F., 2013. A tensor-based method for missing traffic data completion. *Transp. Res. Part C: Emerg. Technol.* 28, 15-27.
- Tan, H., Feng, J., Chen, Z., Yang, F., Wang, W., 2014. Low multilinear rank approximation of tensors and application in missing traffic data. *Adv. Mech. Eng.* 2014.
- Tan, H., Wu, Y., Shen, B., Jin, P.J., Ran, B., 2016. Short-term traffic prediction based on dynamic tensor completion. *IEEE Trans. Intell. Transp. Syst.* 17 (8), 2123-2133.
- Tucker, L., 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31(3), 279-311.
- Wang, Y., Zheng, Y., Xue, Y., 2014. Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *KDD'14. ACM*, pp. 374-383.

*Recovering large-scale incomplete traffic speed
via tensor completion*

Thanks for your listening!