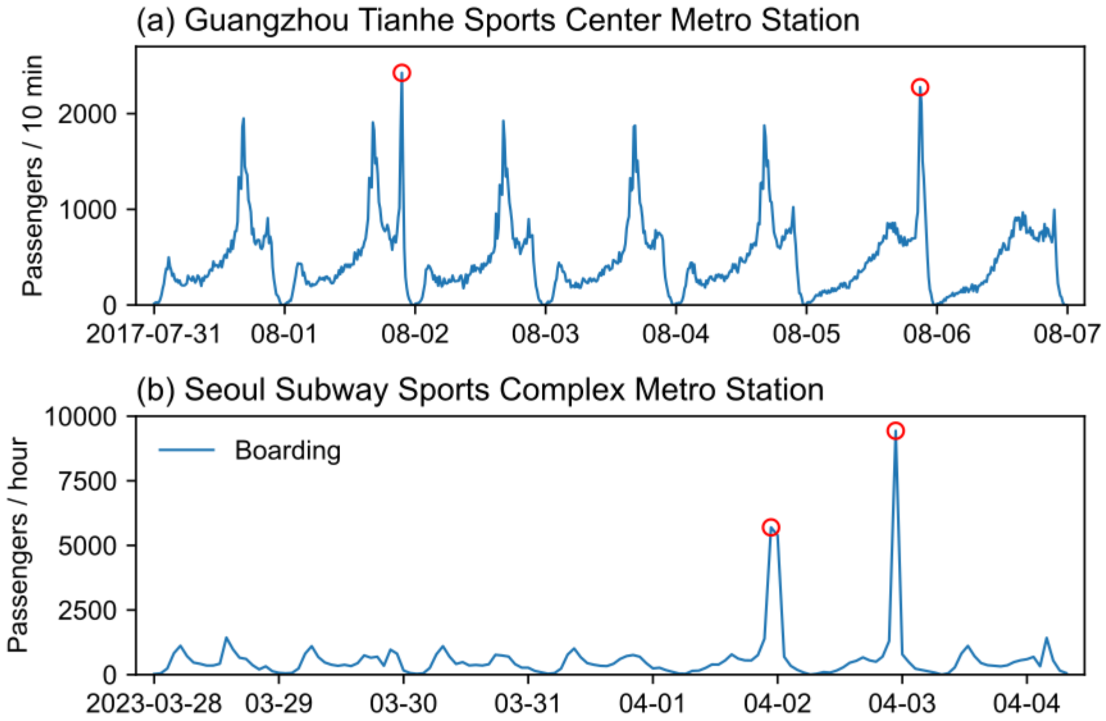
Anomalies in metro passenger demand are predictable --- learning causality with ABTransformer

Transit Data, 1st July 2024

Zhanhong Cheng
Postdoc at Department of Civil Engineering, McGill University
zhanhong.cheng@mcgill.ca

Work conducted with Jiawei Wang, Martin Trépanier and Lijun Sun

Example: abnormal metro demand forecasting



Irregular surges of metro passenger boarding demand.

Forecast abnormal boarding demand

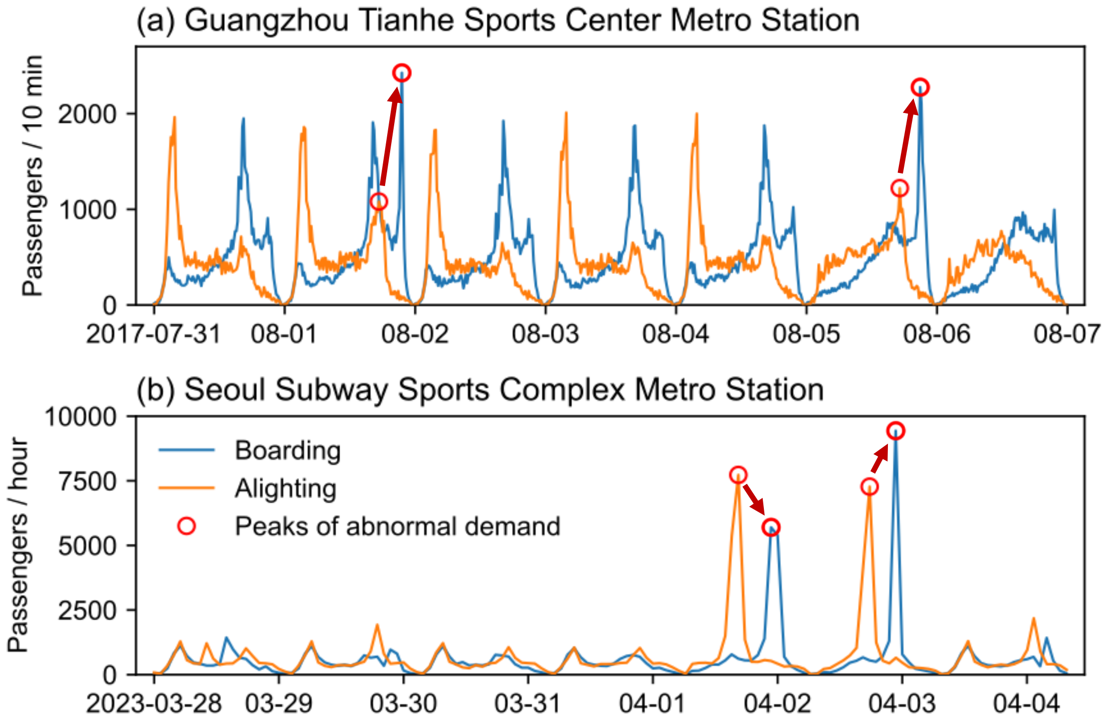
Important:

- It helps operators increase supply and prevent dangers.
- It helps passengers plan their trips.

Difficult:

- Irregular, occasional, often abrupt;
- Require to predict far in advance (hours).

Example: abnormal metro demand forecasting



Irregular surges of metro passenger boarding demand.

Forecast abnormal boarding demand

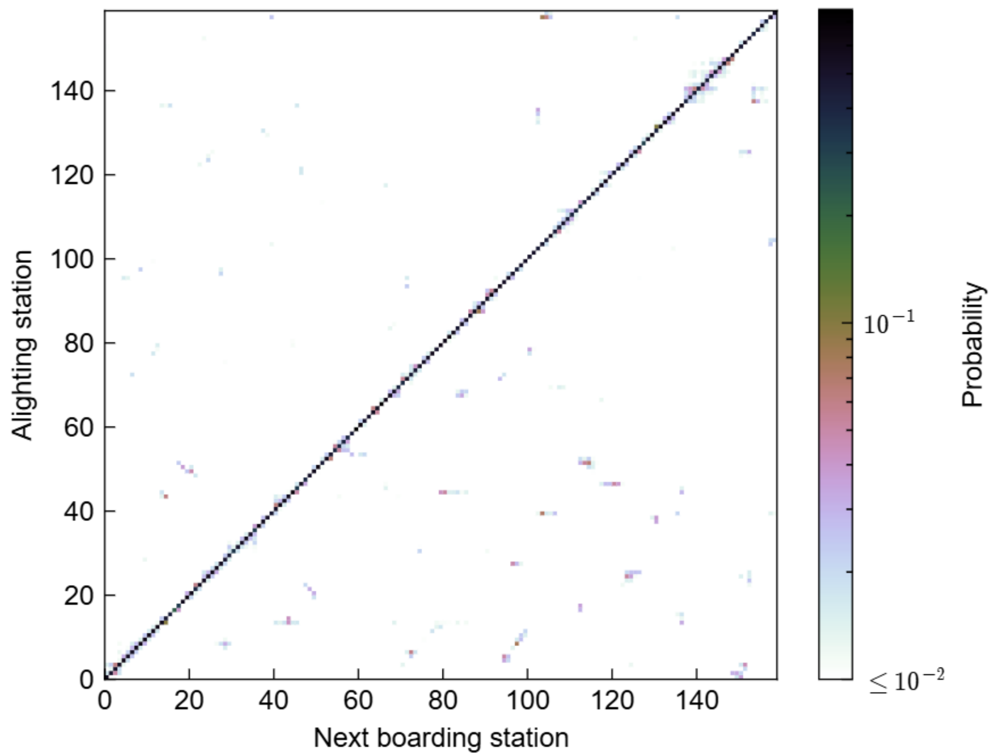
Important:

- It helps operators increase supply and prevent dangers.
- It helps passengers plan their trips.

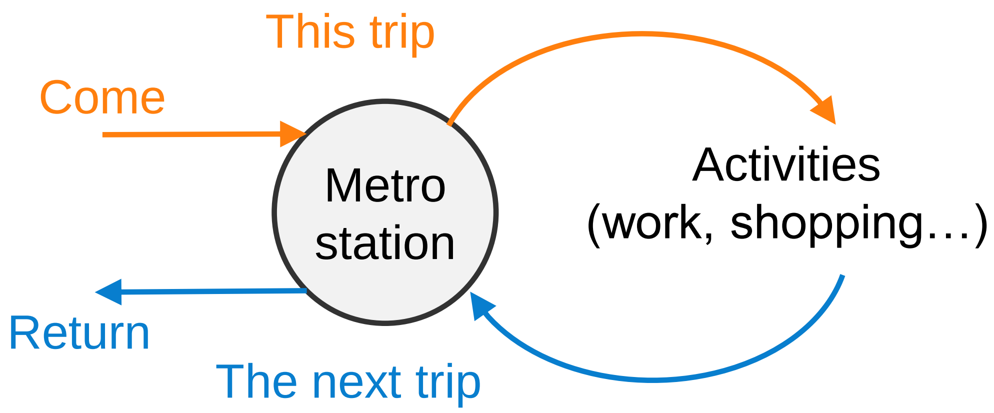
Difficult:

- Irregular, occasional, often abrupt;
- Require to predict far in advance (hours).

The "destination-origin" matrix in metros



The probability of the next boarding station given the current alighting station. (In Guangzhou metro smart card data July 21-28, 2017.)



Two findings

- The next boarding station is often the previous alighting station.
- Long-range correlations and causalities exist in the station boarding and alighting passenger flow.

Outlines

Goal: forecasting method for metro station passenger demand under abnormal events.

Part 1. Abnormal passenger demand identification

For better evaluation.

Part 2. Joint forecasting of boarding and alighting flows

The ABTransformer and other approaches.

Part 3. Results and analyses

How the joint forecast approach works.

Part 4. Discussion & conclusion

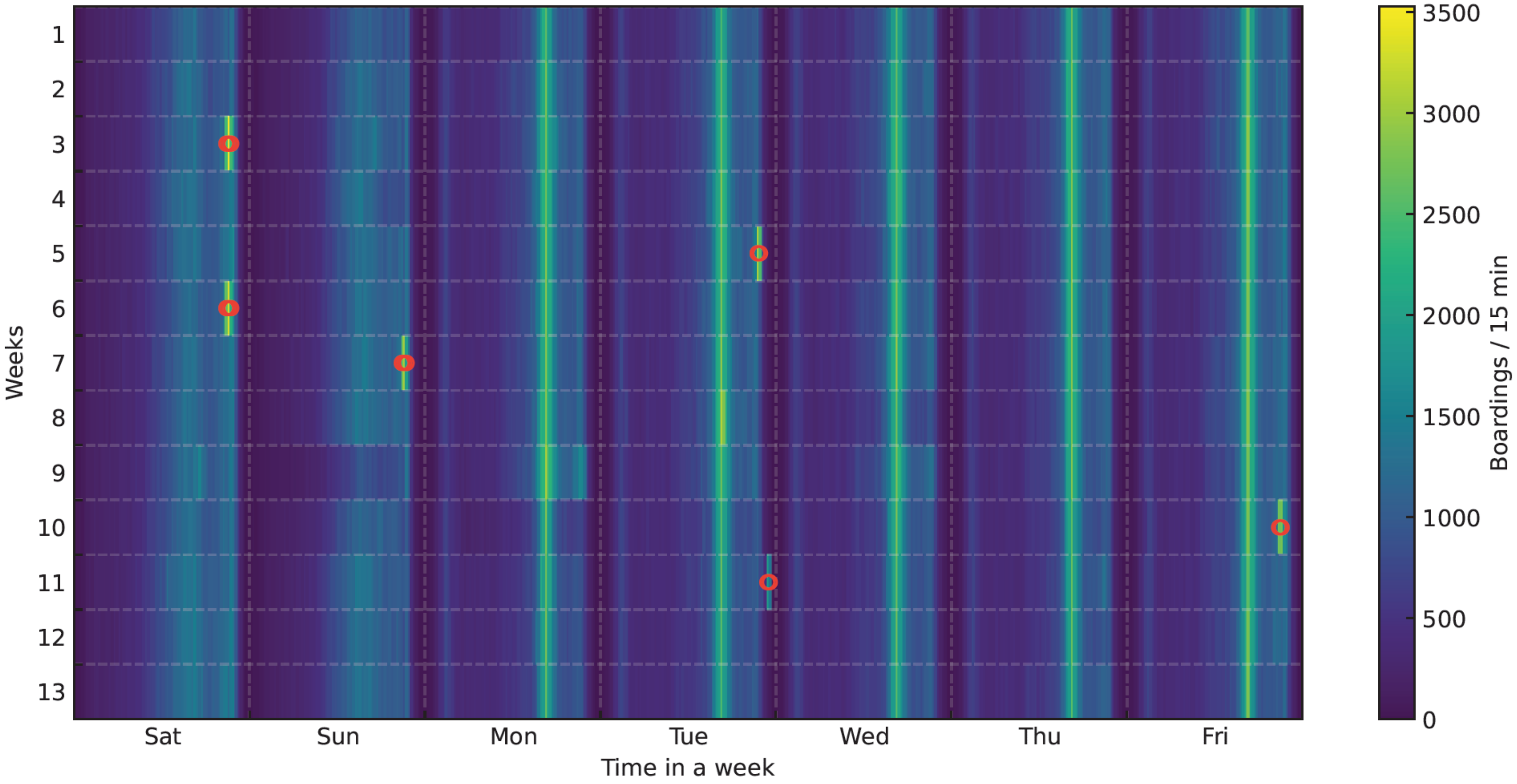
Abnormal passenger demand identification

We use Robust Principal Component Analysis (RPCA) [1] to detect anomalies in passenger flow.

A modification of PCA that works well on data contain outliers.

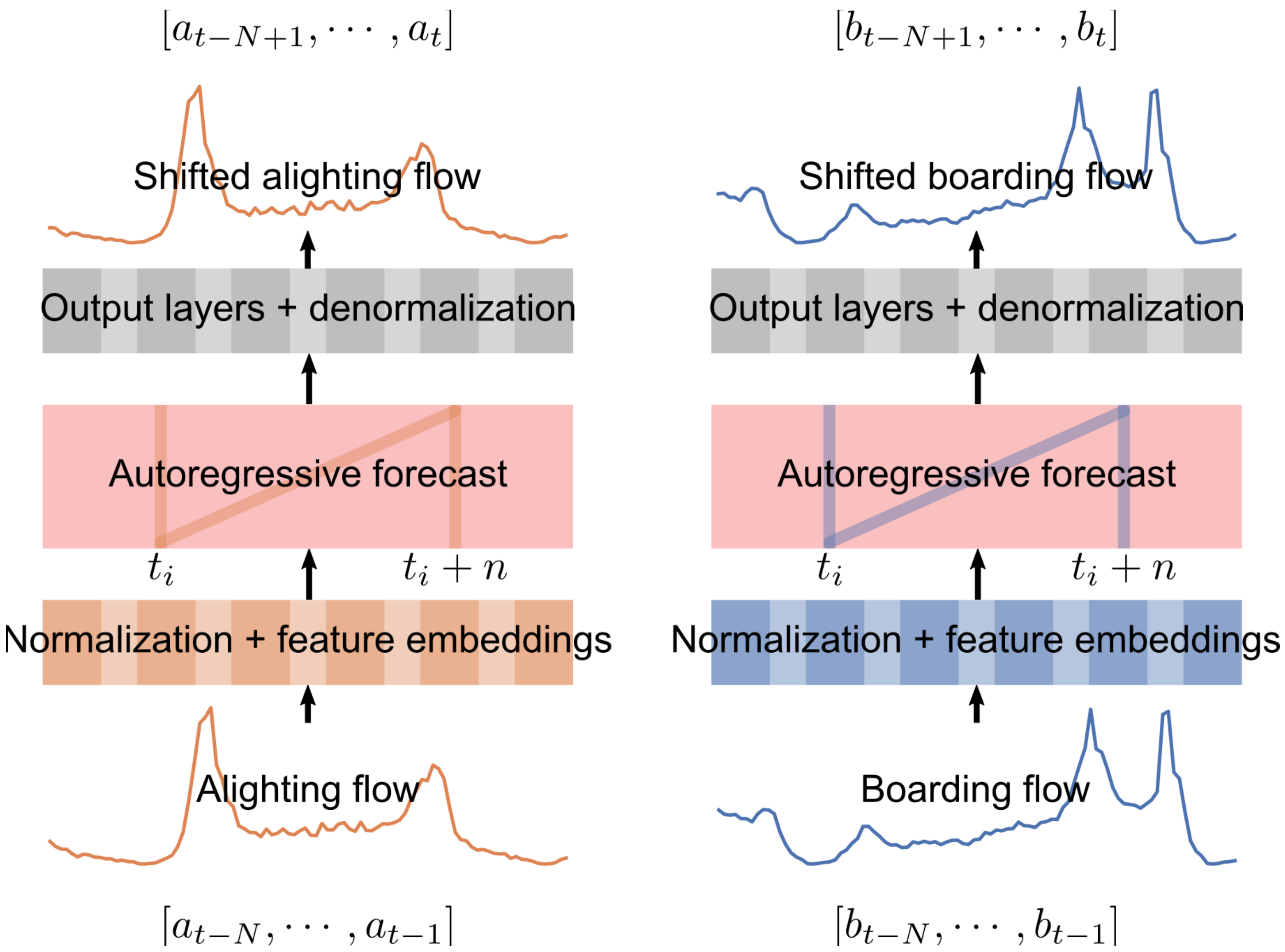
$$\begin{array}{ll} \text{minimize} & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} & L + S = M \end{array}$$

[1] Wright, J., Ganesh, A., Rao, S., Peng, Y., & Ma, Y. (2009). Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. Advances in neural information processing systems, 22.



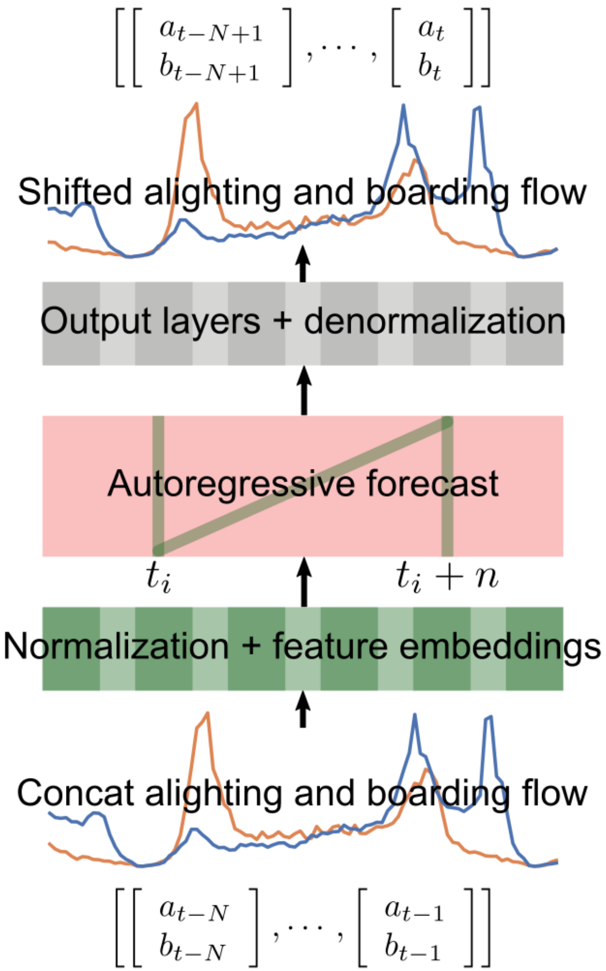
The abnormal boarding demand (marked in red circles) identified by RPCA for the Guangzhou Tianhe Sports Center metro station.

Independent forecast of alighting and boarding flow



- Build independent forecasting models for boarding demand and alighting flow.
- Boarding demand is more important, and the alighting flow forecast is sometime unnecessary.
- Failed to leverage the long-range A&B correlations.

Joint forecasting of boarding and alighting flows (Approach 1)



Approach 1: multivariate forecasting

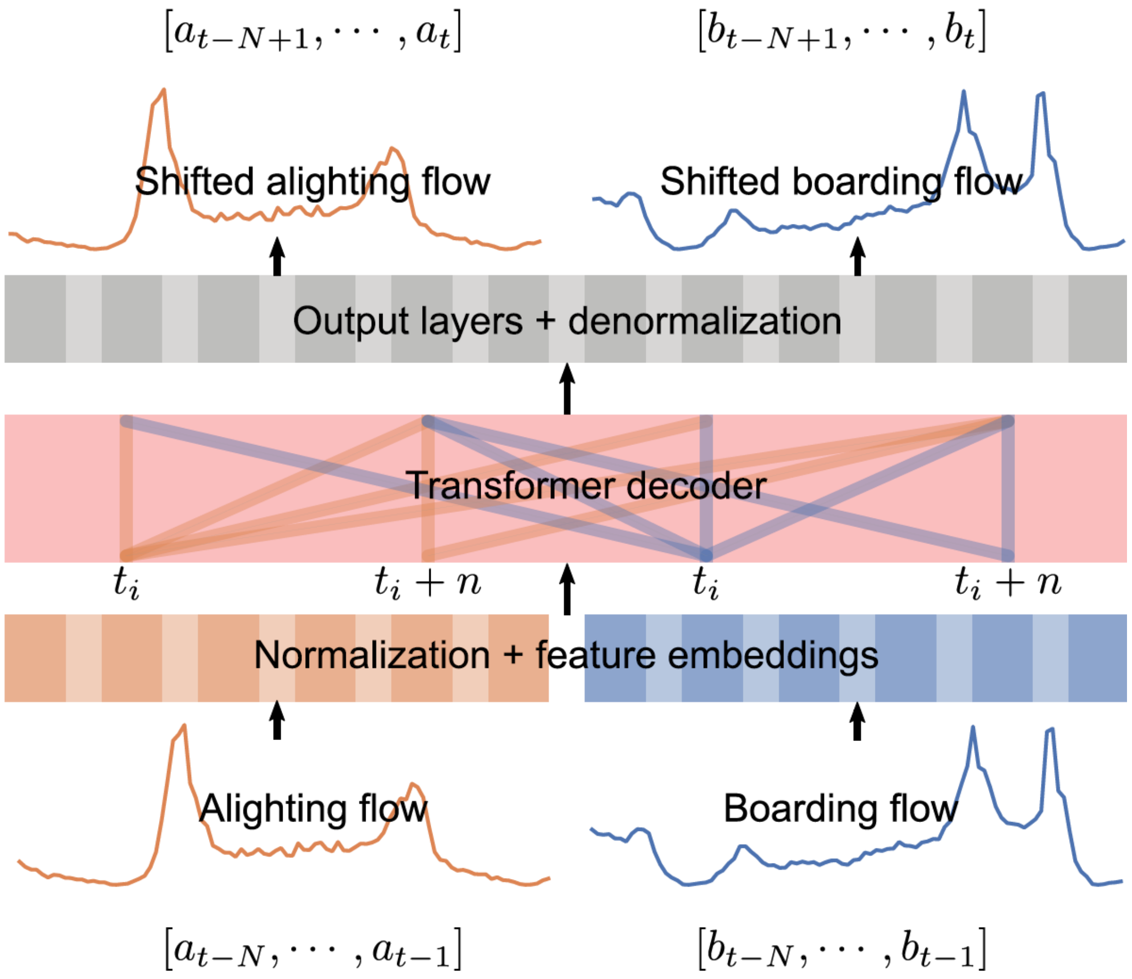
Key argument

- We should **always** forecast the alighting and boarding flow together.

Method

- Using a multivariate forecasting model.
- Many models, such as LSTM and VAR, can be used as the autoregressive core.

Joint forecasting of boarding and alighting flows (Approach 2)



Key argument

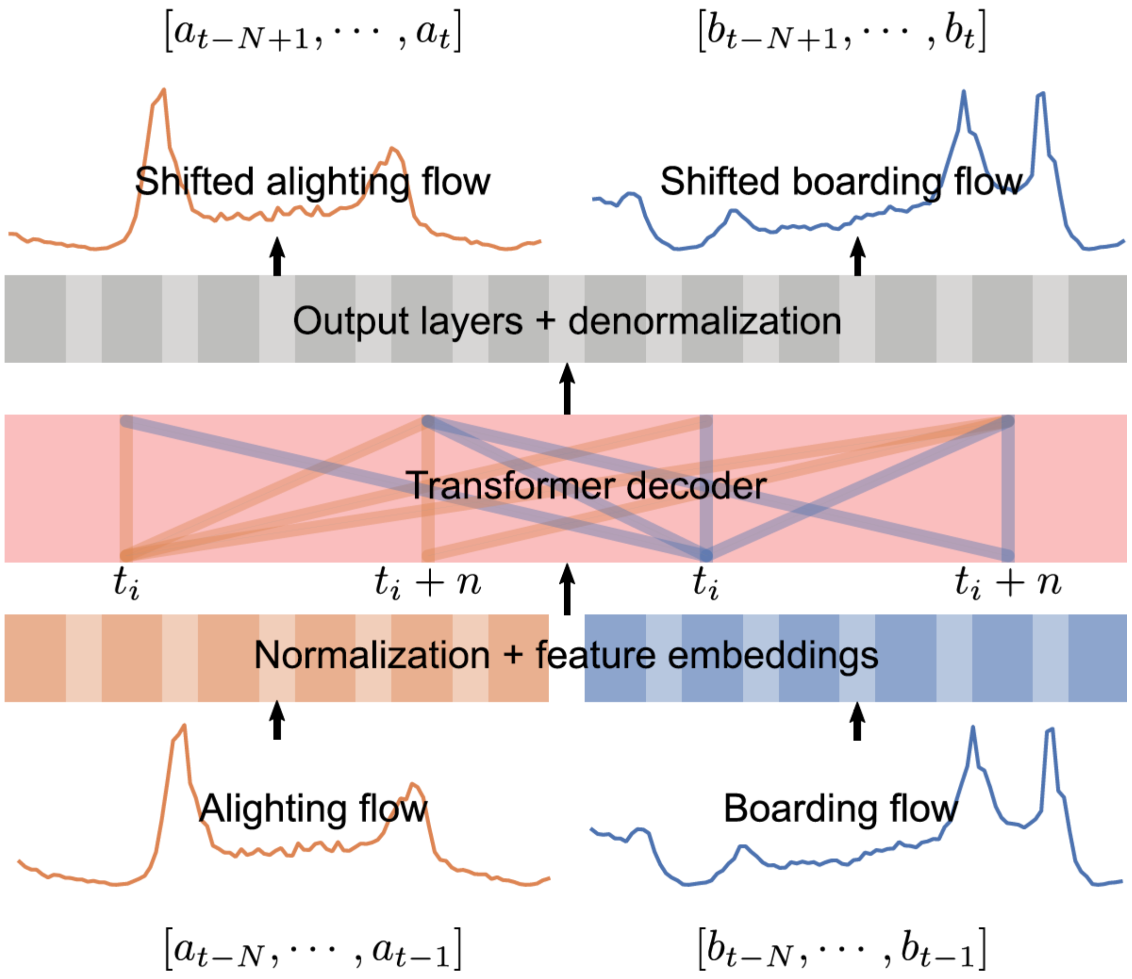
- We should **always** forecast the alighting and boarding flow together.

Method

- Explicitly models A&B correlations.
- Good interpretability, but conventional models do not work.
- Transformer[2] is an ideal way to do it.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Rotary Position Embedding, RoPE[3]



Absolute positional embedding

$$f_{t:t\in\{q,k,v\}}(\boldsymbol{x}_i, i) := \boldsymbol{W}_{t:t\in\{q,k,v\}}(\boldsymbol{x}_i + \boldsymbol{p}_i)$$

$$\begin{cases} \boldsymbol{p}_{i,2t} & = \sin(k/10000^{2t/d}) \\ \boldsymbol{p}_{i,2t+1} & = \cos(k/10000^{2t/d}) \end{cases}$$

RoPE (a relative positional embedding)

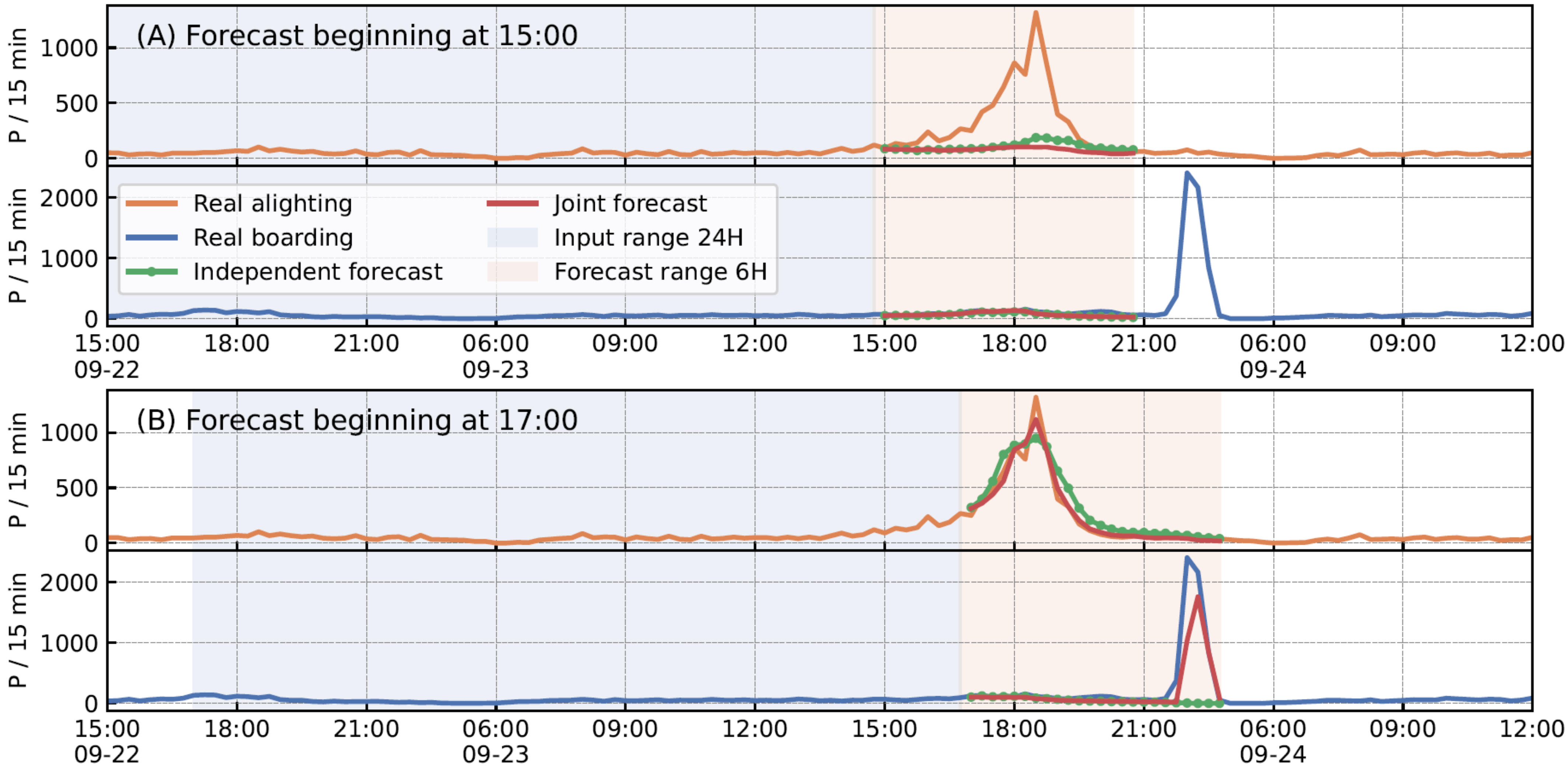
$$f_q(\boldsymbol{x}_m, m) = (\boldsymbol{W}_q \boldsymbol{x}_m) e^{im\theta},$$
$$f_k(\boldsymbol{x}_n, n) = (\boldsymbol{W}_k \boldsymbol{x}_n) e^{in\theta}.$$

$$\boxed{\langle f_q(\boldsymbol{x}_m, m), f_k(\boldsymbol{x}_n, n)\rangle = g(\boldsymbol{x}_m, \boldsymbol{x}_n, m-n)}$$

- We suspect the A&B correlation depends on their relative positions.

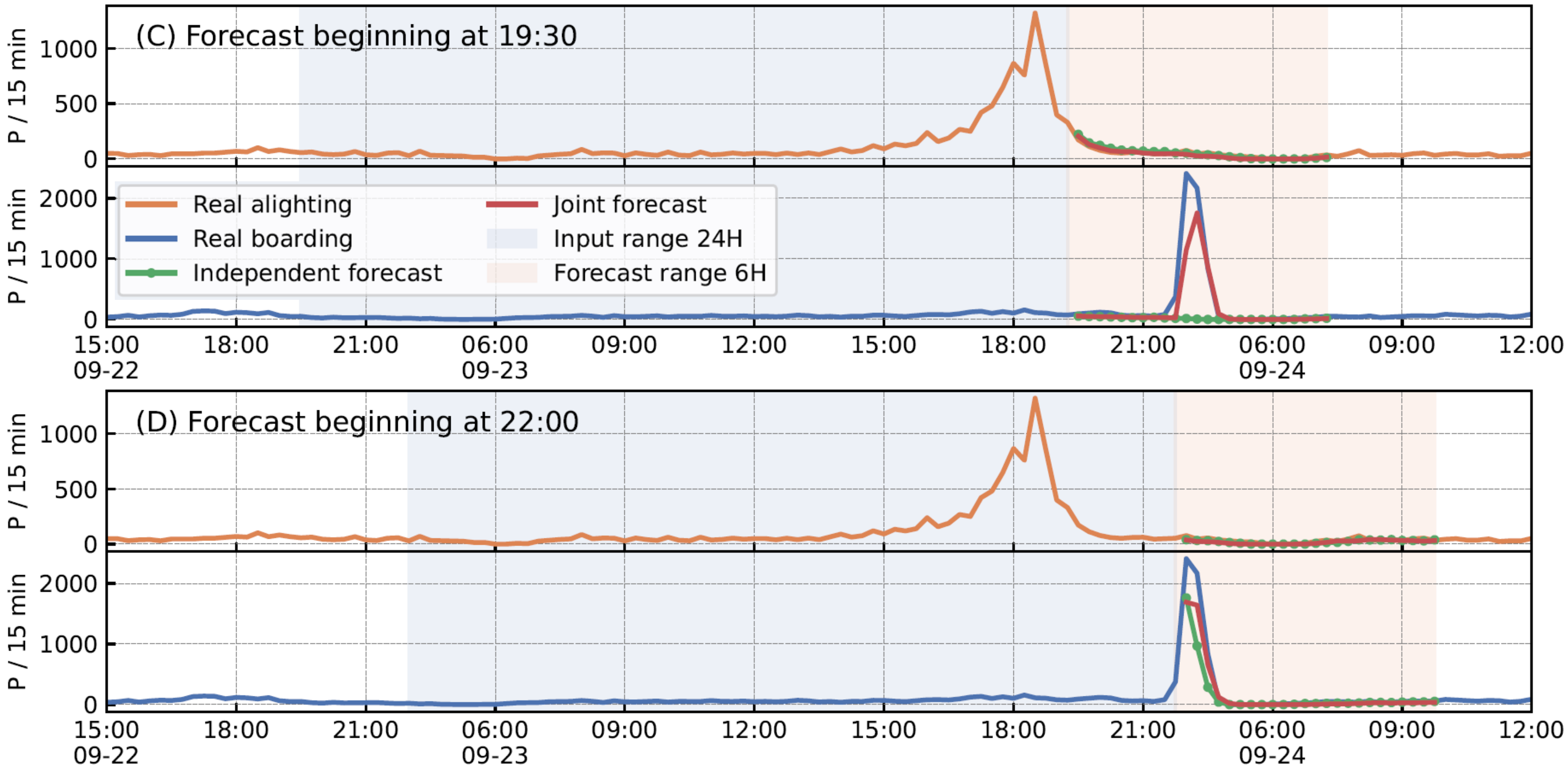
[3] Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., & Liu, Y. (2024). Roformer: Enhanced transformer with rotary position embedding. Neurocomputing, 568, 127063.

Joint vs independent forecast



The forecast of Guangzhou Luogang station

Joint vs independent forecast



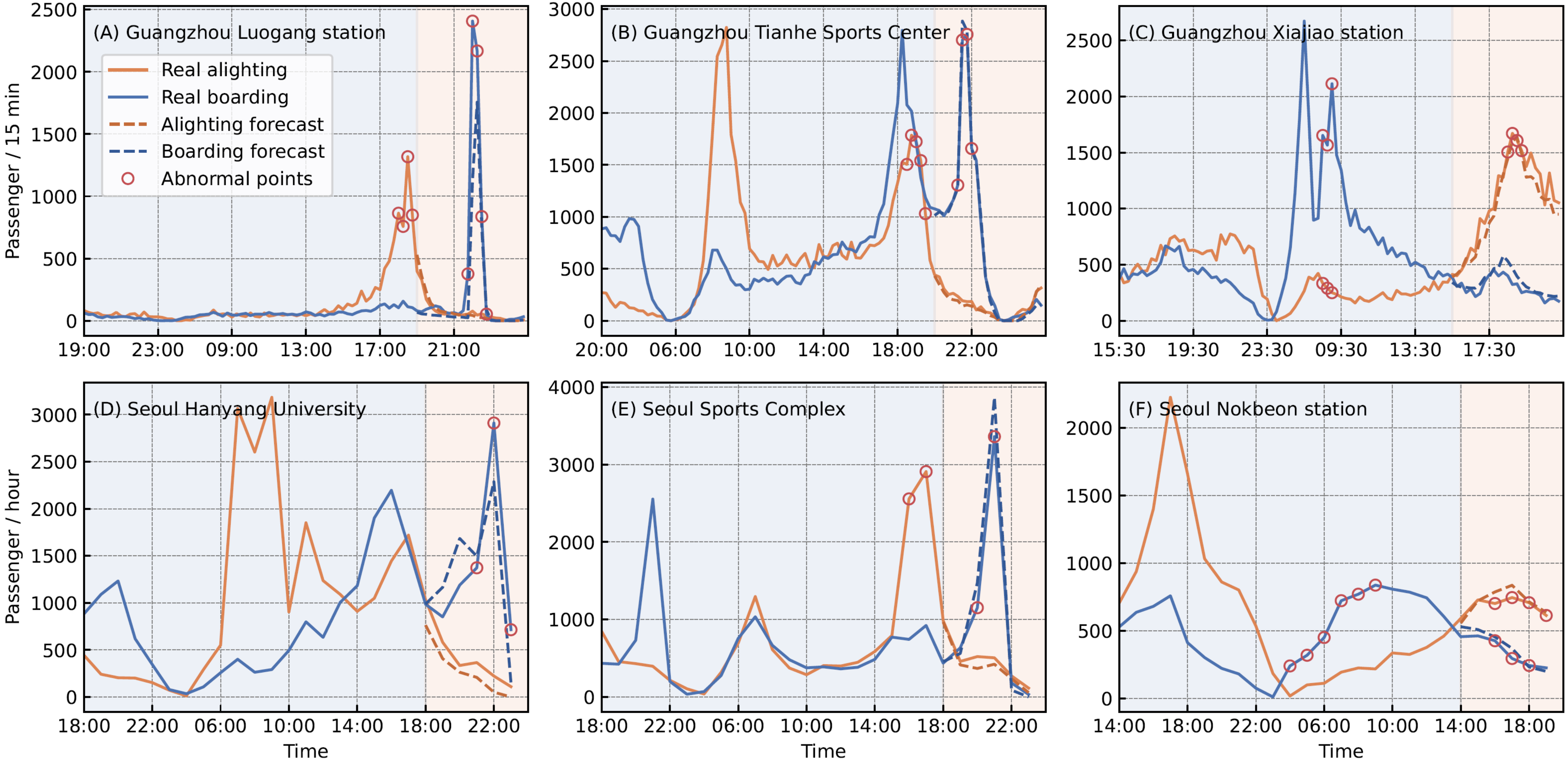
The forecast of Guangzhou Luogang station

Quantitative results

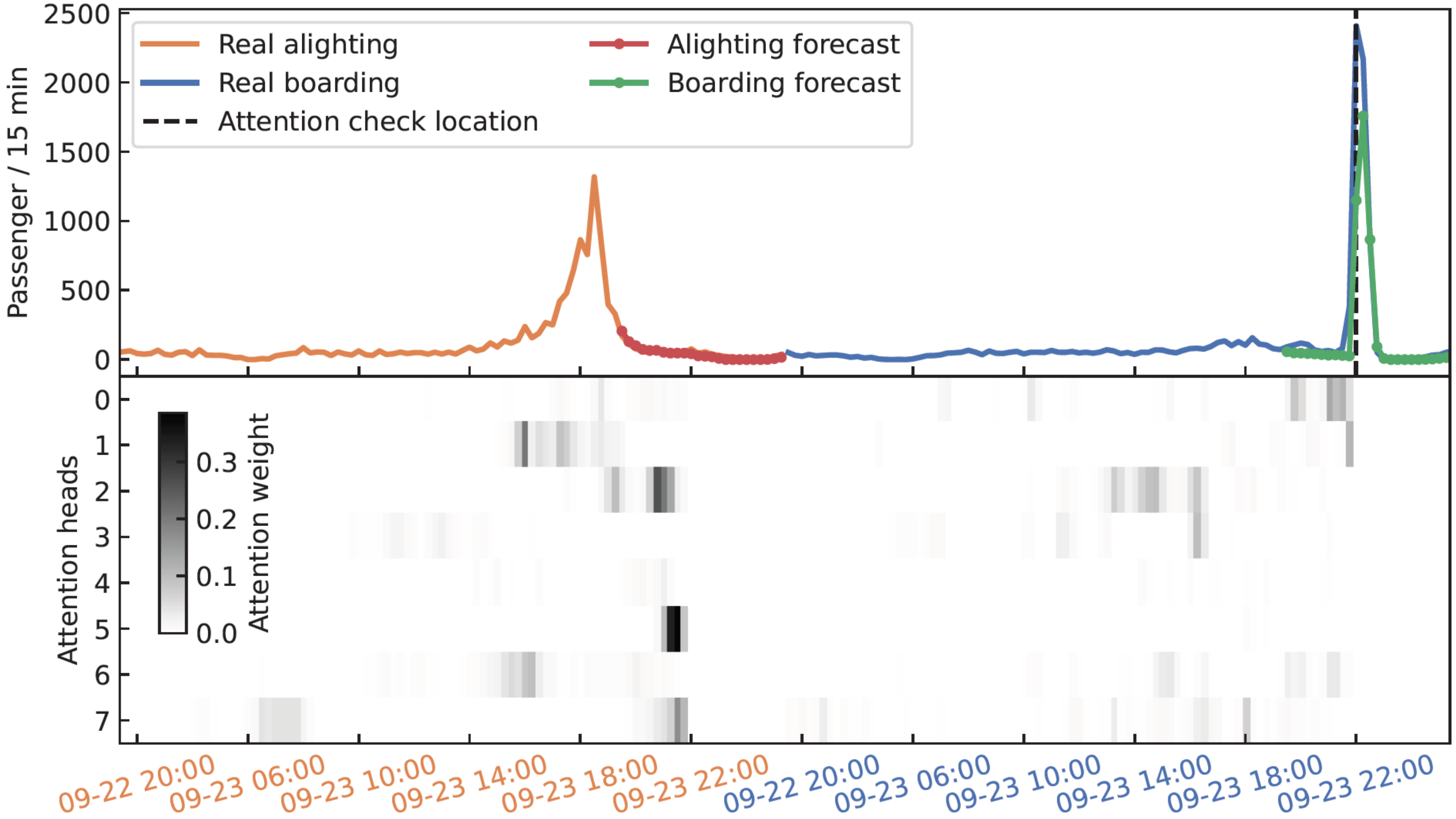
Forecast performance in Guangzhou and Seoul Metros

		Guangzhou						Seoul					
		Inflow			Outflow			Inflow			Outflow		
		Normal	Abnormal	Total	Normal	Abnormal	Total	Normal	Abnormal	Total	Normal	Abnormal	Total
LSTM	MAE	43.1	393.6	43.3	43.6	343.3	43.7	**44.6**	435.0	**46.1**	51.2	467.3	**52.4**
joint 1	WMAPE	10.7%	41.4%	10.8%	10.9%	35.5%	10.9%	**5.5%**	33.1%	**5.7%**	6.3%	36.4%	**6.5%**
LSTM	MAE	48.8	428.1	49.0	41.7	352.1	41.9	48.4	483.6	50.2	51.8	473.2	53.0
independent	WMAPE	12.1%	45.0%	12.2%	10.4%	37.0%	10.4%	6.0%	36.4%	6.2%	6.4%	36.6%	6.6%
Transformer	MAE	**40.2**	366.6	**40.4**	**40.5**	**319.5**	**40.6**	47.2	428.9	48.7	52.1	**442.1**	53.2
join 1	WMAPE	**10.0%**	38.6%	**10.0%**	**10.1%**	**33.0%**	**10.1%**	5.9%	32.4%	6.0%	6.5%	**34.4%**	6.6%
Transformer	MAE	41.7	**352.1**	41.9	42.8	348.1	42.9	48.5	**404.5**	49.9	53.1	458.1	54.2
join 2	WMAPE	10.4%	**37.0%**	10.4%	10.7%	36.0%	10.7%	6.0%	**30.5%**	6.2%	6.6%	35.4%	6.7%
Transformer	MAE	48.8	424.8	49.1	46.5	363.0	46.5	52.1	488.8	53.8	54.8	461.1	56.0
independent	WMAPE	12.2%	44.7%	12.2%	11.6%	37.6%	11.6%	6.5%	36.8%	6.7%	6.8%	35.6%	6.9%

More examples of joint forecast

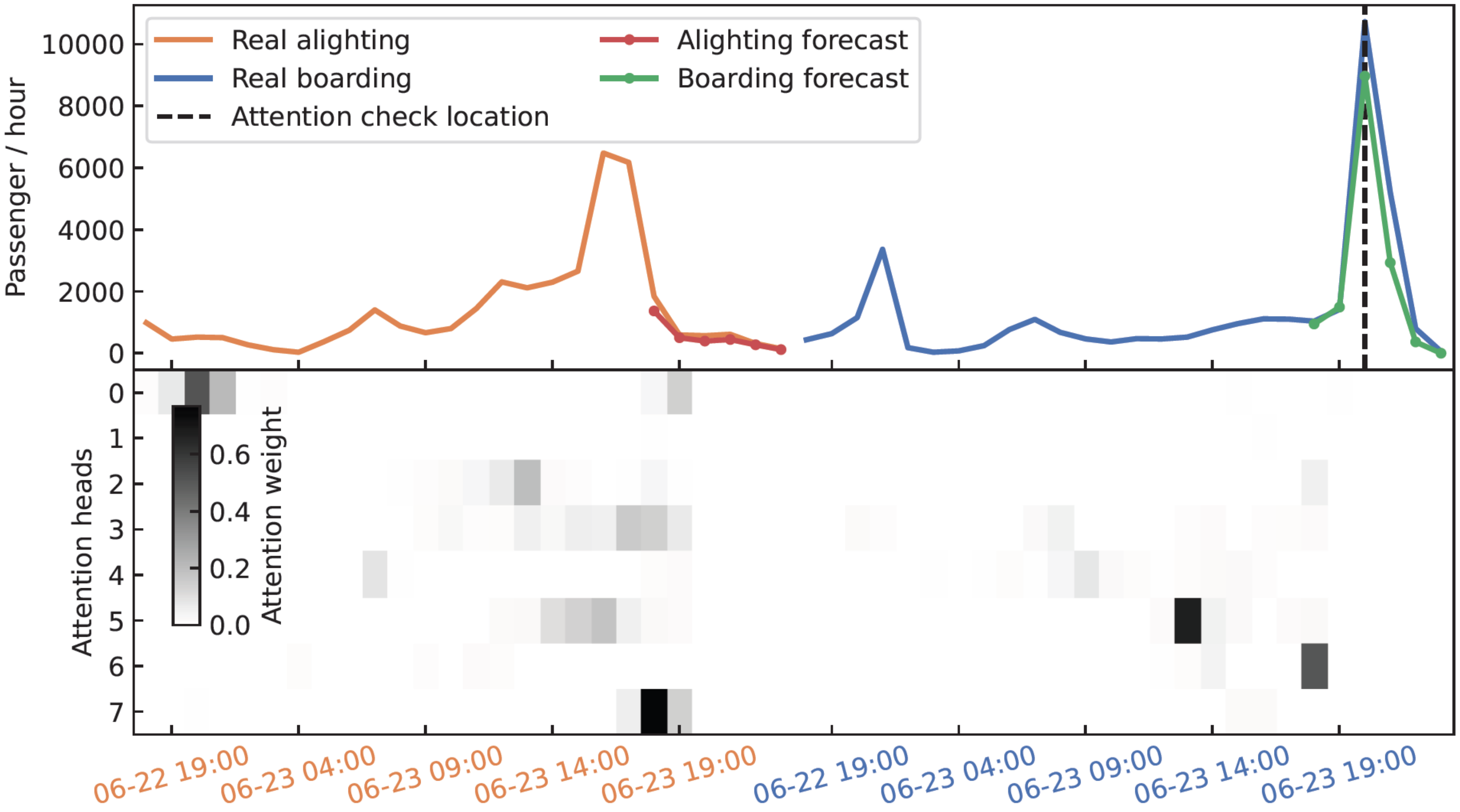


Interpreting forecast results



The attention weights in the forecast of Guangzhou Luogang station

Interpreting forecast results



The attention weights in the forecast of Seoul sports complex

Discussion

Contributions

- We propose a simple and effective way (**always use the joint forecast**) to boost the metro demand forecasting under abnormal situations.
- We developed a Transformer model to understand how deep-learning models capture the A&B correlations.
- We address a difficult but useful problem in transit operation (forecast abnormal metro demand with long lead time).

Limitations

- Holidays and events data, could further enhance the forecast.
- How to quantify the uncertainties in the forecast?
- Using likelihood as a loss function (probabilistic forecasting), on going work. A preprint coming soon.

Thank you!
Questions?

Dr. Zhanhong Cheng
zhanhong.cheng@mcgill.ca
https://chengzhanhong.github.io/