What Drives Predictions in Traffic Forecasting? Data Valuation for Deep Learning on Time Series.

Dylan Chou

Advisor: Peter Freeman, Associate Teaching Professor, Department of Statistics and Data Science

Estimation Method Definitions

Definitions

- A *data region* is a set of consecutive traffic time series observations.
- When we define region types, we classify them into *concept classes*, or data sharing common feature(s).
- *Importance* is estimated as the increase in mean absolute error (MAE) of traffic time series predictions after masking out a reference region *R* with regions in the same *concept class*.

Motivation

- Past work: model accuracy and data importance of individual time steps but did not investigate importance over contiguous time steps.
- We believe that data regions can capture traffic patterns that individual time steps cannot.
- Traffic operators can identify biases and suggest model retraining.

Objective/Research Questions

Objective: Conduct a case study on datasets PeMS04, PeMS08, and A414 Highways England to test our estimation method.

- Q1: Is traffic towards the start or end of a day more important in model predictions?
- **We define the start and end of the day as the first and last four hours of the day.**
- Q2: Is data with larger changes in traffic flow more important than smaller changes on model predictions?
- **Defined by threshold of positive/negative consecutive changes and Euclidean norm.**

Estimation Method

Procedure

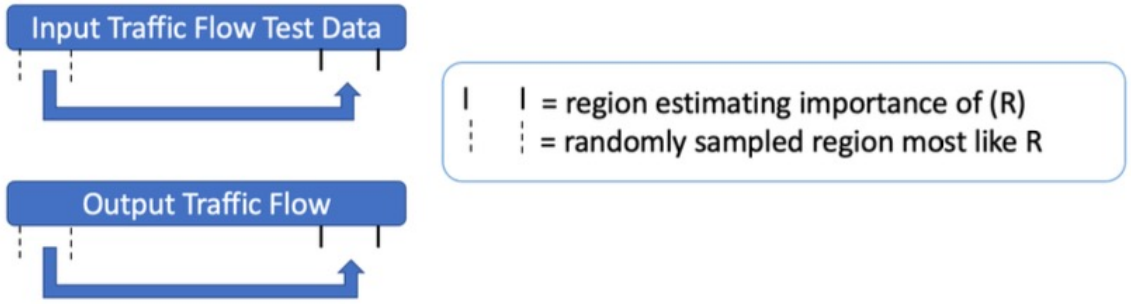
- (1) Define data regions of greatest and least traffic flow changes as well as start and end of day traffic.

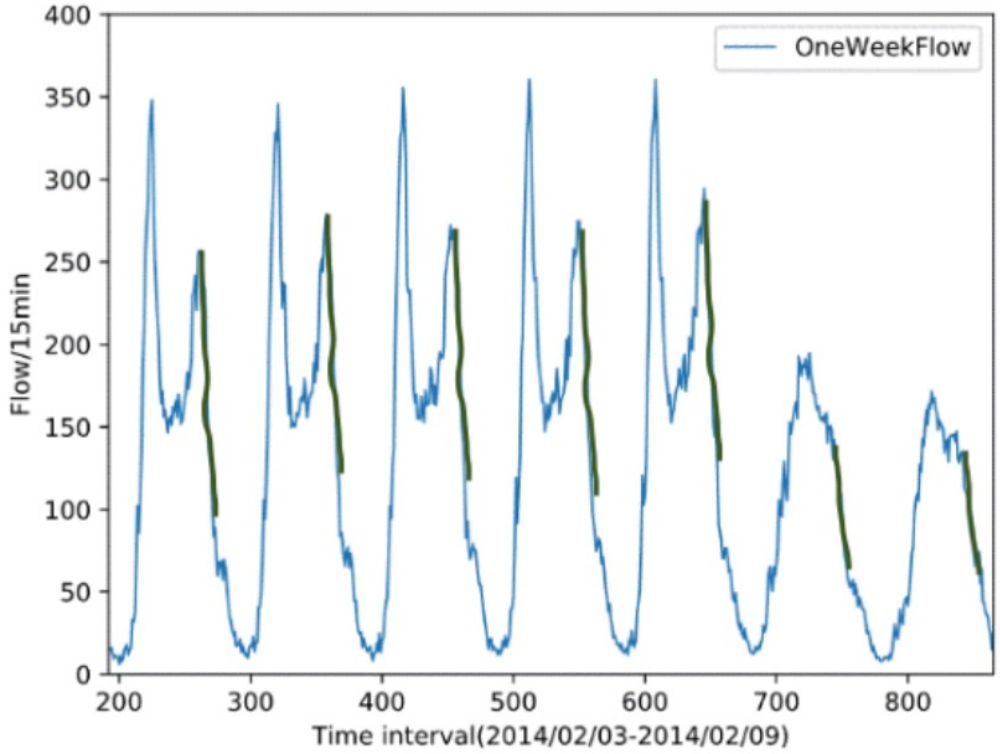
Dataset Region Name	Top K	% Consecutive	# Neighbors in Smoothing	# Observations
PeMS04 [Largest Increases]	50	70%	5	50
PeMS04 [Smallest Increases]	50	70%	15	50
PeMS04 [Largest Decreases]	40	60%	5	50
PeMS08 [Smallest Increases]	10	80%	15	50
PeMS08 [Largest Increases]	15	70%	5	50
PeMS08 [Largest Decreases]	12	60%	5	50
A414 [Largest Decreases]	10	90%	2	16
A414 [Largest Increases]	25	60%	2	16
A414 [Smallest Increases]	20	30%	2	16

Estimation Method (cont.)

Procedure

- (2) Per region *R*, find the set of regions most similar to *R* and repeatedly resample pairs of regions, masking one out with the other and perturbing solely traffic flow data.





Estimation Method (cont.)

Procedure

- (3) Measure the increase in prediction MAE after masking out the region.
- (4) Evaluate *stability,* or how often estimated MAEs result in significant conclusions.
- To measure stability, we have 32 repeated runs, where each run returns 25 prediction MAEs. Per run, we compare the distributions for each region. Significance with Bonferroni correction yields 32 hypothesis tests at level $\alpha = \frac{0.05}{32} = 0.0015625$.

Datasets and Models

- We used three datasets: PeMS04, PeMS08, A414 Highways England
- Model for PeMS: Guo et al's graph convolutional neural network
- Model for A414 Highways England: Du et al's hierarchical CNN-GRU neural network

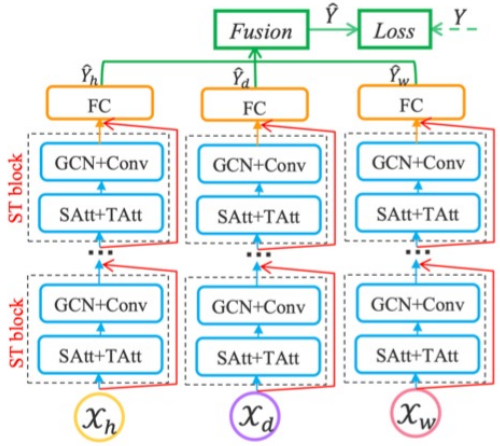
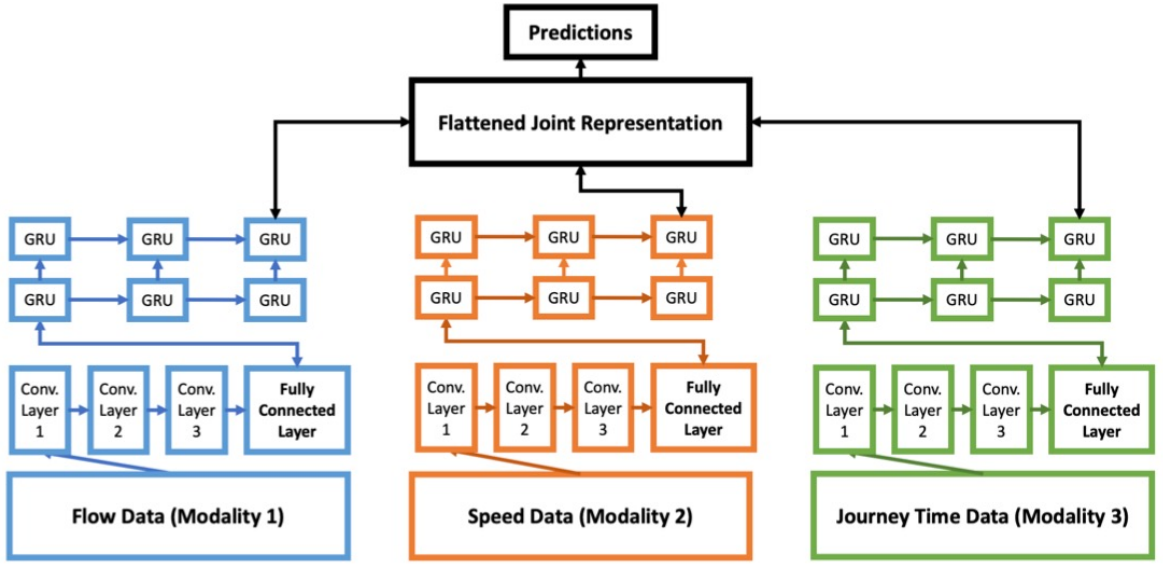


Figure 3: The framework of ASTGCN. SAtt: Spatial Attention; TAtt: Temporal Attention GCN: Graph Convolution; Conv: Convolution; FC: Fully-connected; ST block: Spatial-Temporal block.



Dataset Definitions

PeMS04/PeMS08

- Traffic time series observations recorded every 5 minutes.
- PeMS04 (1/1/2018-2/28/2018), PeMS08 (7/1/2016-8/31/2016)

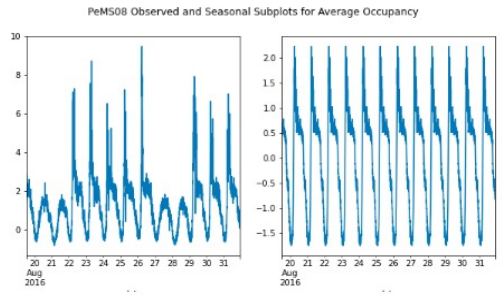
A414 Highways England

- Traffic observations recorded every 15 minutes.
- A414 (1/1/2013-12/31/2013 for training set, 2/1/2014-2/28/2014 for testing set)

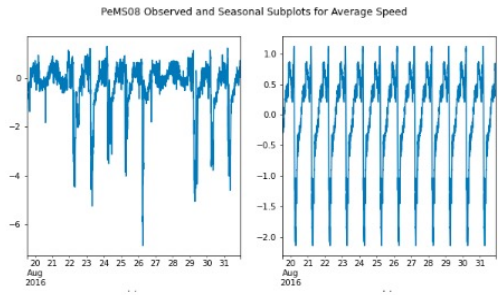
PeMS04	PeMS08	A414
Traffic Flow, Average Speed, Average Occupancy	Traffic Flow, Average Speed, Average Occupancy	Traffic Flow, Average Speed, Average Journey Time
Training: 10,181 Testing: 3,394	Training: 10,699 Testing: 3,567	Training: 34,876 Testing: 2,688

Resampling Similar Regions in Traffic Flow

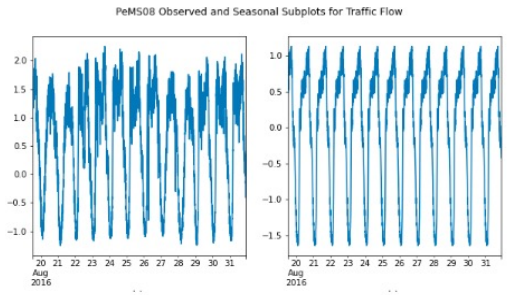
PeMS datasets



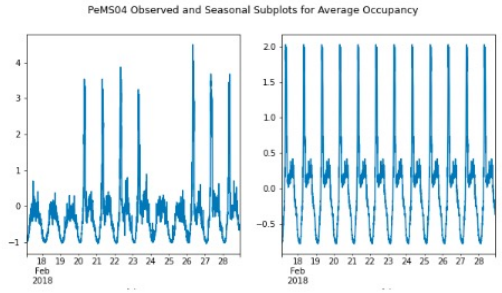
(a)



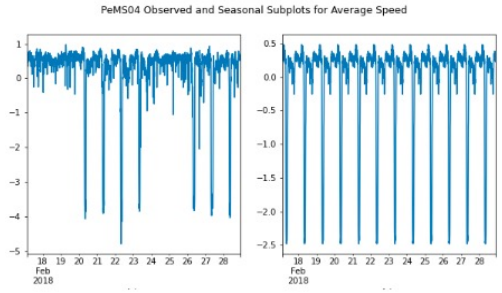
(b)



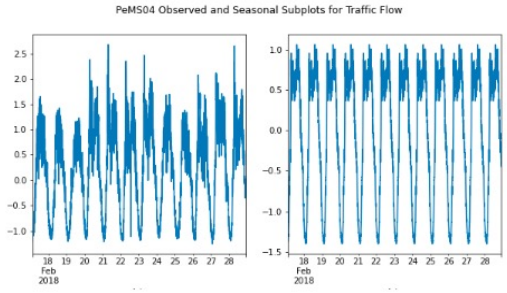
(c)



(d)

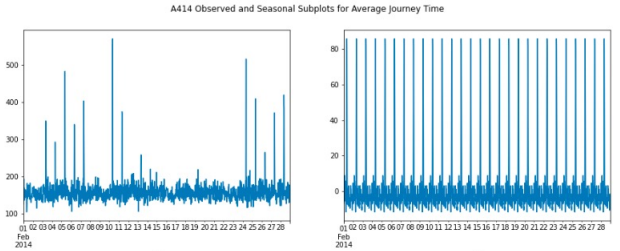


(e)

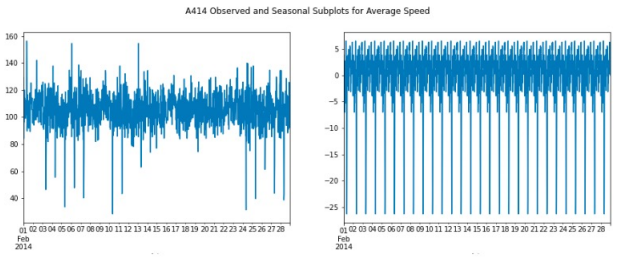


(f)

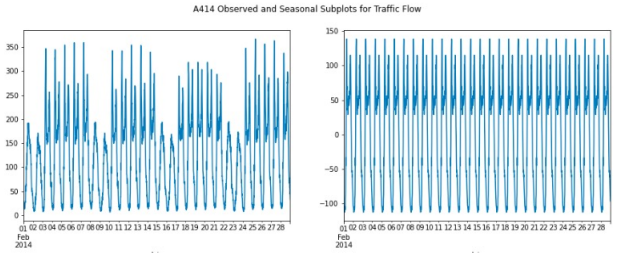
A414 dataset



(a)



(b)



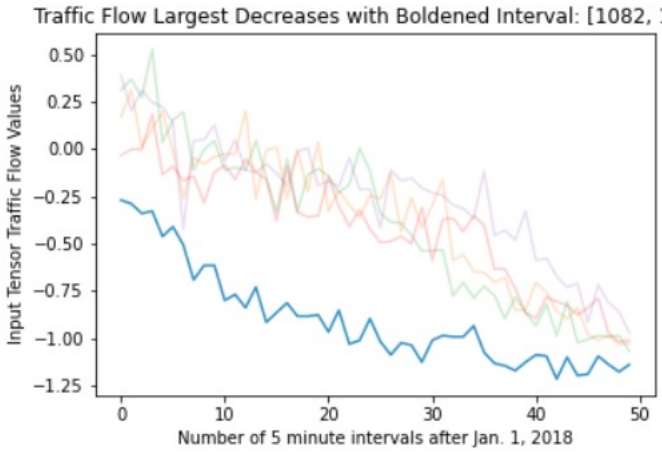
(c)

Results Definitions

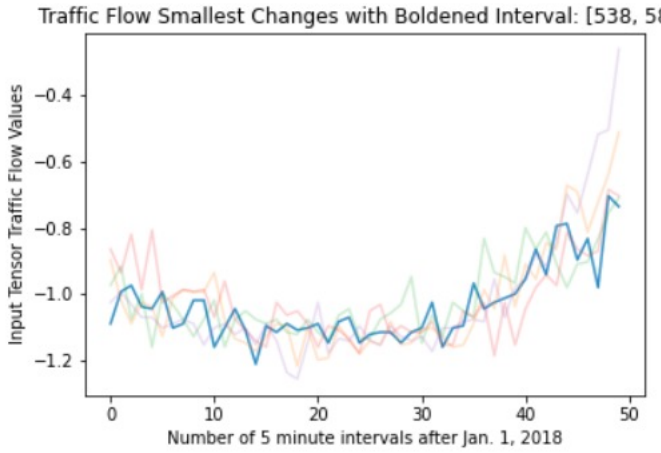
Definitions

- R_1 = largest increases in traffic flow
- R_2 = smallest changes
- R_3 = largest decreases
- R_4 = start of day traffic
- R_5 = end of day traffic

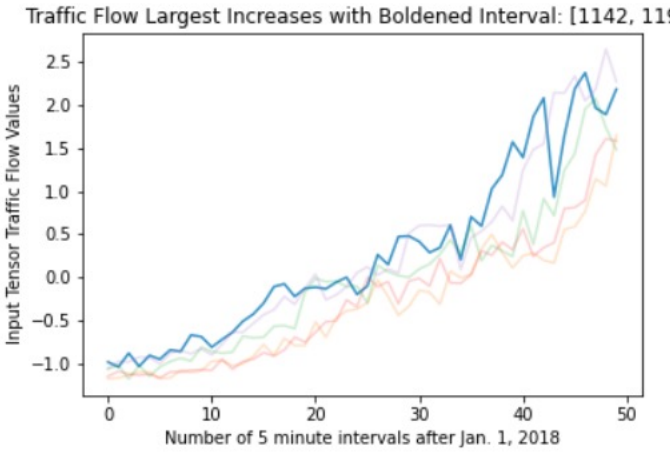
Results (Defined Regions)



(a)



(b)



(c)

Results (Stability)

- **A414**: runs were not statistically significant (0/32) for all pairs of regions.
- **PeMS04/PeMS08**: nearly stable test results ($R_3 - R_1$: 27/32, $R_5 - R_1$: 29/32) or stable results among the 32 runs ($R_3 - R_2$: 32/32).

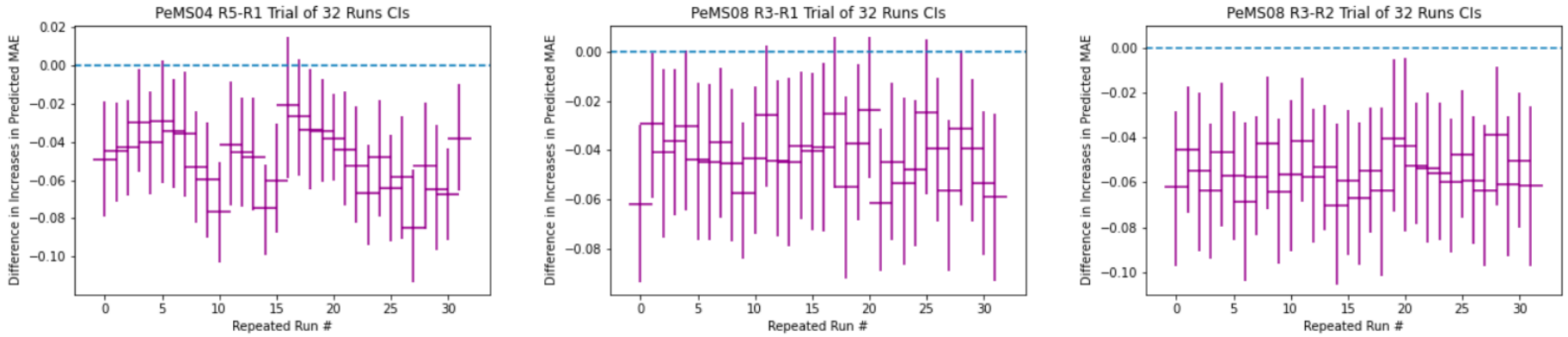
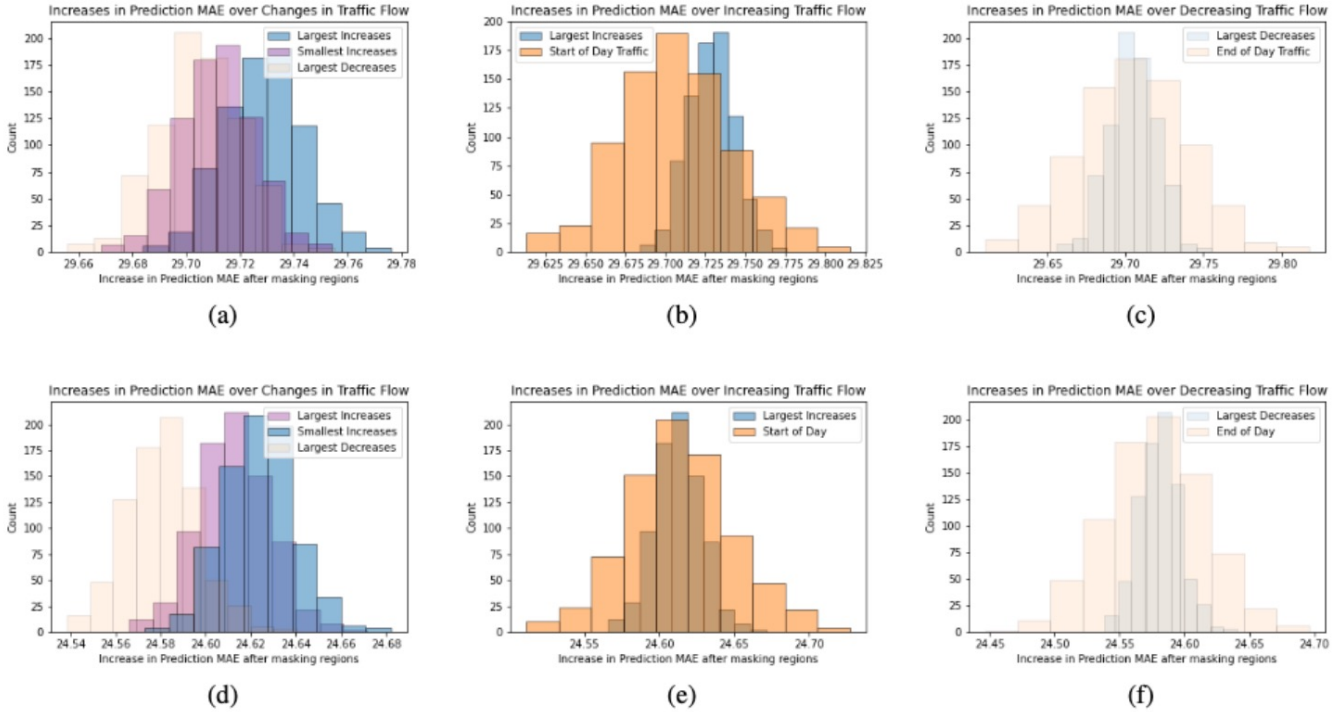
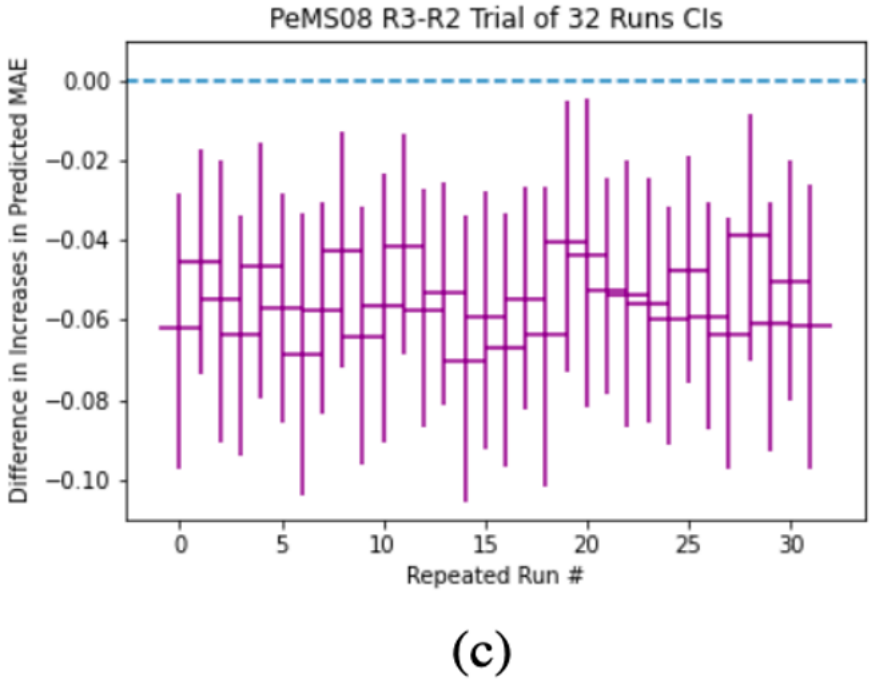


Figure 11: The graph on the left displays the confidence intervals (CIs) from 32 runs on the PeMS04 dataset, where each run had 20,000 bootstrapped sampled difference in average prediction MAEs. The graph in the middle contains the CIs from 32 runs on the PeMS08 dataset of the difference in prediction MAE between the regions defined by the largest decreases and increases in traffic flow. The plot on the right displays confidence intervals across the 32 runs of the difference in prediction MAEs: $R_3 - R_2$. The confidence intervals are provided as results of the repeated runs and the stability of the estimation method.

Results (Stability cont.)

- “start of day” traffic appears to be more important to traffic flow forecasts than “end of day” traffic for the PeMS datasets.
- Regions with the smallest changes in traffic flow consistently appear to be more important than largest decreases in PeMS08.





Conclusion

- **Answers to the Two Questions:**
  - Smaller changes in traffic flow appears to contribute more to traffic flow predictions than larger changes in PeMS08 only.
  - “start of day” traffic appears to contribute more to traffic flow forecasts than “end of day” traffic in the PeMS datasets.
- **Stability:**
  - **A414**: runs were not statistically significant.
  - **PeMS04**: 29/32 of runs’ largest increases in traffic flow are deemed more important than “end of day” traffic regions.
  - **PeMS08**: 32/32 of the runs resulted in smallest changes > largest decreases in traffic flow.

Limitations and Future Work

Limitations

- We assume the resampled regions of data are roughly equivalent (same *concept class*) to reference region *R* and could be substituted.
- Assume importance of data is the increase in prediction MAE after masking.

Future Work

- Test on newer datasets (e.g., data in connected vehicle environments (CVE)).
- Perturb all features (e.g., average speed, average journey time) and account for causal dependencies between features.
- Evaluate *interpretability,* or improvement in traffic operator understanding of model after using estimation method.