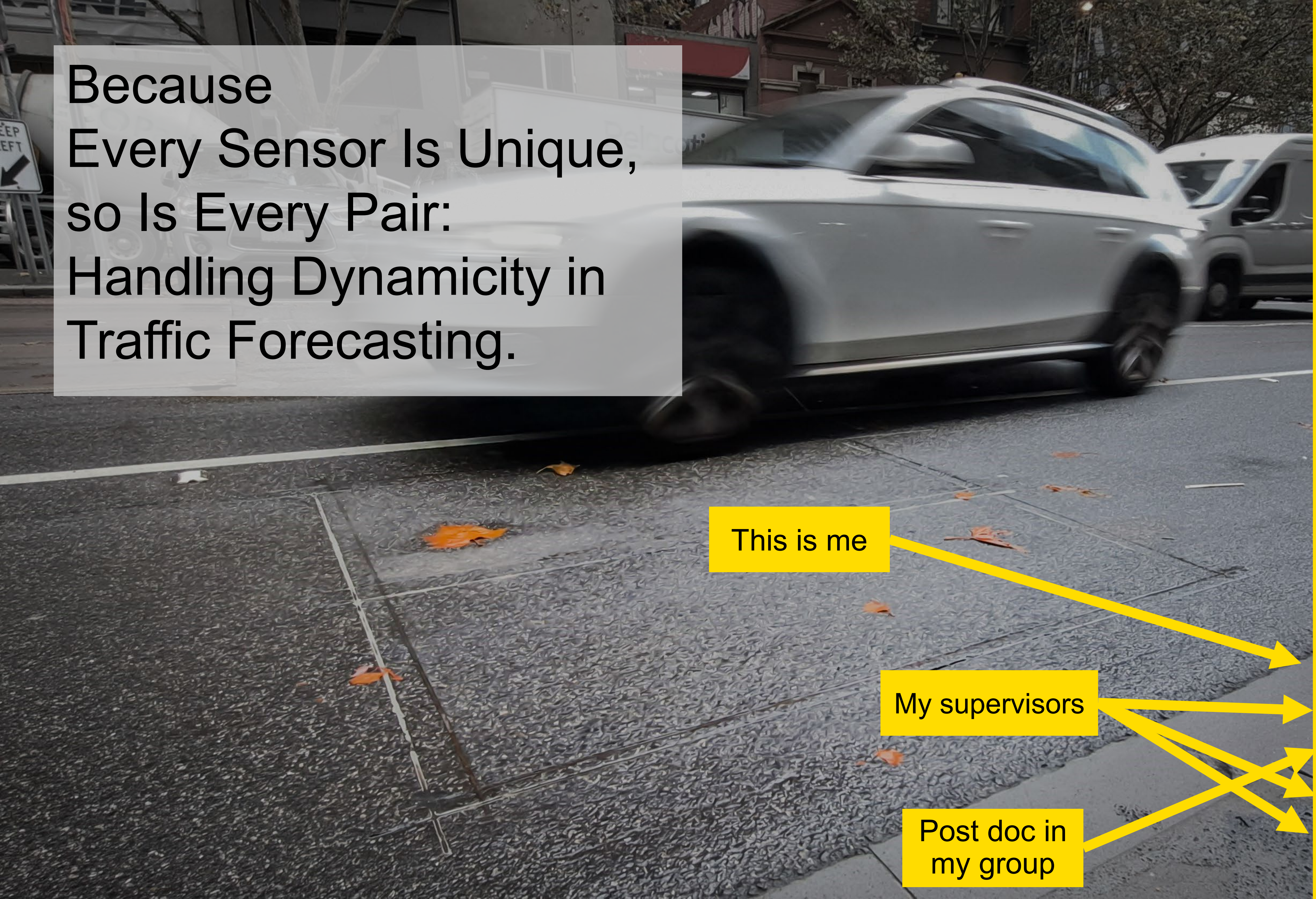
Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in Traffic Forecasting.

Arian Prabowo,
Wei Shao,
Hao Xue,
Piotr Koniusz, and
Flora D. Salim.

UNSW SYDNEY
CSIRO
RMIT UNIVERSITY



Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in Traffic Forecasting.

This is me

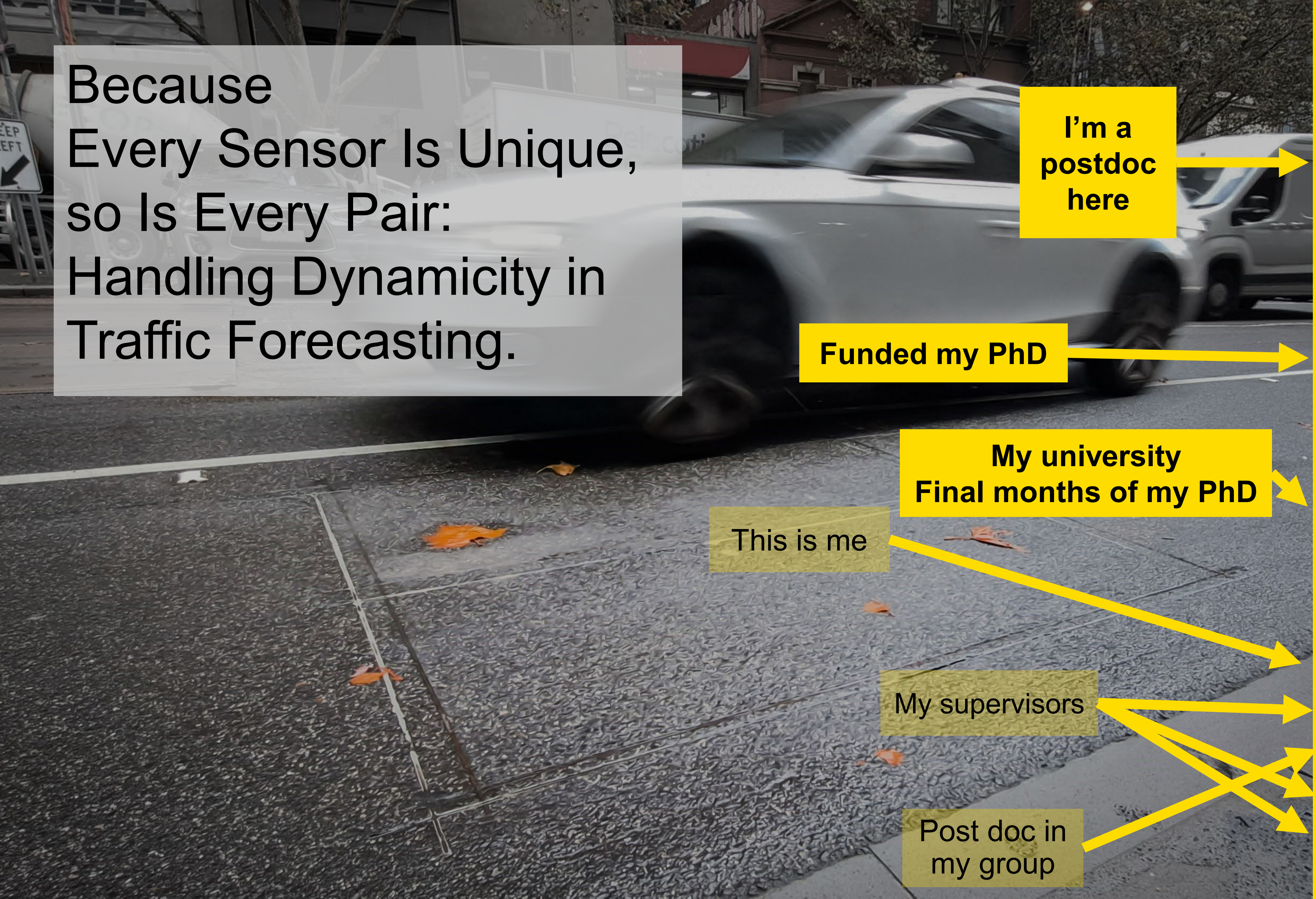
My supervisors

Post doc in my group

Arian Prabowo,
Wei Shao,
Hao Xue,
Piotr Koniusz,
Flora D. Salim.

UNSW SYDNEY
CSIRO
RMIT UNIVERSITY

Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in Traffic Forecasting.



I'm a postdoc here

UNSW SYDNEY

Funded my PhD

CSIRO

My university
Final months of my PhD

RMIT UNIVERSITY

This is me

My supervisors

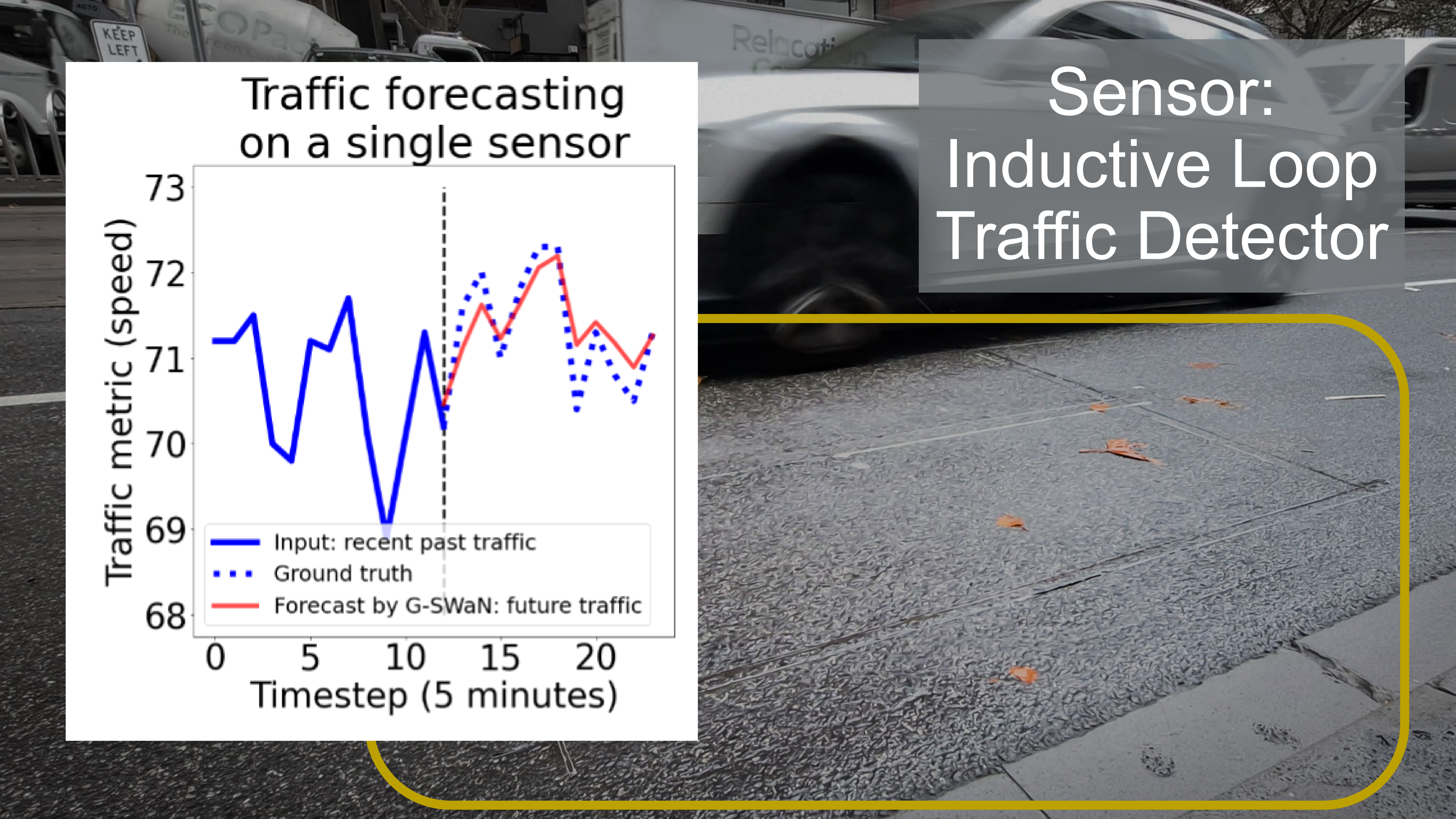
Post doc in my group

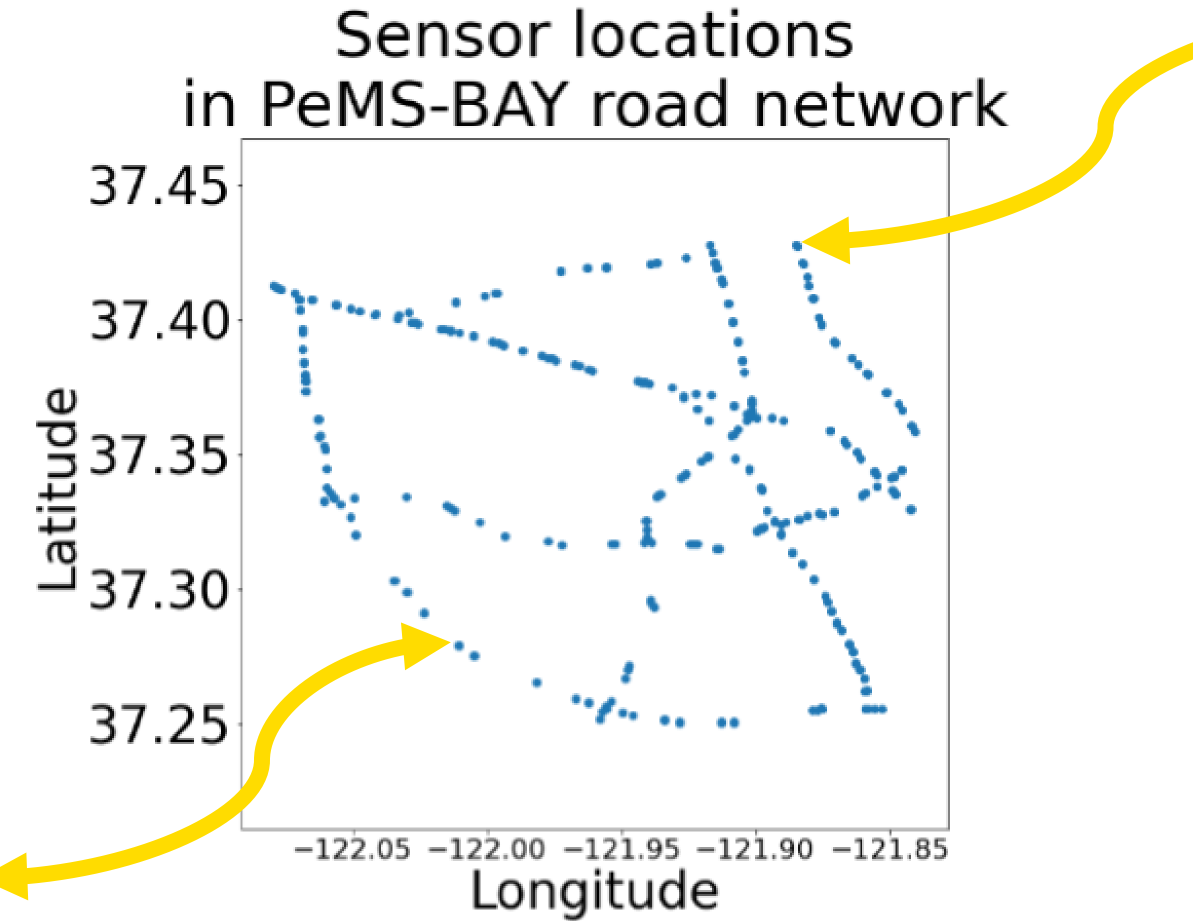
Arian Prabowo,
Wei Shao,
Hao Xue,
Piotr Koniusz,
Flora D. Salim.

UNSW SYDNEY

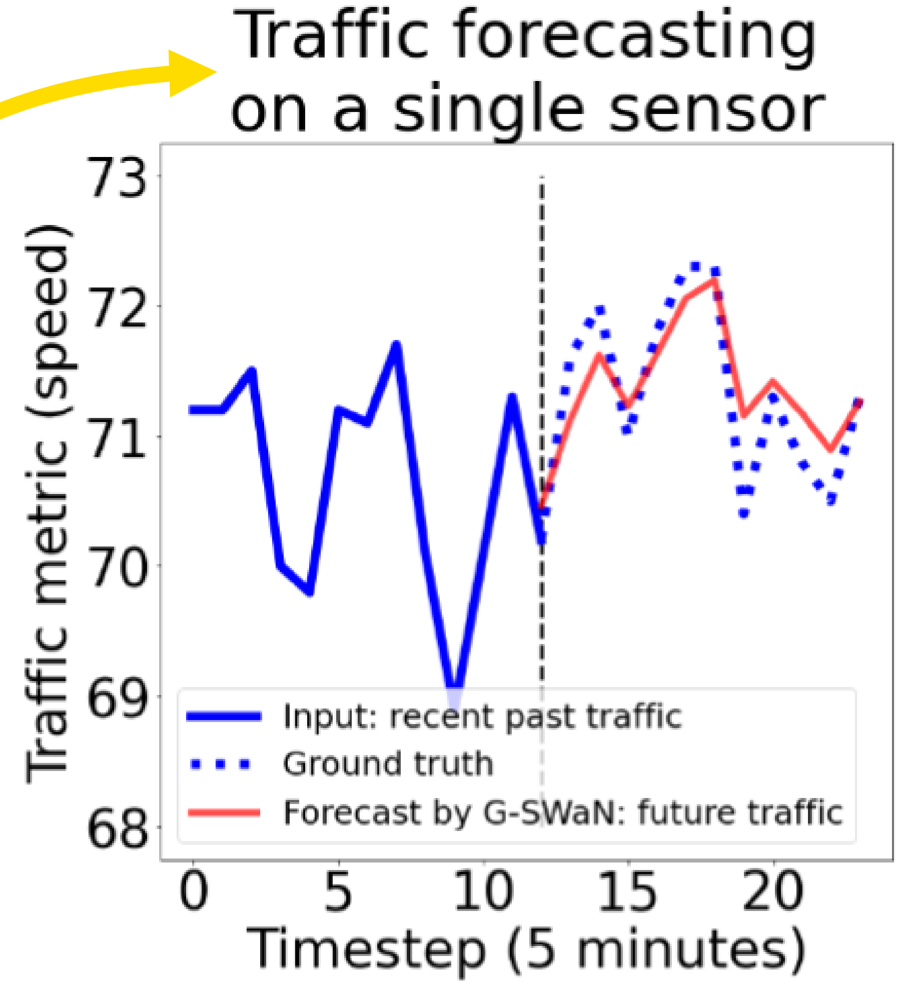
Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in **Traffic Forecasting**


Sensor:
Inductive Loop
Traffic Detector
KEEP
LEFT


KEEP LEFT
Traffic forecasting on a single sensor
Traffic metric (speed)
73
72
71
70
69
68
0
5
10
15
20
Timestep (5 minutes)
Input: recent past traffic
Ground truth
Forecast by G-SWaN: future traffic
Sensor: Inductive Loop Traffic Detector



(a) Locations of the sensors on the Californian highway network surrounding the bay area. Installing a network of sensors on a road infrastructure enables traffic forecasting and smarter cities.



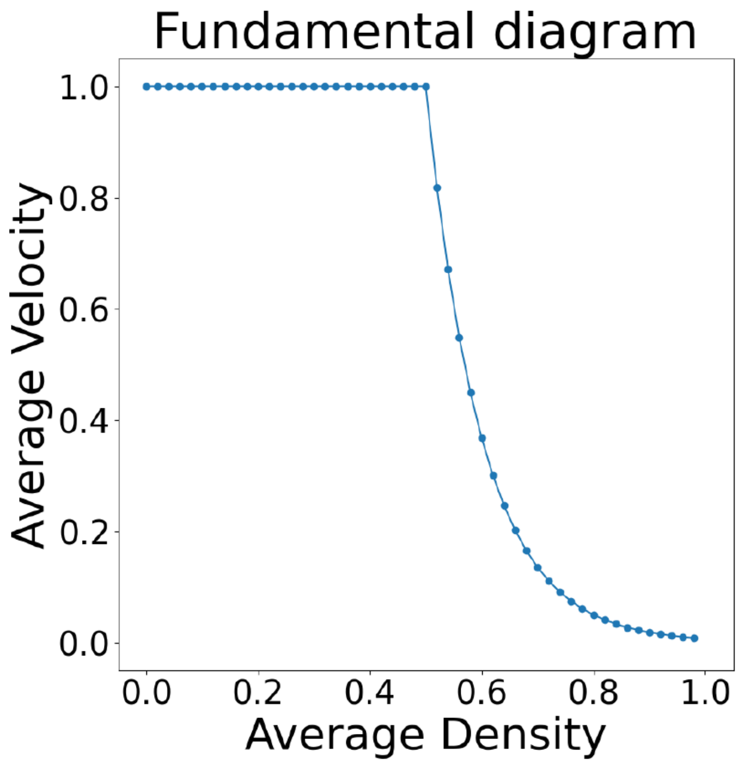
(b) At each sensor, traffic forecasting uses the recent sensor readings (solid blue line) to predict the future traffic (red line). This forecast is made by our proposed architecture Graph Self-attenion WaveNet (G-SWaN). Our forecasts accurately predict the future traffic (dotted blue line).

Fig. 1. Visual abstract of the traffic forecasting task.

Because
Every Sensor Is Unique,
so Is Every Pair:
Handling Dynamicity in
Traffic Forecasting

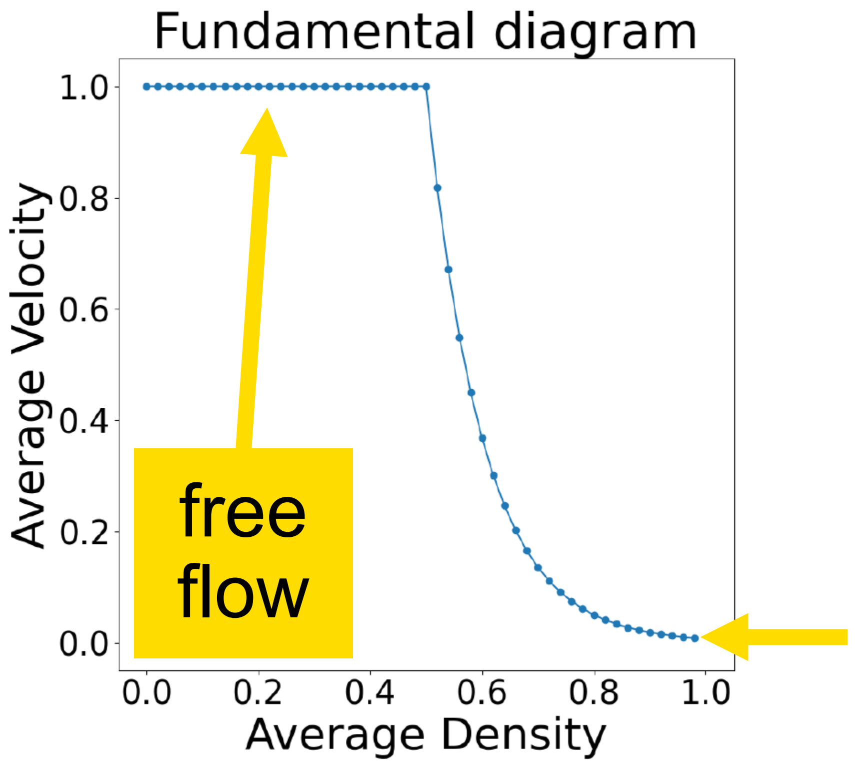
Because **Every Sensor Is Unique**, so Is Every Pair: Handling Dynamicity in Traffic Forecasting

Every Sensor Is Unique



(a) Idealised fundamental diagram according to [15].

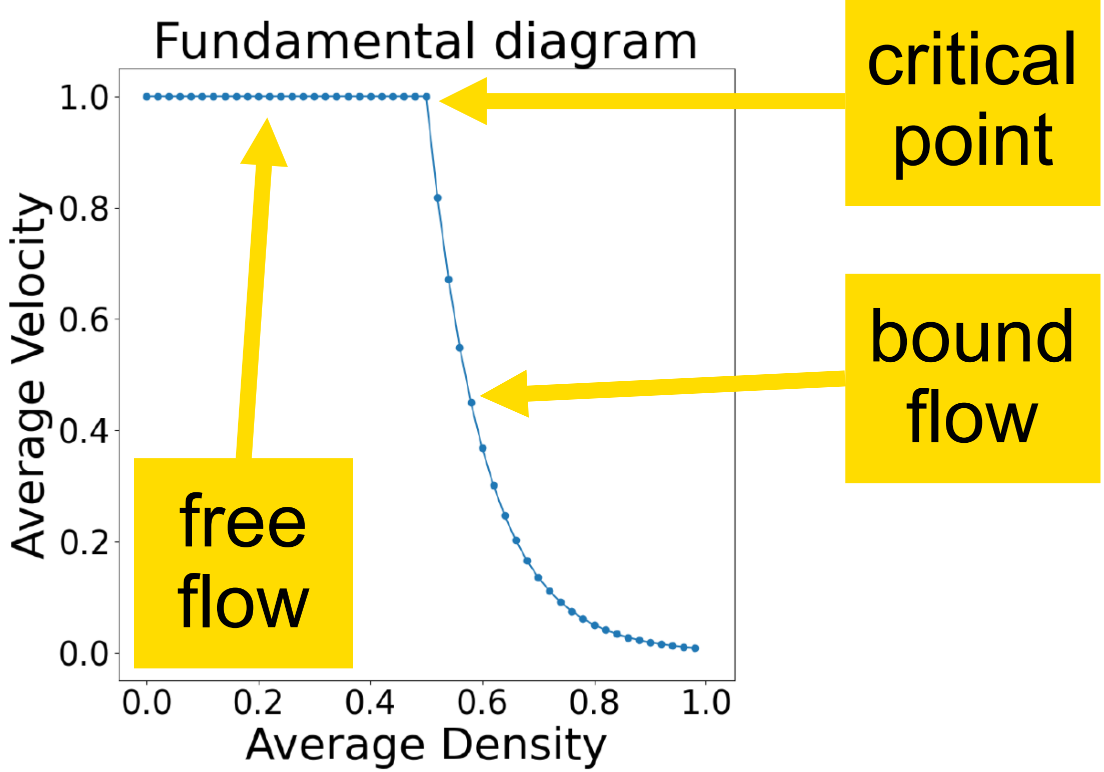
Every Sensor Is Unique



(a) Idealised fundamental diagram according to [15].

1 car = speed limit
2 cars = speed limit

Every Sensor Is Unique

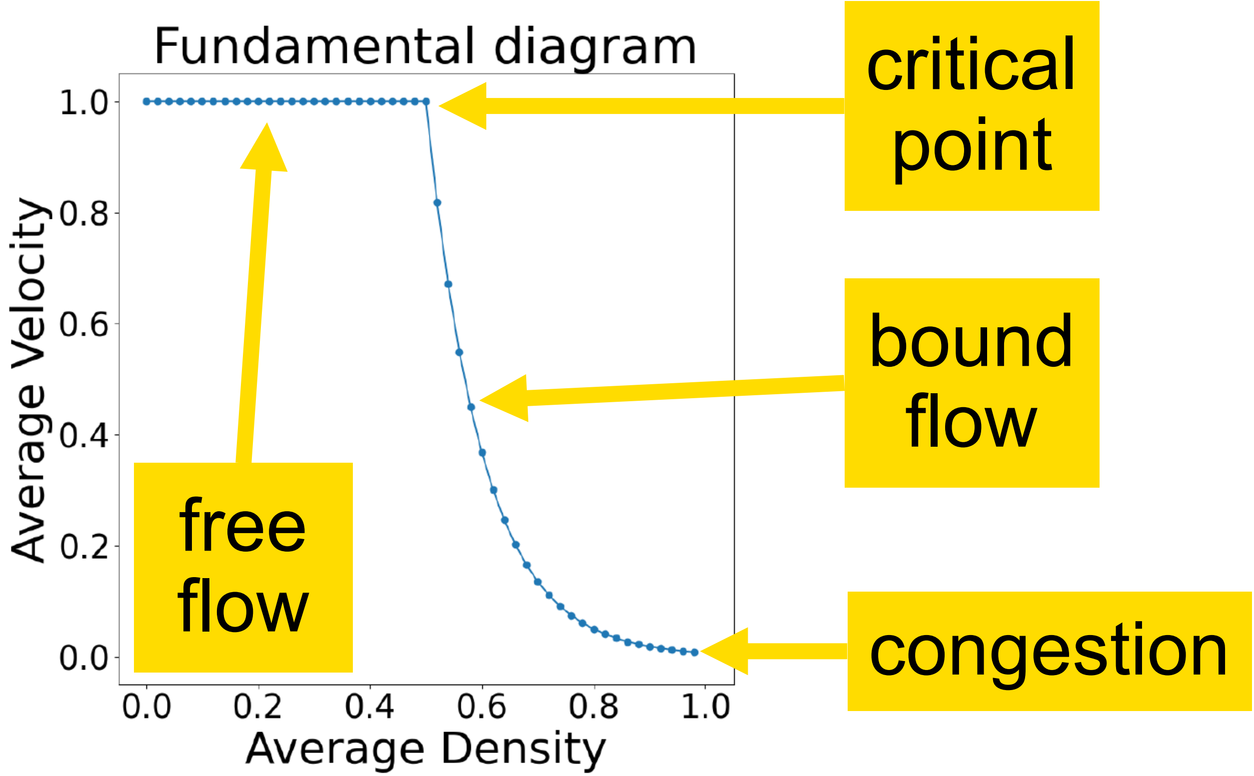


(a) Idealised fundamental diagram according to [15].

1 car = speed limit
2 cars = speed limit

C cars = speed limit
C+1 cars = slower

Every Sensor Is Unique



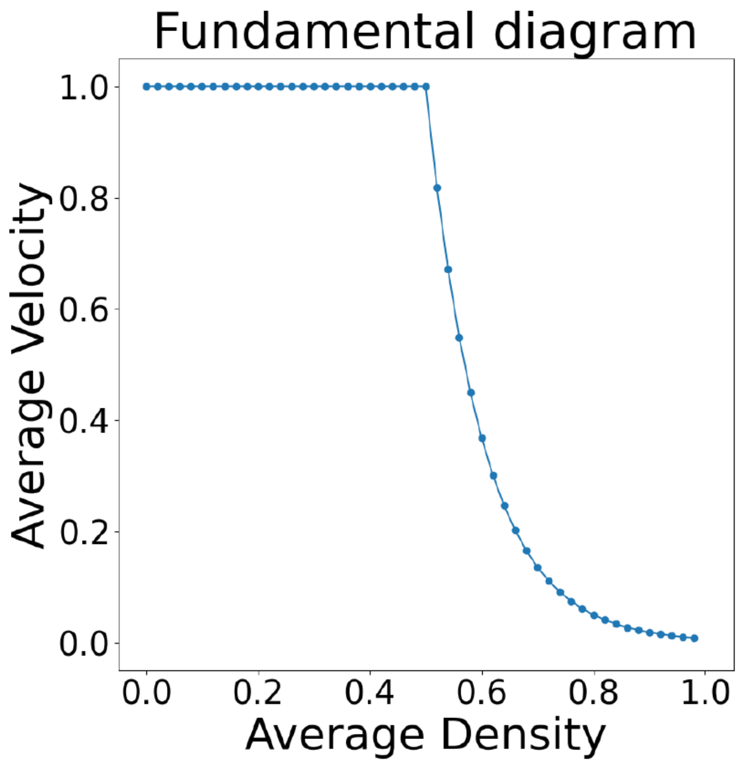
(a) Idealised fundamental diagram according to [15].

1 car = speed limit
2 cars = speed limit

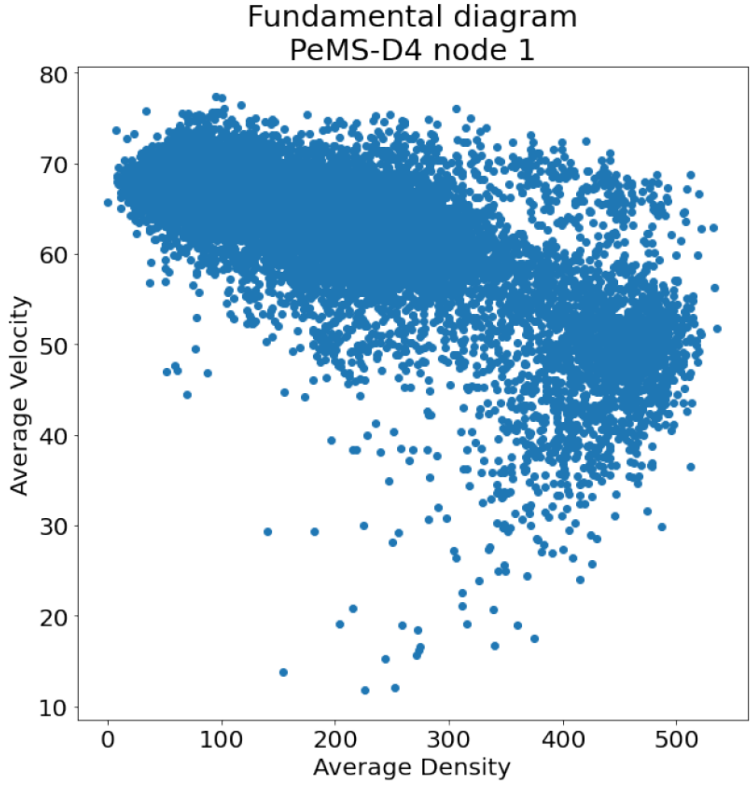
C cars = speed limit
C+1 cars = slower

C++ cars = congestion

Every Sensor Is Unique



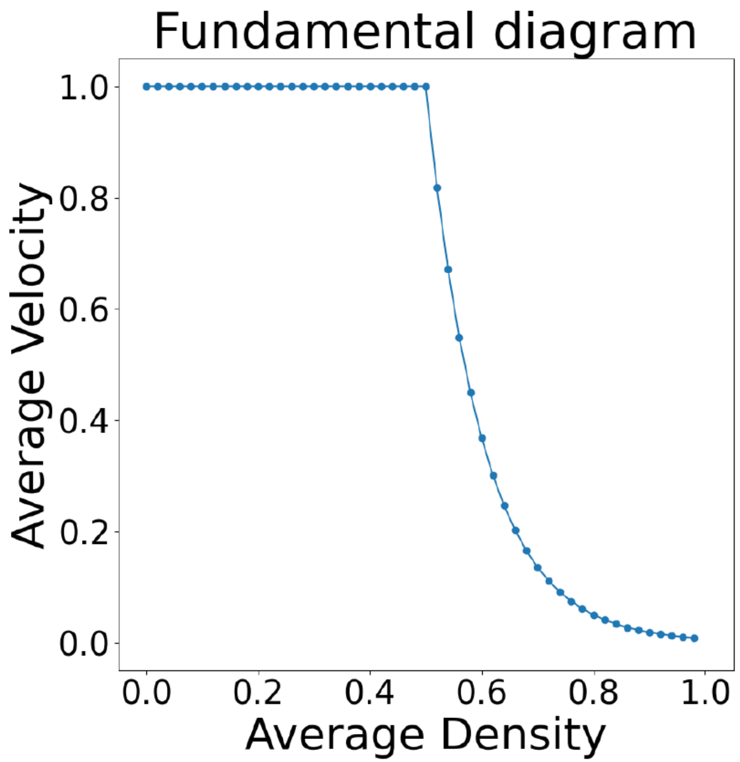
(a) Idealised fundamental diagram according to [15].



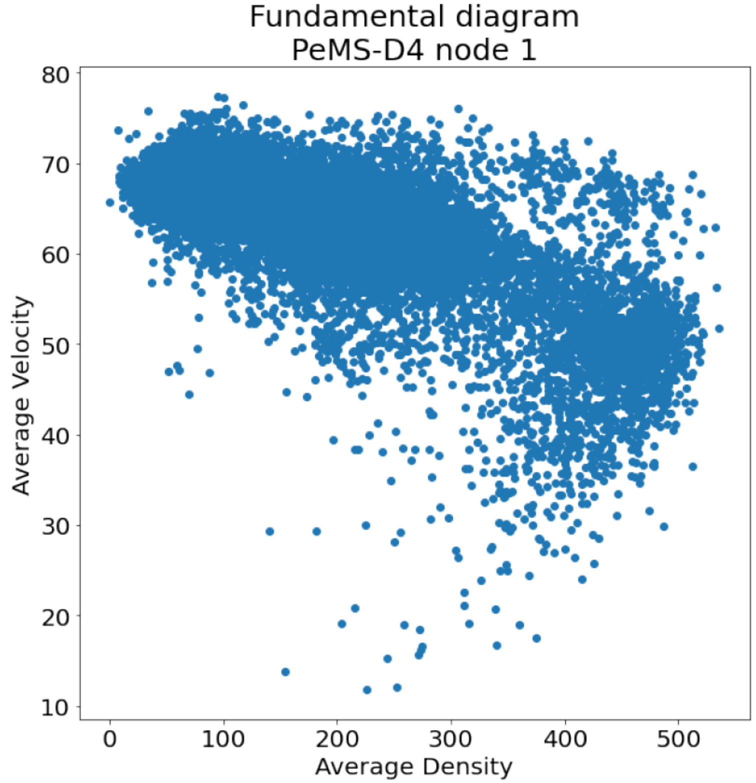
(b) Fundamental diagram of a sensor in the PeMS-D4 dataset.

Shape looks very different (with some similarities).

Every Sensor Is Unique



(a) Idealised fundamental diagram according to [15].

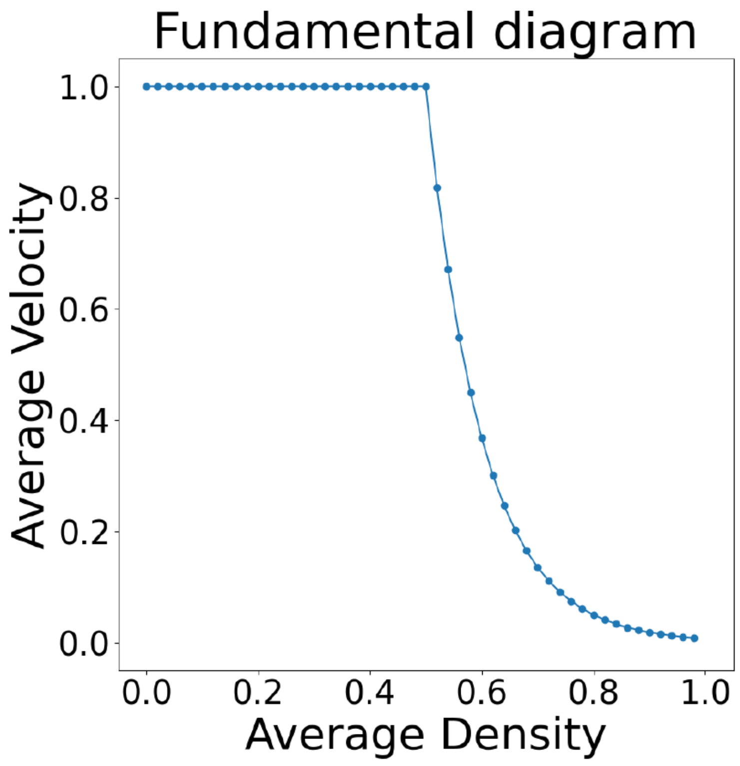


(b) Fundamental diagram of a sensor in the PeMS-D4 dataset.

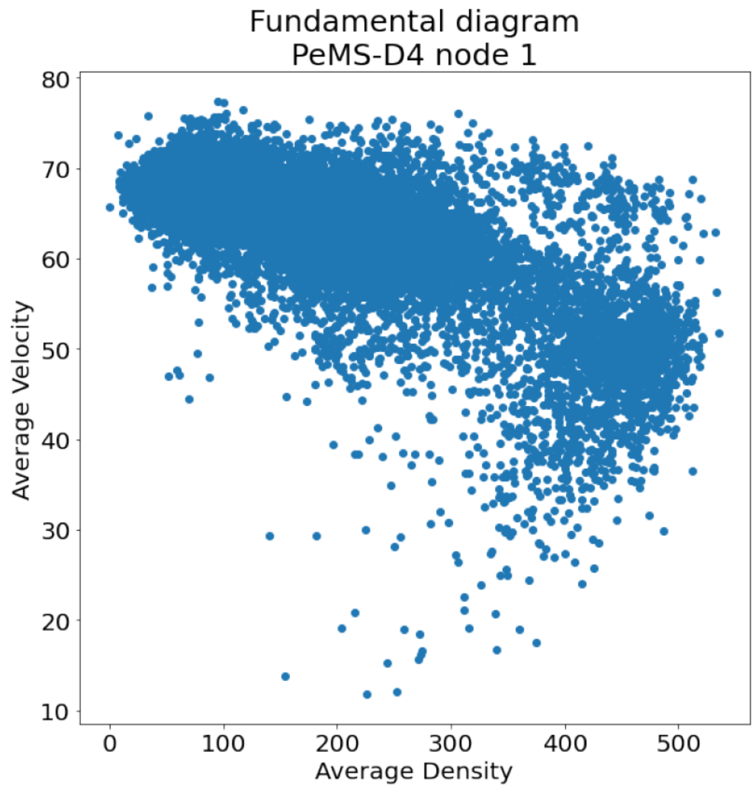
Shape looks very different (with some similarities).

Wider distributions.

Every Sensor Is Unique



(a) Idealised fundamental diagram according to [15].



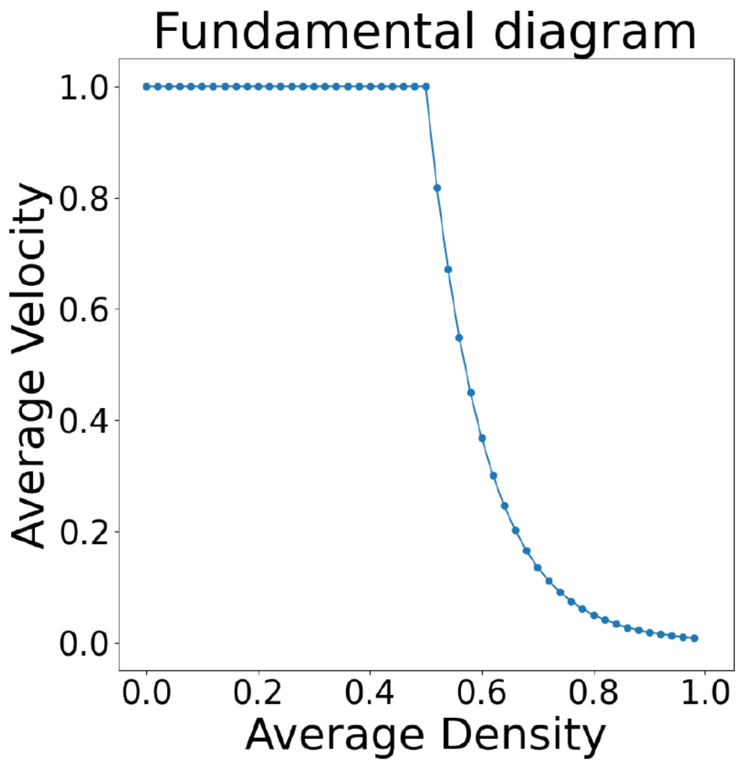
(b) Fundamental diagram of a sensor in the PeMS-D4 dataset.

Shape looks very different (with some similarities).

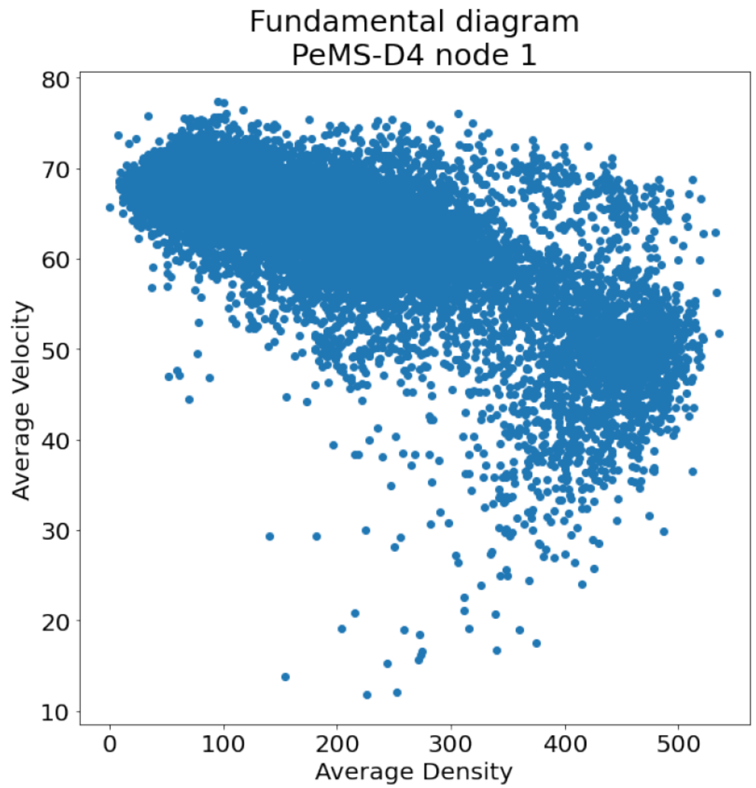
Wider distributions.

Maybe we picked an anomalous sensor?

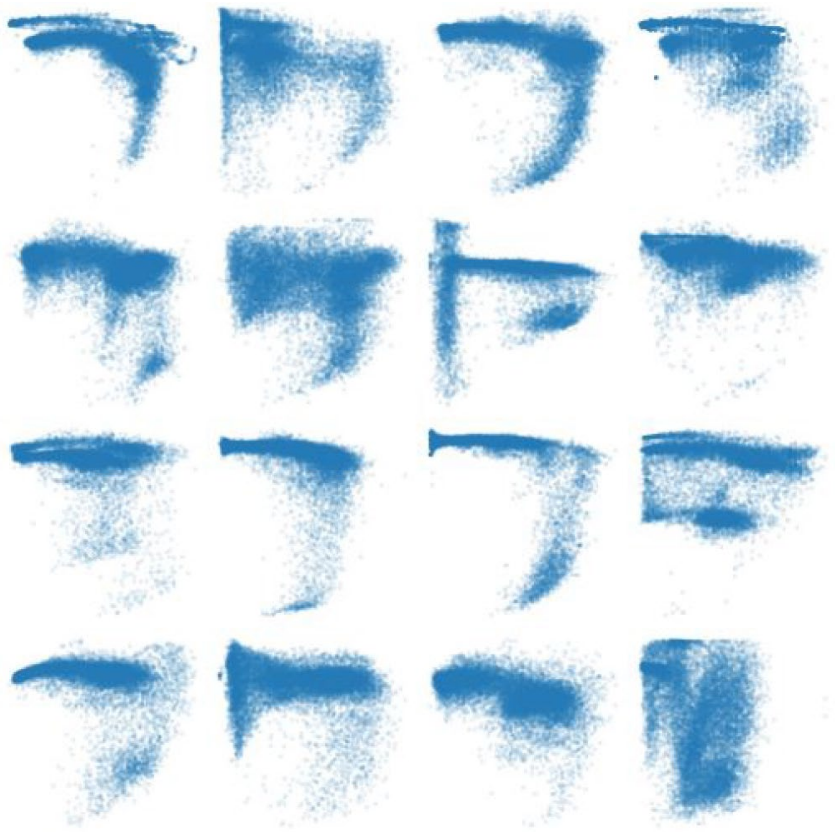
Every Sensor Is Unique



(a) Idealised fundamental diagram according to [15].

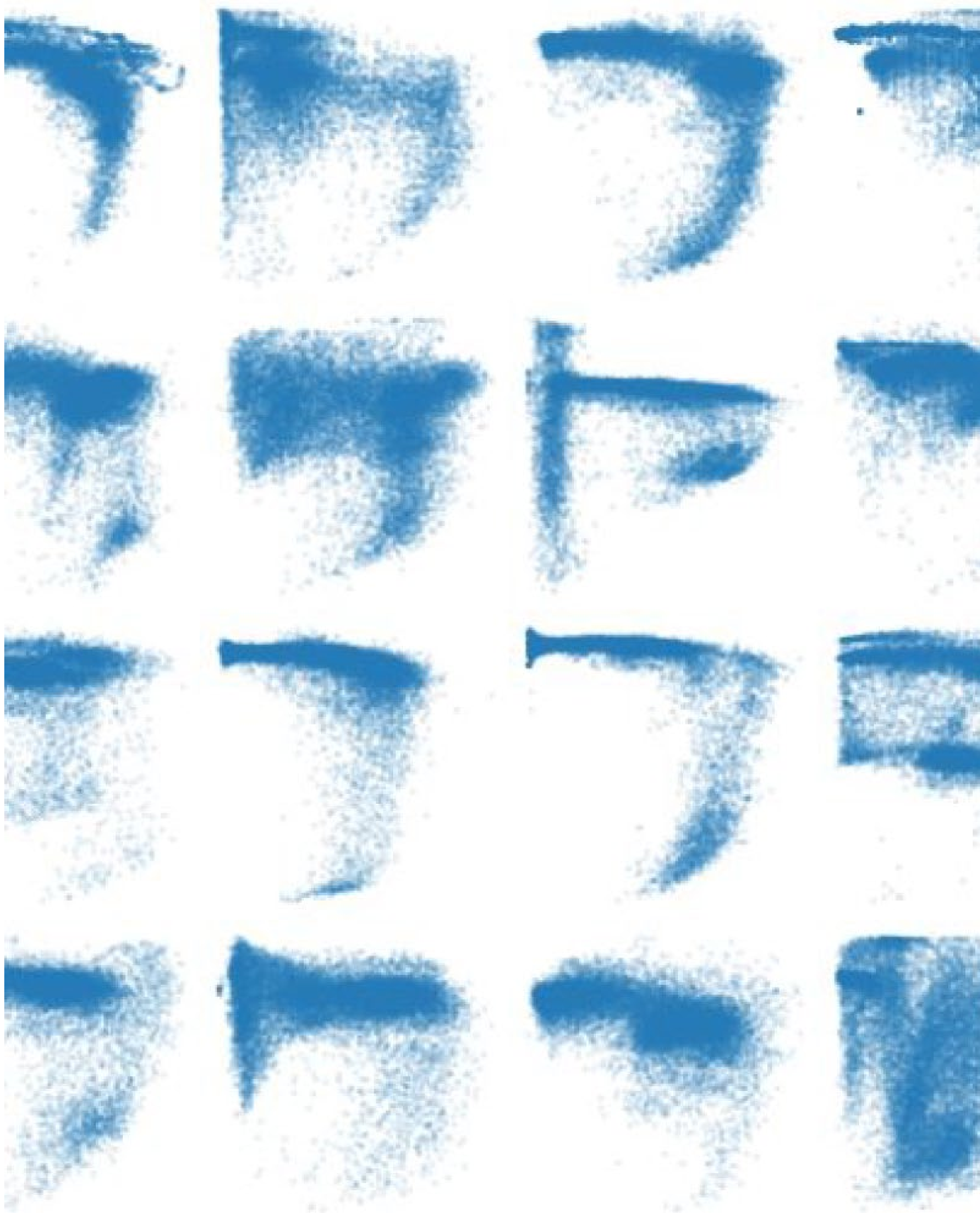


(b) Fundamental diagram of a sensor in the PeMS-D4 dataset.



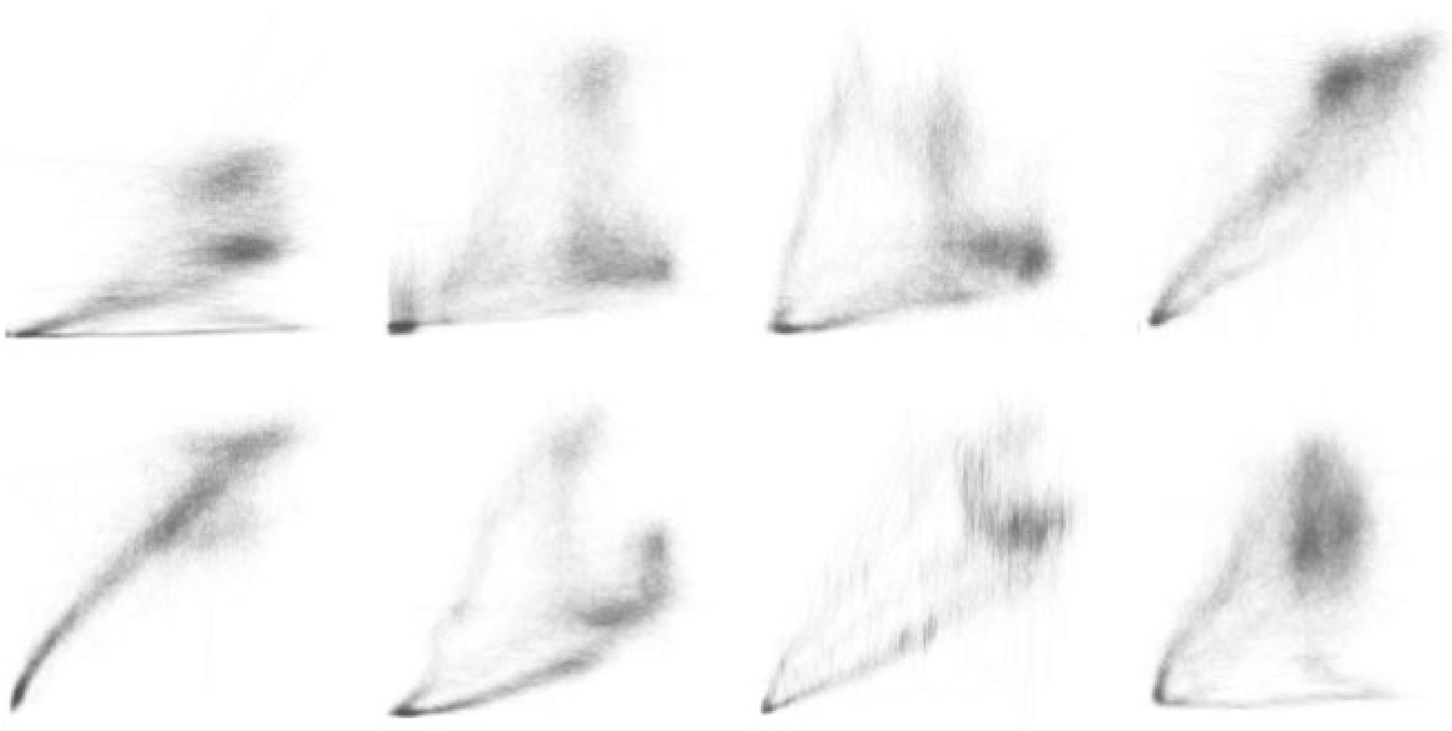
(c) Fundamental diagrams of 16 selected sensors in the PeMS-D4 dataset, showing great diversity.

Because **Every Sensor Is Unique**, so Is Every Pair: Handling Dynamicity in Traffic Forecasting



Because
Every Sensor Is Unique,
so Is Every Pair:
Handling Dynamicity in
Traffic Forecasting

so Is Every Pair



(a) Association plots of 8 random sensor pairs.

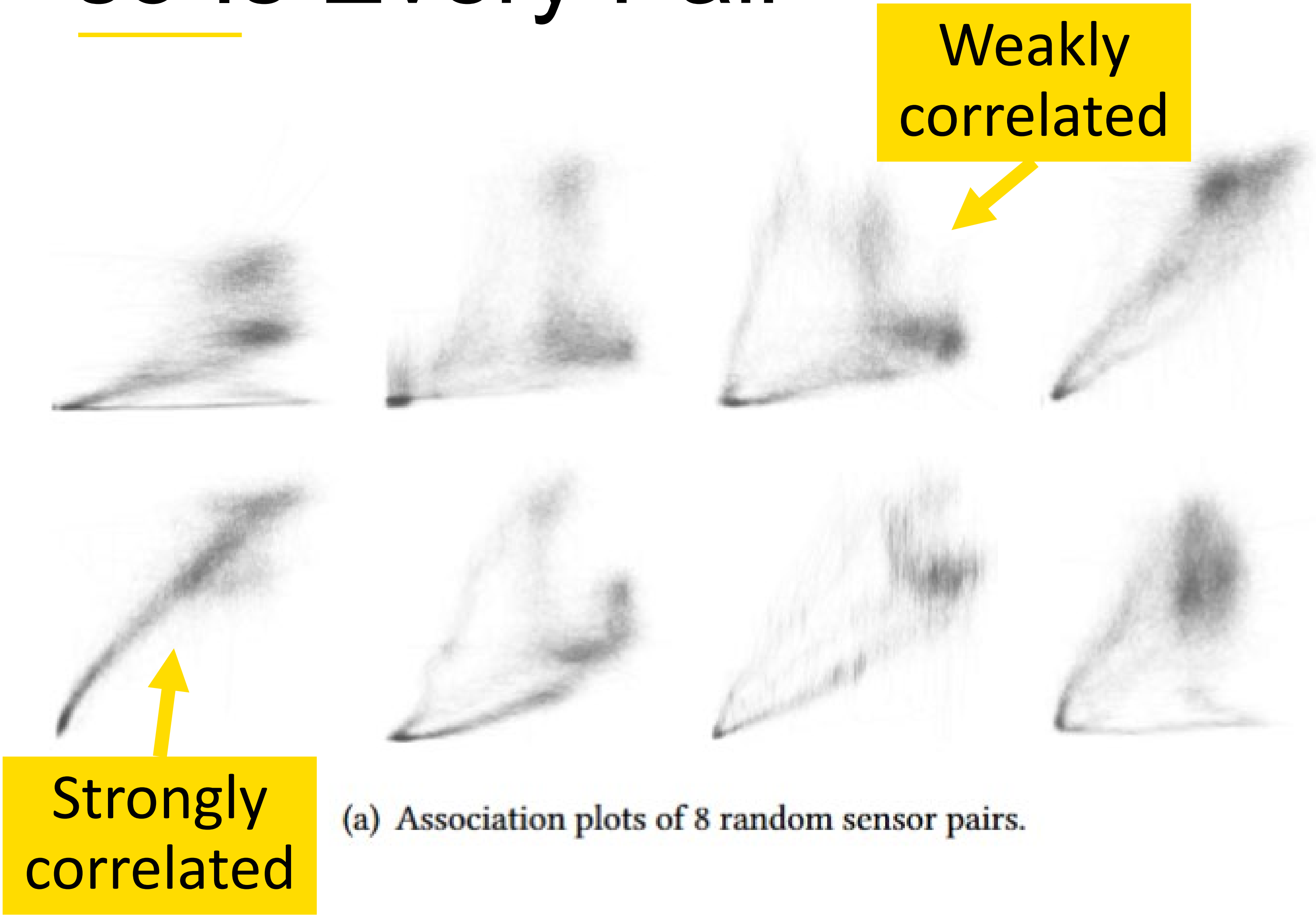
For every plot:

- x-axis: flow at sensor A
- y-axis: flow at sensor B

A and B are randomly selected locations

(flow ~ count ~ density)

so Is Every Pair



(a) Association plots of 8 random sensor pairs.

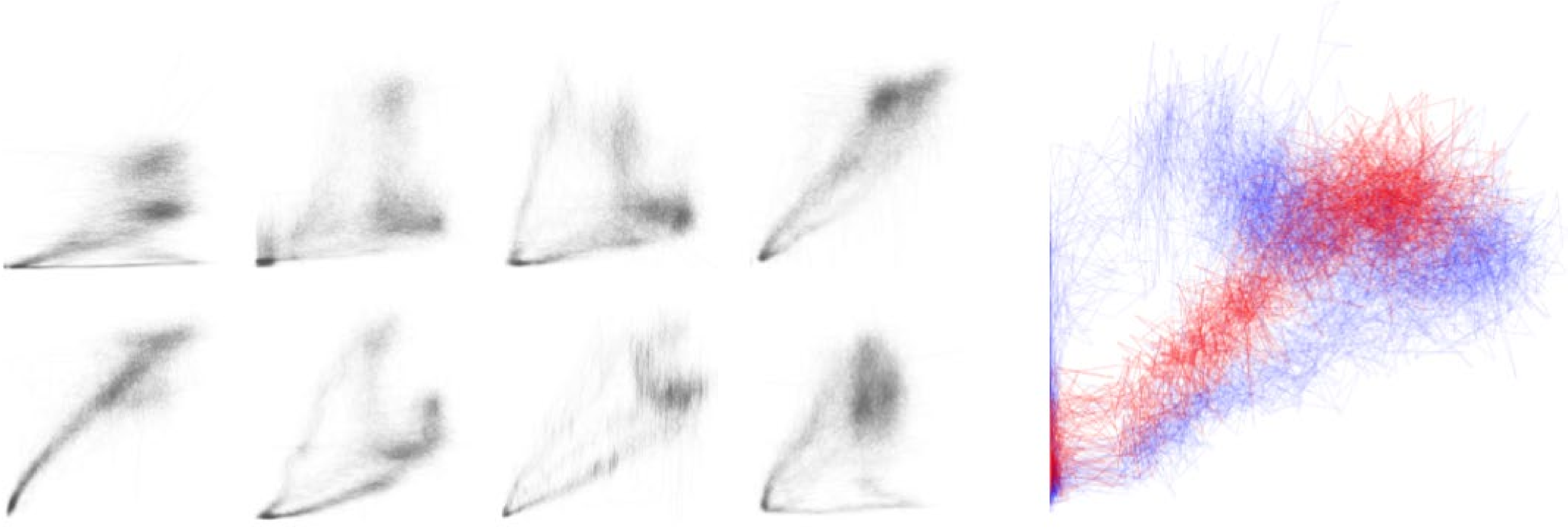
For every plot:

- x-axis: flow at sensor A
- y-axis: flow at sensor B

(A and B are randomly selected locations)

(flow ~ count ~ density)

so Is Every Pair **(Also change over time!)**

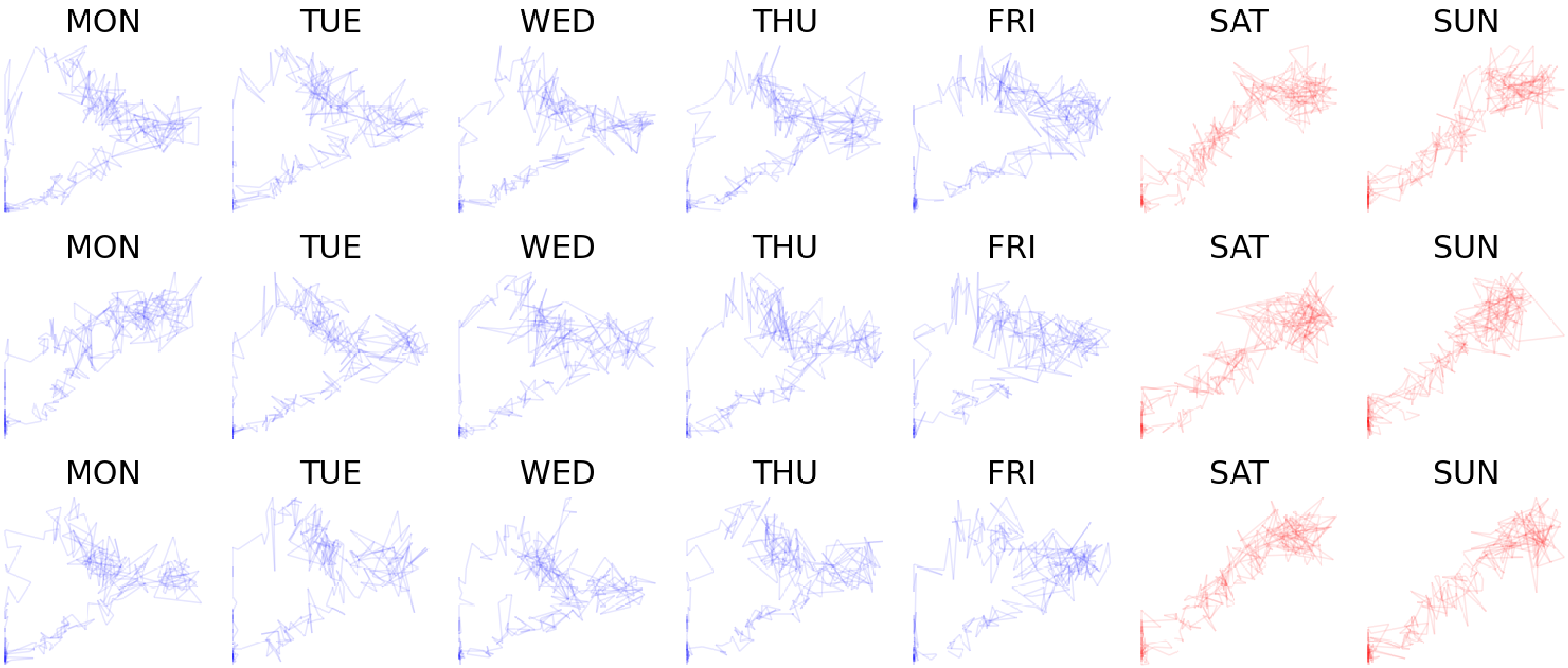


(a) Association plots of 8 random sensor pairs.

(b) Association plots of a sensor pair over three weeks. Blue lines are weekdays, red lines are weekends.

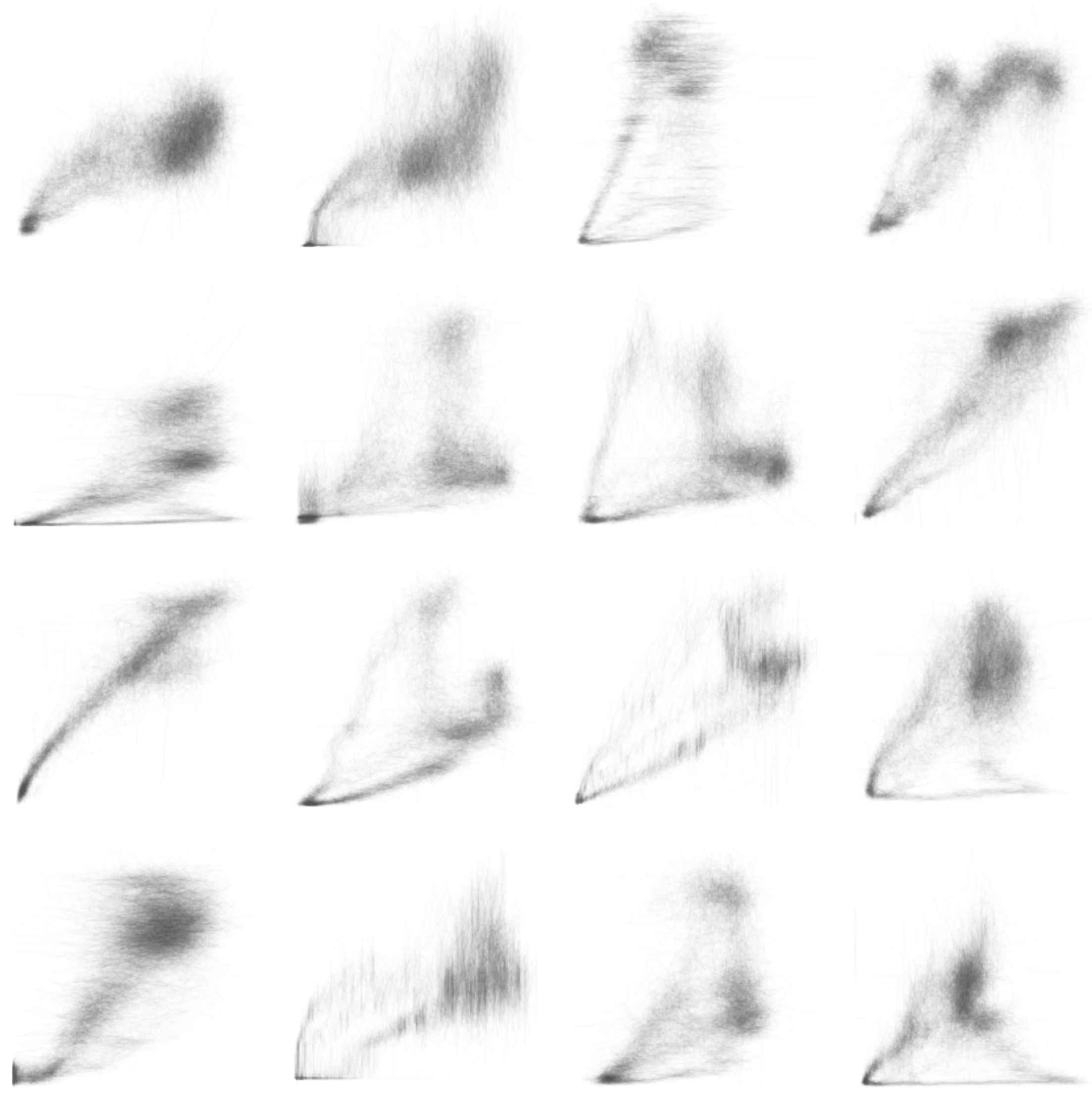
Fig. 3. Association plots of different pairs of sensor readings in PeMS-D4. The x-value of a point is the flow at one sensor, while the y-value is the flow at the other sensor. Consecutive data points are connected by a line.

so Is Every Pair **(Also change over time!)**



UNSW SYDNEY

Because Every Sensor Is Unique, **<u>so Is Every Pair</u>**: Handling Dynamicity in Traffic Forecasting



G-SWaN: Graph Self-attention WaveNet

Because Every Sensor Is Unique, so Is Every Pair: **Handling Dynamicity** in Traffic Forecasting

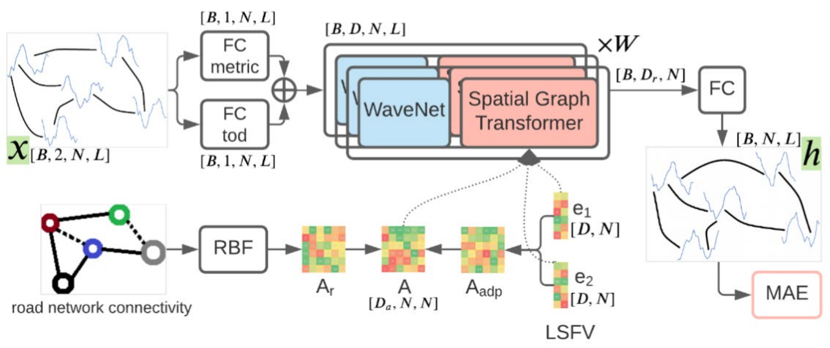


Fig. 4. System architecture of Graph Self-attention WaveNet (G-SWaN). Spatial Graph Transformers (SGT) is the novel module proposed that uses the node embeddings e_1 and e_2 to capture the unique sensor dynamics in the self-attention mechanisms. The notations are described in Table 1.

G-SWaN architecture

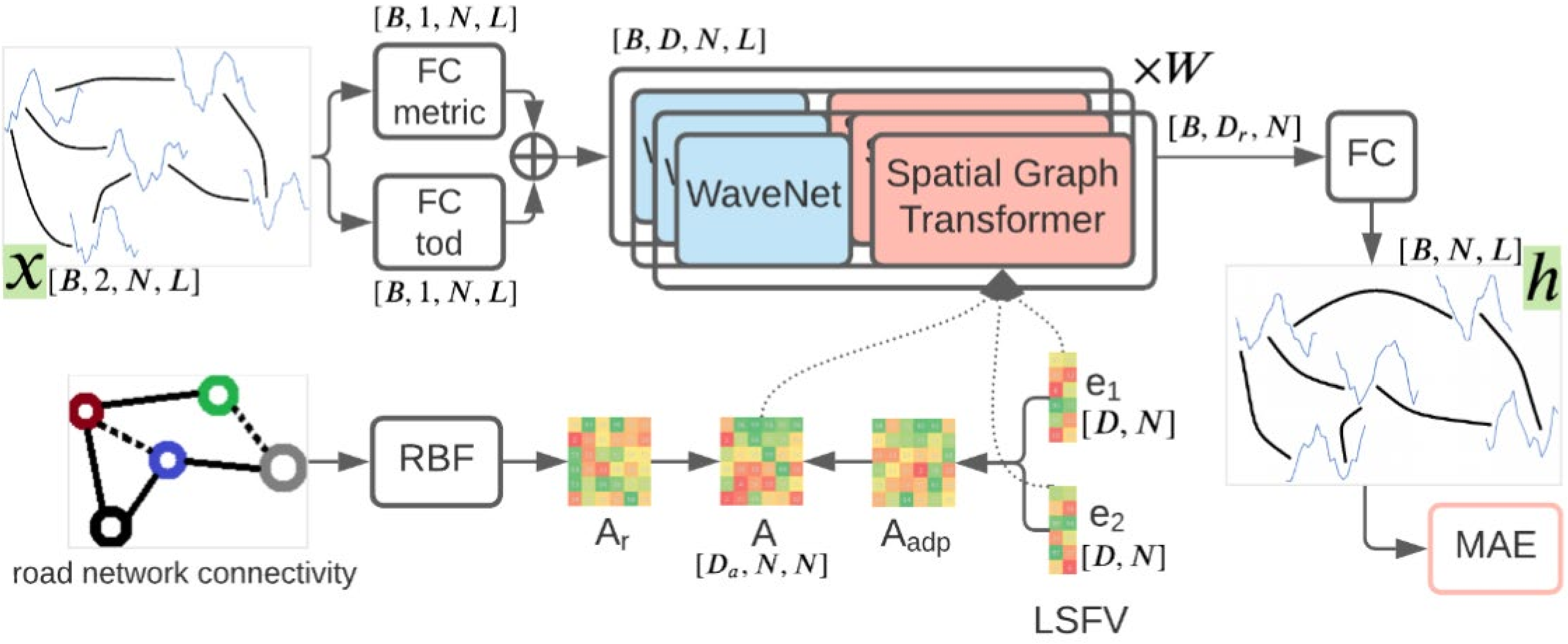


Fig. 4. System architecture of Graph Self-attention WaveNet (G-SWaN). Spatial Graph Transformers (SGT) is the novel module proposed that uses the node embeddings e_1 and e_2 to capture the unique sensor dynamics in the self-attention mechanisms. The notations are described in Table 1.

G-SWaN layer and the SGT module

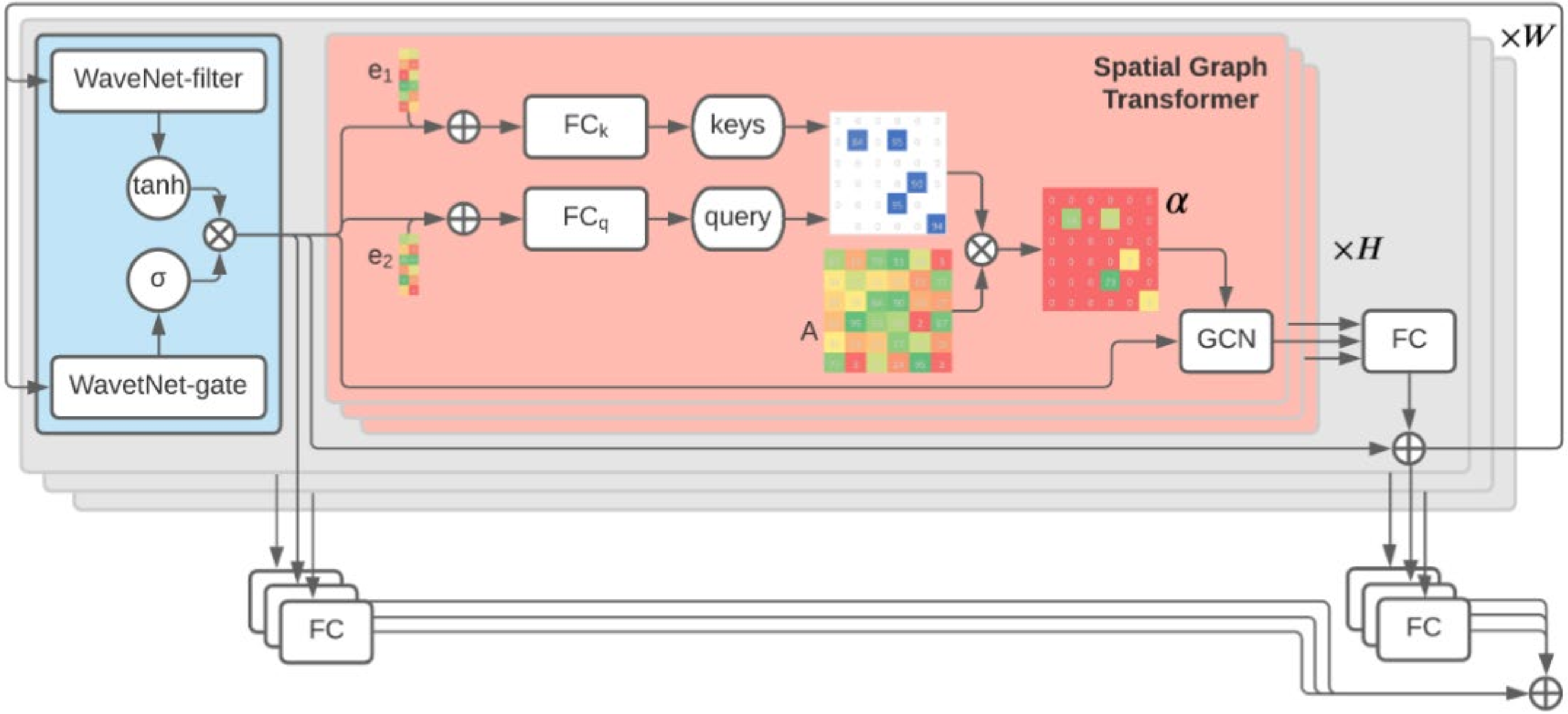
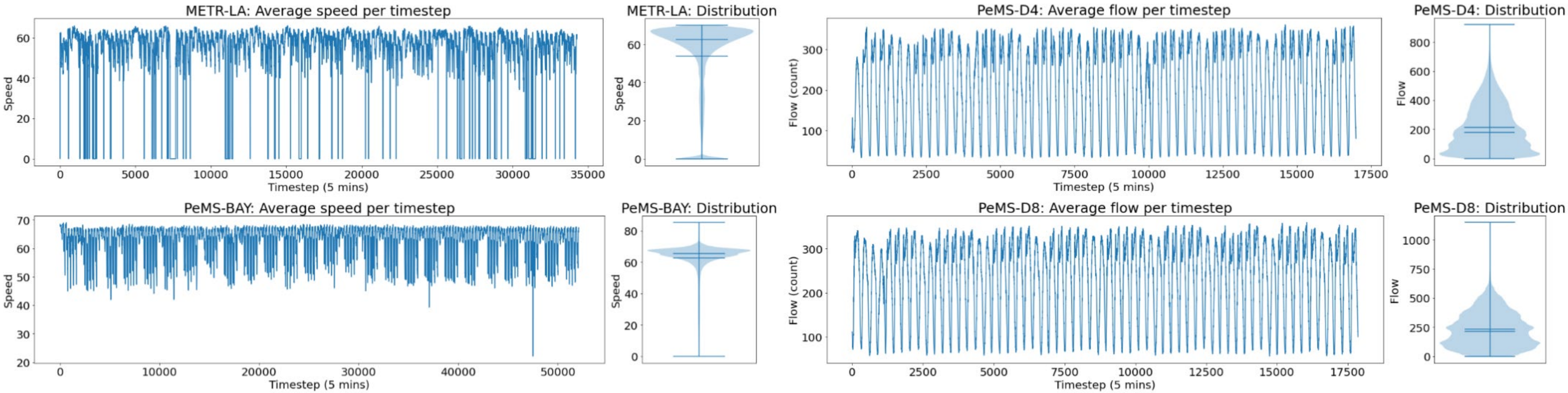


Fig. 5. A layer of G-SWaN contains a WaveNet temporal block (blue) and a Spatial Graph Transformer (SGT) block (red). SGT uses node embedding (e_1 and e_2) to apply multi-headed a query and key self-attention mechanism on the adjacency matrix A. This way, the self-attention mechanism is sensitive to the unique dynamics of every pair of sensors. There are H attention heads and W layers. Some details such as activation functions and batch normalization are also not shown.

Table 2. Dataset description.

Dataset	Spatial		Temporal		Value		Size	
	Sensors	Edges	Timesteps	Range (duration in days)	Metric	Mean ± Std	Entry	Compressed (MB)
METR-LA	207	1,515	34,272	1 Mar 12 - 30 Jun 12 (121)	speed	53.72± 20.26	7,094,304	54
PeMS-BAY	325	2,694	52,116	1 Jan 17 - 30 Jun 17 (180)	speed	62.61± 9.59	16,937,700	130
PeMS-D4	307	340	16,969	1 Jan 18 - 28 Feb 18 (58)	flow	211.70± 158.07	5,209,483	31
PeMS-D8	170	277	17,833	1 Jul 16 - 31 Aug 16 (61)	flow	230.68± 146.22	3,031,610	18



Results

Table 3. Performance comparison on speed metric using METR-LA and PeMS-BAY datasets. Since all the metrics are error metrics, lower means better. Prediction horizon = 15 / 30 / 60 minutes. **Bold** means the best performance within the metric. Underline means the second best performance.

(Metric: speed) Model	METR-LA MAE	METR-LA RMSE	METR-LA MAPE (%)	PeMS-BAY MAE	PeMS-BAY RMSE	PeMS-BAY MAPE (%)
HA	4.16	7.80	13.00	2.88	5.59	6.80
ARIMA	3.99/5.15/6.90	8.21/10.45/13.23	9.60/12.70/17.40	1.62/2.33/3.38	3.30/4.76/6.50	3.50/5.40/8.30
VAR	4.42/5.41/6.52	7.89/9.13/10.11	10.20/12.70/15.80	1.74/2.32/2.93	3.16/4.25/5.44	3.60/5.00/6.50
SVR	3.99/5.05/6.72	8.45/19.87/13.76	9.30/12.10/16.7	1.85/2.48/3.28	3.59/5.18/7.08	3.80/5.50/8.00
FNN	3.99/4.23/4.49	7.94/8.17/8.69	9.90/12.90/14.00	2.20/2.30/2.46	4.42/4.63/4.98	5.19/5.43/5.89
FC-LSTM	3.44/3.77/4.37	6.30/7.23/8.69	9.60/10.90/13.20	2.05/2.20/2.37	4.19/4.55/4.96	3.80/5.20/5.70
DCRNN	2.77/3.15/3.60	5.38/6.45/7.60	7.30/8.80/10.50	1.38/1.74/2.07	2.95/3.97/4.74	2.90/3.90/4.90
STGCN	2.88/3.47/4.59	5.74/7.24/9.40	7.62/9.57/12.70	1.36/1.81/2.49	2.96/4.27/5.69	2.90/4.17/5.79
Graph WaveNet	2.69/3.07/3.53	5.15/6.22/7.37	6.90/8.37/10.01	**1.30**/1.63/1.95	2.74/3.70/4.52	2.73/3.67/4.63
ST-MetaNet	2.69/3.10/3.69	5.17/6.28/7.52	6.91/8.57/10.63	1.36/1.76/2.20	2.90/4.02/5.06	2.82/4.00/5.45
ASTGCN	4.86/5.43/6.51	9.27/10.61/12.52	9.21/10.13/11.64	1.52/2.01/2.61	3.13/4.27/5.42	3.22/4.48/6.00
STSGCN	3.31/4.13/5.06	7.62/9.77/11.66	8.06/10.29/12.91	1.44/1.83/2.26	3.01/4.18/5.21	3.04/4.17/5.40
AGCRN	2.87/3.23/3.62	5.58/6.58/7.51	7.70/9.00/10.38	1.37/1.69/1.96	2.87/3.85/4.54	2.94/3.87/4.64
GMAN	2.80/3.12/**3.44**	5.55/6.49/7.35	7.41/8.73/10.07	1.34/1.63/**1.86**	2.91/3.76/**4.32**	2.86/3.68/**4.37**
MTGNN	2.69/3.05/3.49	5.18/6.17/**7.23**	6.86/8.19/9.87	1.32/1.65/1.94	2.79/3.74/4.49	2.77/3.69/4.53
G-SWaN (ours)	**2.65/3.02**/3.47	**5.05/6.12**/7.27	**6.72/8.13/9.86**	**1.30/1.61**/1.91	**2.72/3.64**/4.37	**2.69/3.62**/4.49

Results

Table 4. Performance comparison on flow metric using PeMS-D4 and PeMS-D8 datasets. Since all the metrics are error metrics, lower means better. **Bold** means the best performance within the metric. Underline means the second best performance.

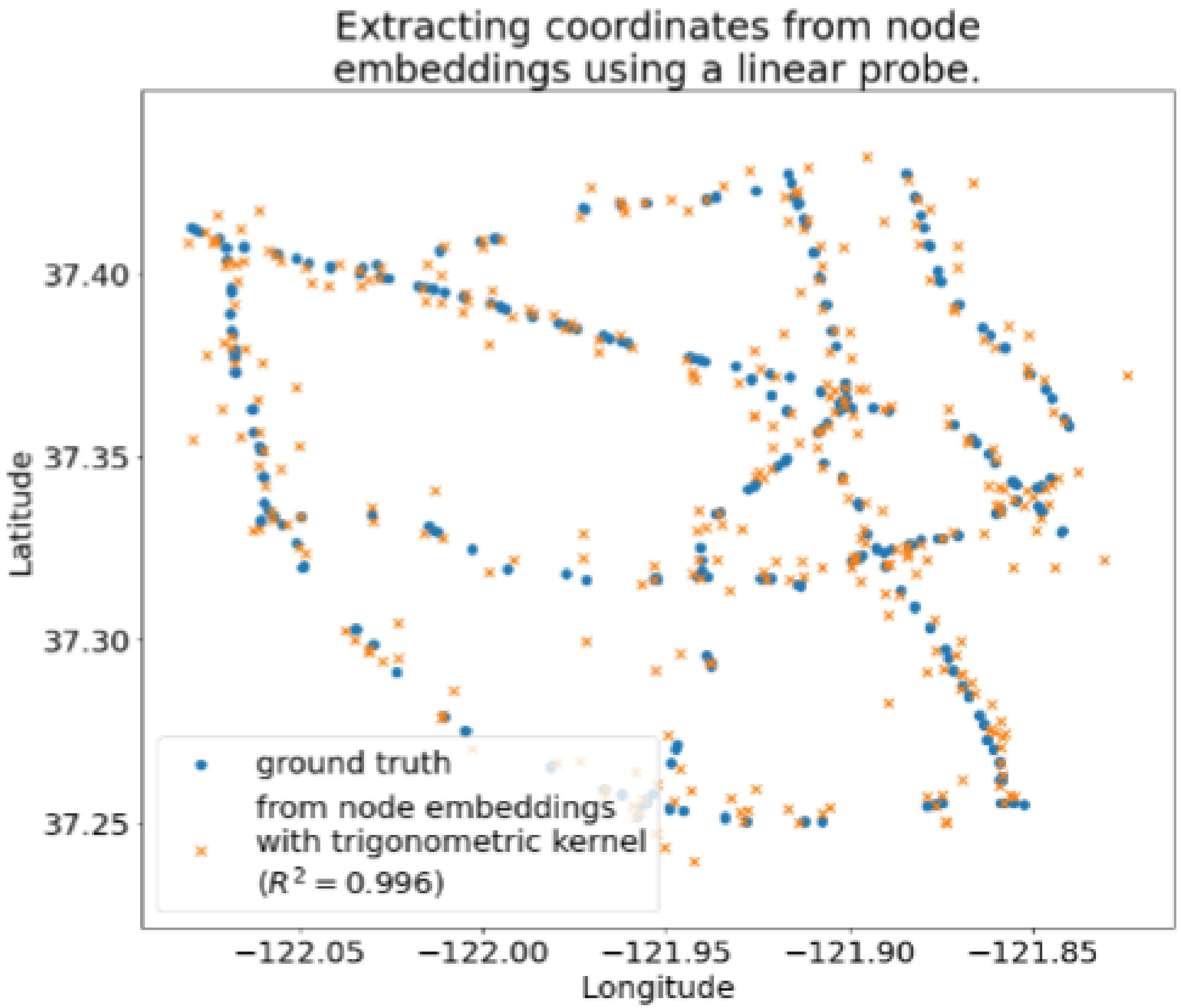
(Metric: flow)	PeMS-D4			PeMS-D8		
Model	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA	38.03	59.24	27.88	34.86	52.04	24.07
VAR	24.54	38.61	17.24	19.19	29.81	13.10
GRU-ED	23.68	39.27	16.44	22.00	36.23	13.33
DSANet	22.79	35.77	16.03	17.14	26.96	11.32
DCRNN	21.22	33.44	14.17	16.82	26.36	10.92
STGCN	21.16	34.89	13.83	17.50	27.09	11.29
Graph WaveNet	28.98	42.08	30.80	20.52	30.04	16.20
ASTGCN	22.93	35.22	16.56	18.25	28.06	11.64
STSGCN	21.19	33.65	13.90	17.13	26.86	10.96
AGCRN	<u>19.83</u>	<u>32.26</u>	<u>12.97</u>	<u>15.95</u>	<u>25.22</u>	<u>10.09</u>
G-SWaN (ours)	**18.48**	**30.51**	**12.59**	**14.05**	**23.00**	**9.08**

Ablation

Table 5. Ablation study on PeMS-D8. **Bold** means the worst performance, showing the importance of the missing component. Underline means the second worst.

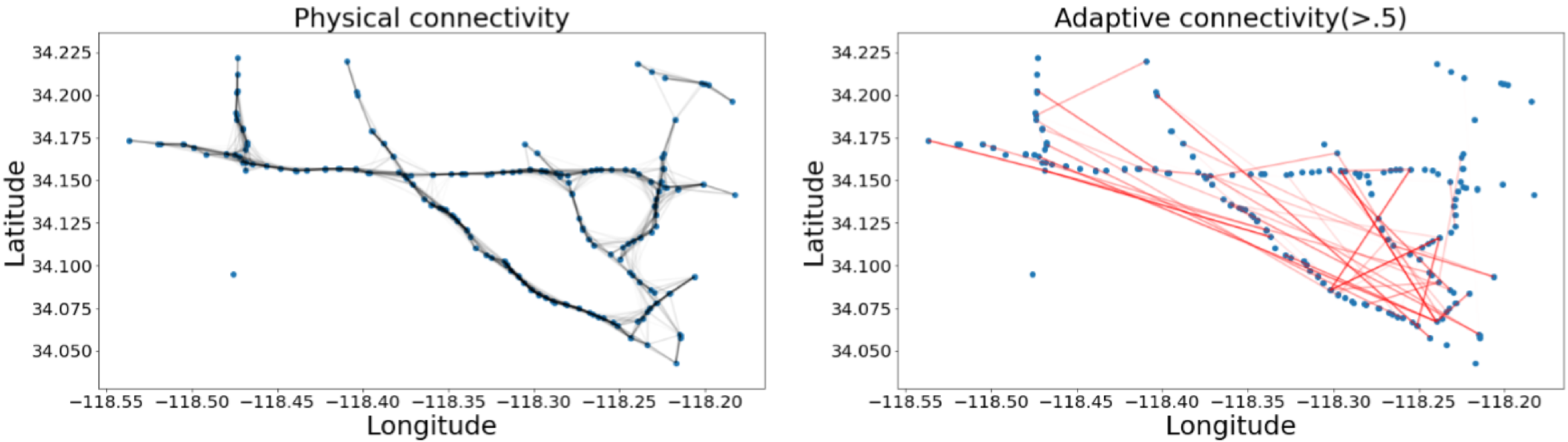
Model	MAE	MAPE (%)	RMSE
G-SWaN	14.05	9.08	23.00
w/o spatial occlusion	14.12	9.14	23.10
w/o temporal permutation	14.21	9.18	23.15
w/o uniform noise	14.11	9.22	23.15
w/o node embeddings	14.29	9.22	23.11
Single head attention	14.21	9.14	23.05
GCN w/o SGT	**14.62**	**9.52**	**23.34**

Node embeddings encode the coordinate from the traffic dynamics alone.



(a) Recovering sensor coordinates from node embeddings using a linear probe with trigonometric kernels.

Adaptive connectivity connect distant nodes.



(b) Qualitative comparison between physical and adaptive adjacency matrix in METR-LA dataset. The line transparency is proportional to the edge weight.

Thank You. Any Questions?

G-SWaN

Because
Every Sensor Is Unique,
so Is Every Pair:
Handling Dynamicity in
Traffic Forecasting

Link to GitHub https://github.com/aprbw/g-swan

Arian Prabowo,
Wei Shao,
Hao Xue,
Piotr Koniusz, and
Flora D. Salim

Traffic Forecasting: Problem Definition