Motorway Traffic Flow Prediction using Advanced Deep Learning

by A.S. Mihaita, Haowen Li, Zongyang He, Marian-Andrei Rizoiu

SPEAKER'S NAME:

Adriana-Simona MIHAITA
University of Technology Sydney, Australia

29th of October 2019

Auckland, New Zealand
IEEE ITSC 2019

www.fmlab.org

Summary

1. **Introduction**

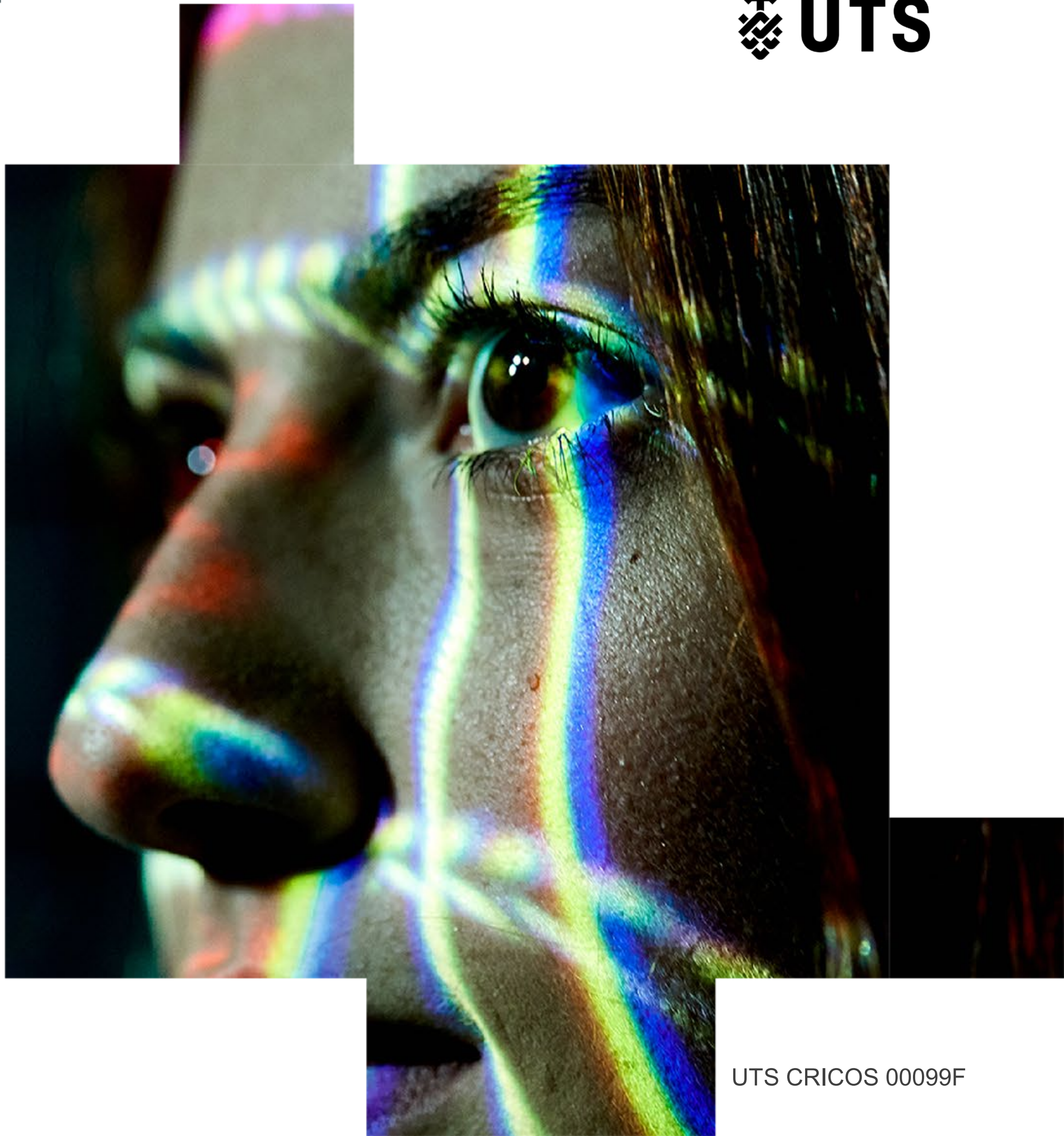
2. **Methodology**
    1. Network identification and data set preparation
    2. Data profiling and Outlier Identification
    3. Feature Construction
    4. Deep Learning Model Development

3. **Case Study**
    1. Sydney M7 motorway traffic flow
    2. Daily Profile and traffic flow map

4. **Experimental setup**

5. **Results**
    1. Model performances
    2. Past vs future time horizon analysis

6. **Conclusions**



Introduction

Globally:

Traffic congestion has reached unprecedented peaks in majority of large urban areas in the world!

Top 10 most congested cities have reached up to 140% of congestion in October 2019.

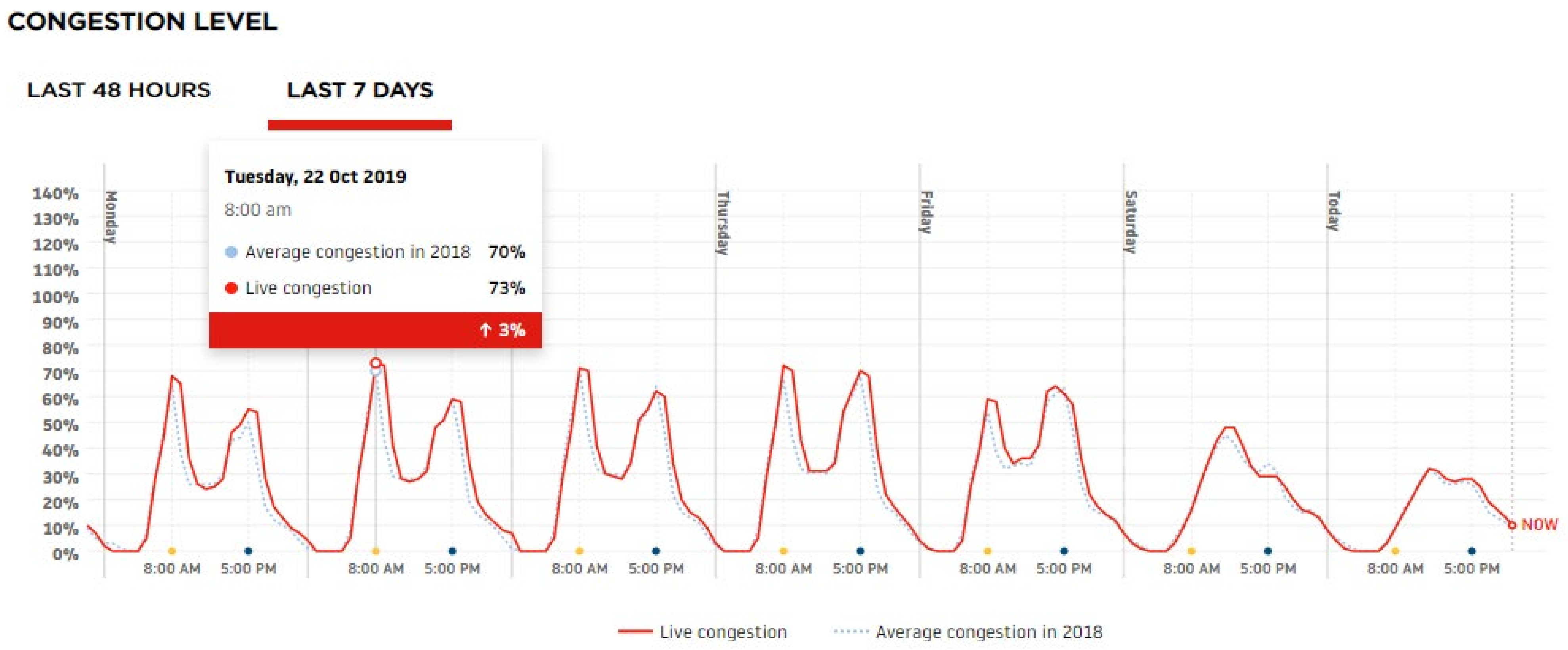
1	1	**Mumbai**	India	**65%**	↓ 1%
2	2	**Bogota**	Colombia	**63%**	↑ 1%
3	3	**Lima**	Peru	**58%**	↑ 8%
4	4	**New Delhi**	India	**58%**	↓ 4%
5	5	**Moscow region (oblast)**	Russia	**56%**	↓ 1%
6	6	**Istanbul**	Turkey	**53%**	↓ 6%
7	7	**Jakarta**	Indonesia	**53%**	↓ 8%
8	8	**Bangkok**	Thailand	**53%**	↓ 2%
9	9	**Mexico City**	Mexico	**52%**	– 0%

Source: TomTom Live Congestion Index https://www.tomtom.com/en_gb/traffic-index/ranking

Introduction

Australia:

- Sydney is the country's most congested city when average speeds are compared to free-flow speeds
- The cost of congestion to the national economy is projected to rise to **$37.3 billion by 2030** without major policy changes



Source: TomTom Live Congestion Index https://www.tomtom.com/en_gb/traffic-index/ranking

Open Questions:

a) **How to efficiently predict road traffic congestion using extensive data-driven techniques which can adapt to real-time big-data sets?**

- Parametric models: Kalman filtering, ARIMA[2]. SARIMA[4], ARIMAX[4], etc. Stochastic and disruptive events can affect accuracy of parametric models.
- Non-parametric models : k-nearest neighbours [6], support vector regressions [7], artificial neural networks, [8], Gaussian Processes [9], etc.

b) **What are the best techniques that can capture the spatial-temporal correlations arising in complex traffic networks?**

Deep Learning gained increase popularity over the last years: CNN, LSTM, hybrid modelling

c) **Why are some models efficient for short-term traffic prediction, but not for long-term prediction?**

[2] M. V. D. Voort, M. Dougherty, and S. Watson, “Combining kohonen maps with arima time series models to forecast traffic flow,” Transportation Research Part C, vol. 4, no. 5, pp. 307 – 318, 1996.
[4] A. M. Khoei, A. Bhaskar, and E. Chung, “Travel time prediction on signalised urban arterials by applying sarima modelling on Bluetooth data,” in Australasian Transport Research Forum, 2013.
[5] B. M. Williams, “Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling,” Trans. Res. Rec., vol. 1776, no. 1, pp.194–200, 2001.
[6] H. Chang, Y. Lee, B. Yoon, and S. Baek, “Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences,” IET Intel. Transport Systems, vol. 6, no. 3, pp. 292–305, 2012.
[7] Y. Jeong, Y. Byon, M. M. Castro-Neto, and S. M. Easa, “Supervised weighting-online learning algorithm for short-term traffic flow prediction,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1700–1707, 2013.
[8] M. Karlaftis and E. Vlahogianni, “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights,” Trans. Research Part C, vol. 19, no. 3, pp. 387 – 399, 2011.
[9] D. Ide and S. Kato, “Travel-time prediction using gaussian process regression: A trajectory,” in International Conference on Data Mining, 2009, pp. 1185–1196.

Deep Learning Challenges:

a) Scalability of DL models for large scale and real-life deployment

b) Relationship between the training and the prediction horizons

c) Deploying hybrid deep learning models that combine both spatial and temporal modelling

Summary

1. Introduction

2. **Methodology**
    1. Network identification and data set preparation
    2. Data profiling and Outlier Identification
    3. Feature Construction
    4. Deep Learning Model Development

3. Case Study
    1. Sydney M7 motorway traffic flow
    2. Daily Profile and traffic flow map

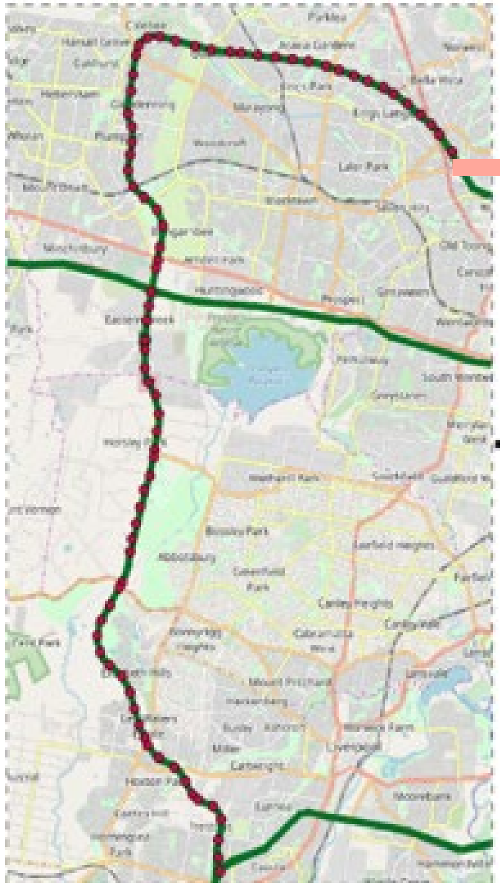
4. Experimental setup

5. Results
    1. Model performances
    2. Past vs future time horizon analysis

6. Conclusions

Methodology

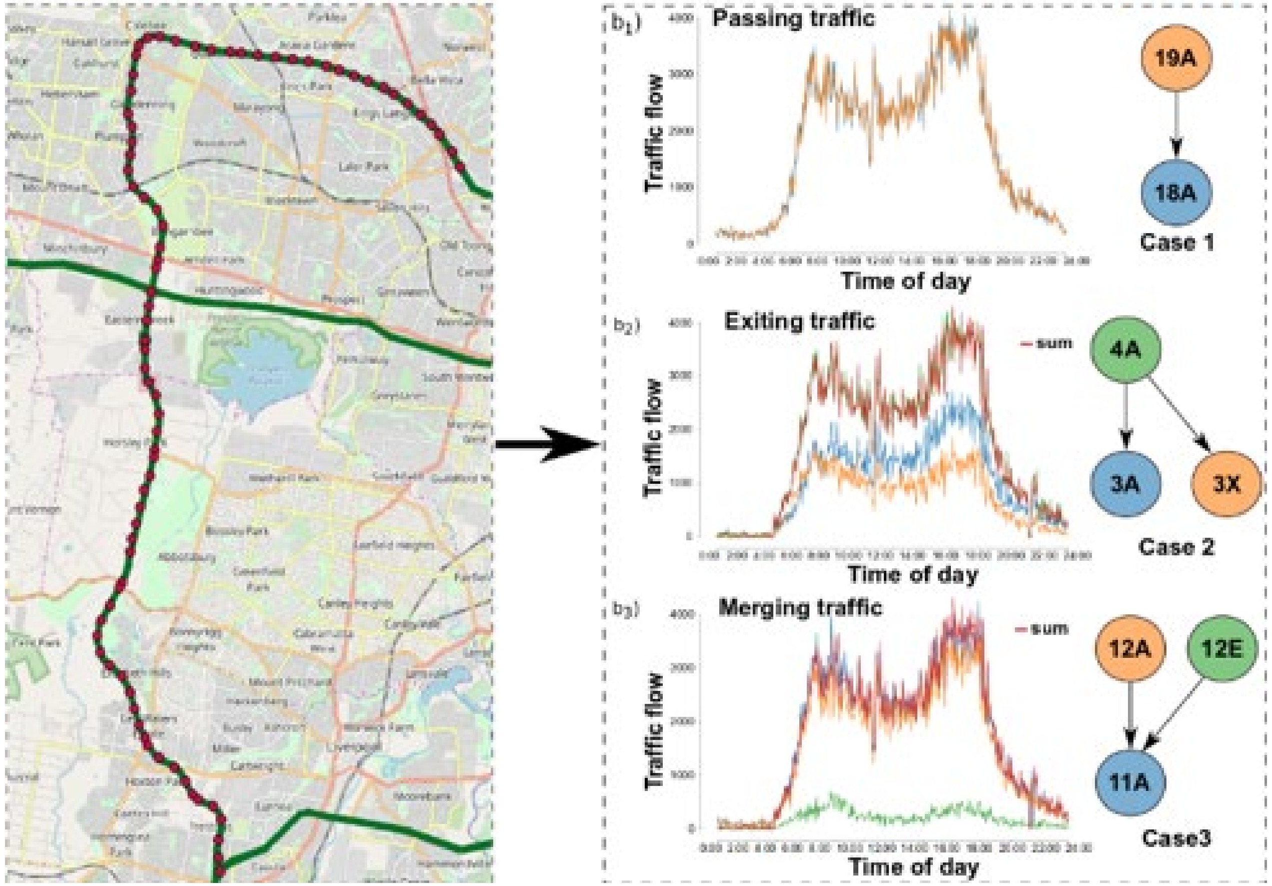
a) Motorway road insfrastructure



Methodology

a) Motorway road insfrastructure

b) Data profiling and Outlier identification

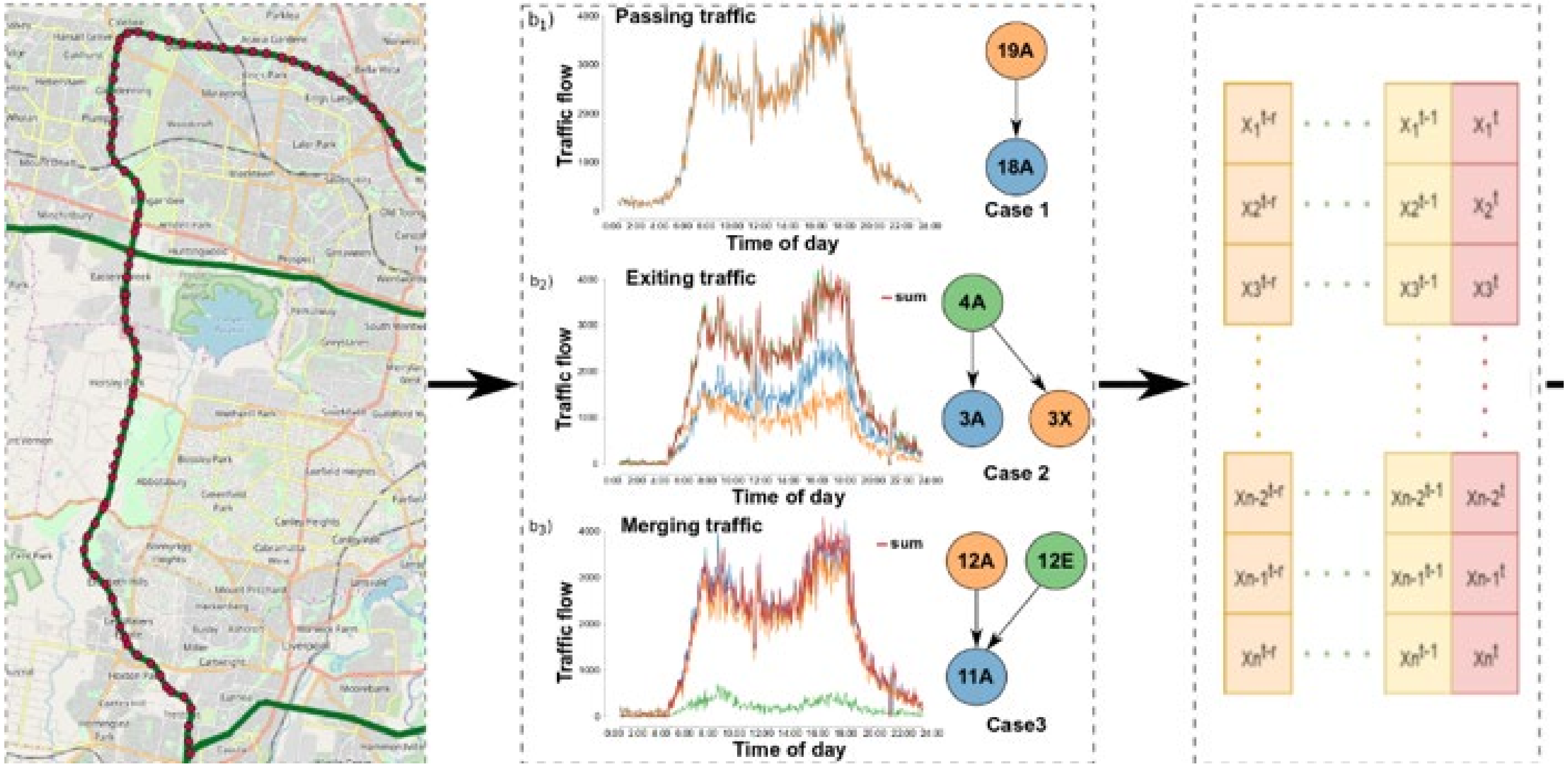


Methodology

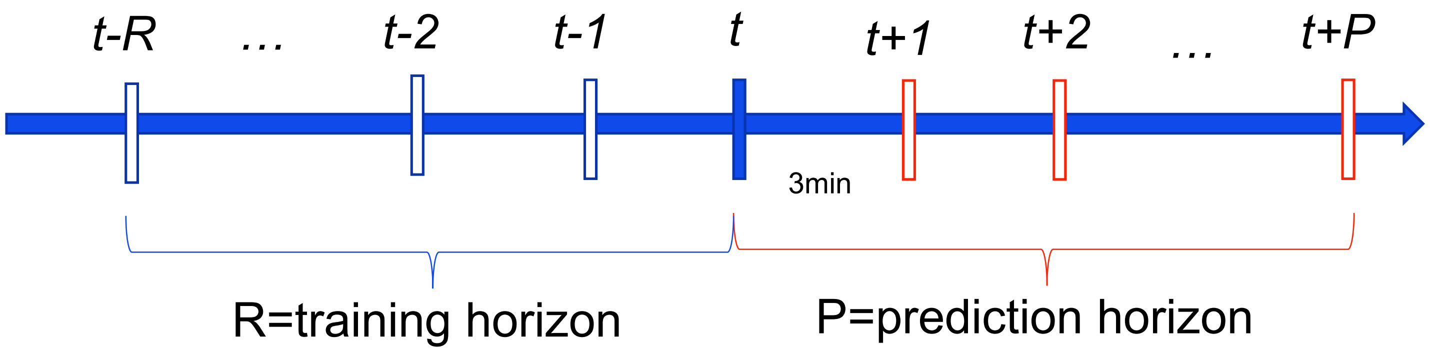
a) Motorway road insfrastructure

b) Data profiling and Outlier identification

c) Multiple couting stations Feature generation



Methodology



$$X^t = \begin{bmatrix} \vec{X}_1^{\,t} \\ \vec{X}_2^{\,t} \\ \dots \\ \vec{X}_N^{\,t} \end{bmatrix} = \begin{bmatrix} x_1^{t-R+1} & \dots & x_1^{t-1} & x_1^t \\ x_2^{t-R+1} & \dots & x_2^{t-1} & x_2^t \\ \dots & \dots & \dots & \dots \\ x_N^{t-R+1} & \dots & x_N^{t-1} & x_N^t \end{bmatrix}$$

$$\hat{X}^{t+P} = [\hat{X}_1^{t+P}, \hat{X}_2^{t+P}, \dots, \hat{X}_n^{t+P}]^T$$

TABLE I: Summary of notations.

Notation	Interpretation
N	the total number of stations used in this study ($N =$ 208 comprised of 104 in each direction).
TI	a 3-min *Time Interval*; the time is discretized into 3-minute time intervals (480 TI per day).
R	the length of the time window in the past; the number of TI used as historic information.
P	future prediction horizon; predictions will be made for the P^{th} TI in the future.
x_j^{t-i}	the traffic flow of station j at TI $= t-i, i \in \{0, \dots R\}$.
$\vec{S}^{t-i}$	the traffic flow for ALL stations at TI $= t-i, i \in \{0, \dots R\}$; an N-dimensional column vector $\vec{S}^i = [x_1^{t-i}, x_2^{t-i}, \dots, x_N^{t-i}]^T$.
X^t	the observed traffic flow, for all stations for the past R TI, an $R \times N$ matrix (see Eq. (1)).
$RMSE$	the Root Mean Square Error evaluation metric; $RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(f_j - \hat{f}_j)^2}$
$ReLU(x)$	The ReLU function; $ReLU(x) = max(x, 0)$

Methodology

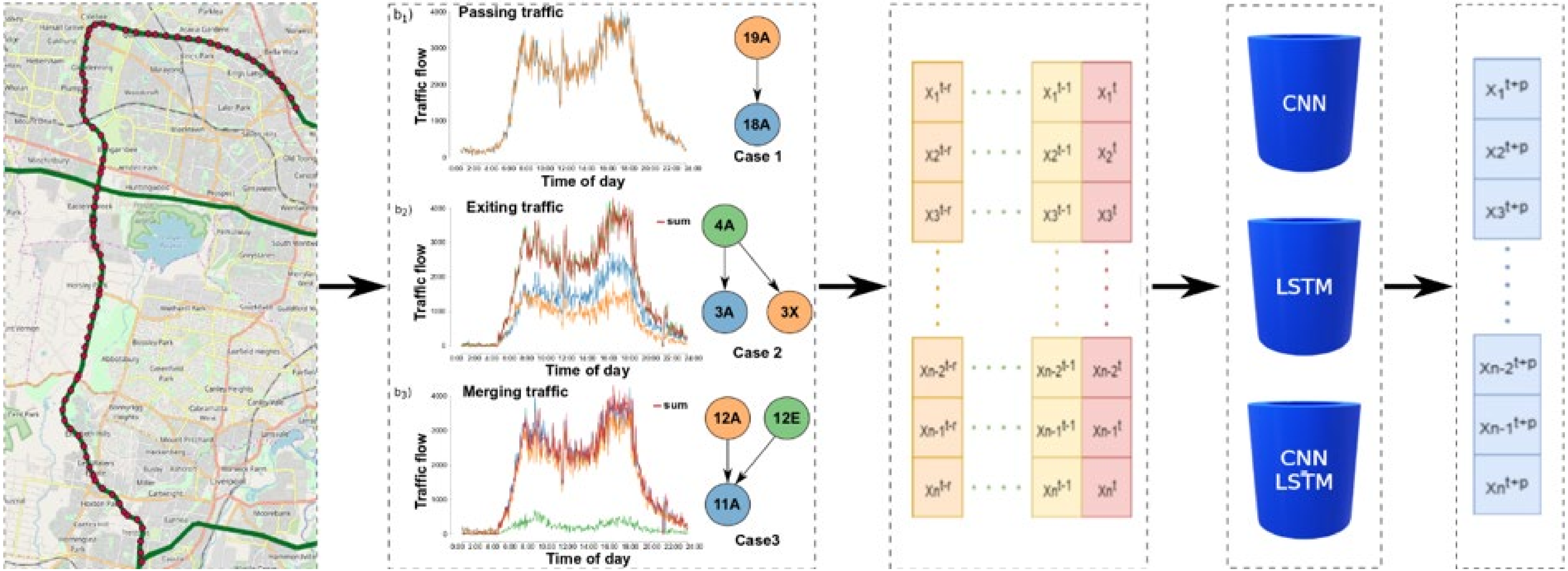
a) Motorway road insfrastructure

b) Data profiling and Outlier identification

c) Multiple couting stations Feature generation

d) Deep Learning Prediction

e) Prediction Output



Methodology

a) Back-propagation Neuronal Networks (BPNN):

BPNN consists of two fully-connected layers. The input of first layer is the historical information of all stations, and the last layer's output is the prediction of the traffic flow across all monitoring stations.

In this work, BPNN is mainly used as a lower bound DL performance measure, and it serves to assess the performance gains obtained when implementing the more complex models detailed here below.

b) Convolutional Neural Networks (CNN):

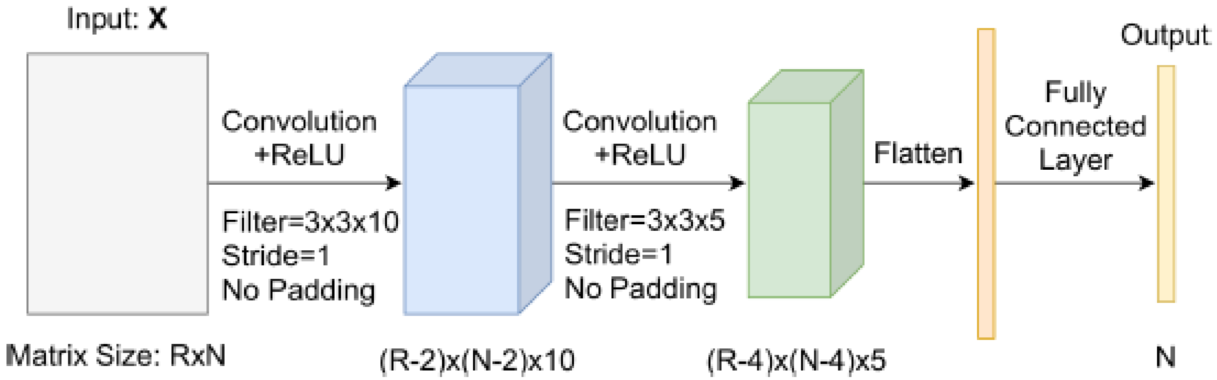


Fig. 2: CNN model for traffic flow prediction.

Methodology

c) Long Short-Term Memory Networks (LSTM):

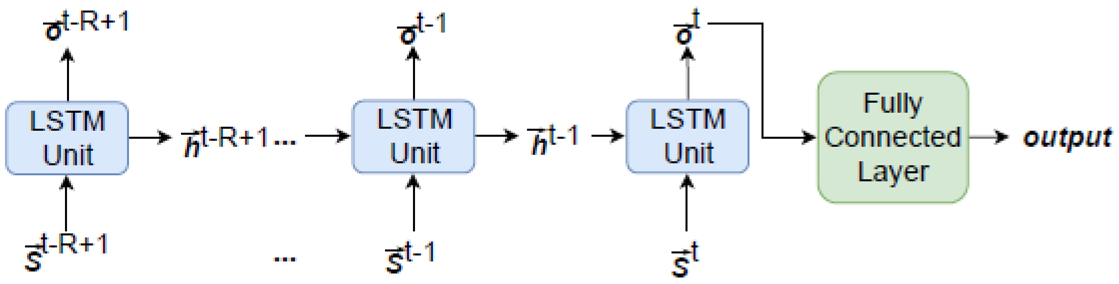


Fig. 3: LSTM model for traffic flow prediction.

d) Hybrid CNN-LSTM prediction:

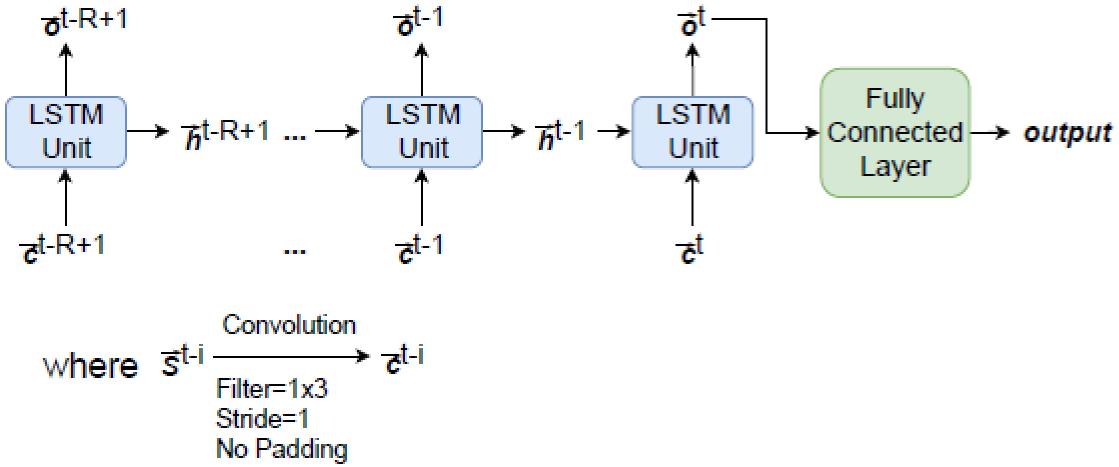


Fig. 4: CNN-LSTM hybrid model for traffic flow prediction.

Summary

1. Introduction

2. Methodology
    1. Network identification and data set preparat
    2. Data profiling and Outlier Identification
    3. Feature Construction
    4. Deep Learning Model Development

3. **Case Study**
    1. Sydney M7 motorway traffic flow
    2. Daily Profile and traffic flow map

4. Experimental setup

5. Results
    1. Model performances
    2. Past vs future time horizon analysis

6. Conclusions

Daily profiling and outlier identification

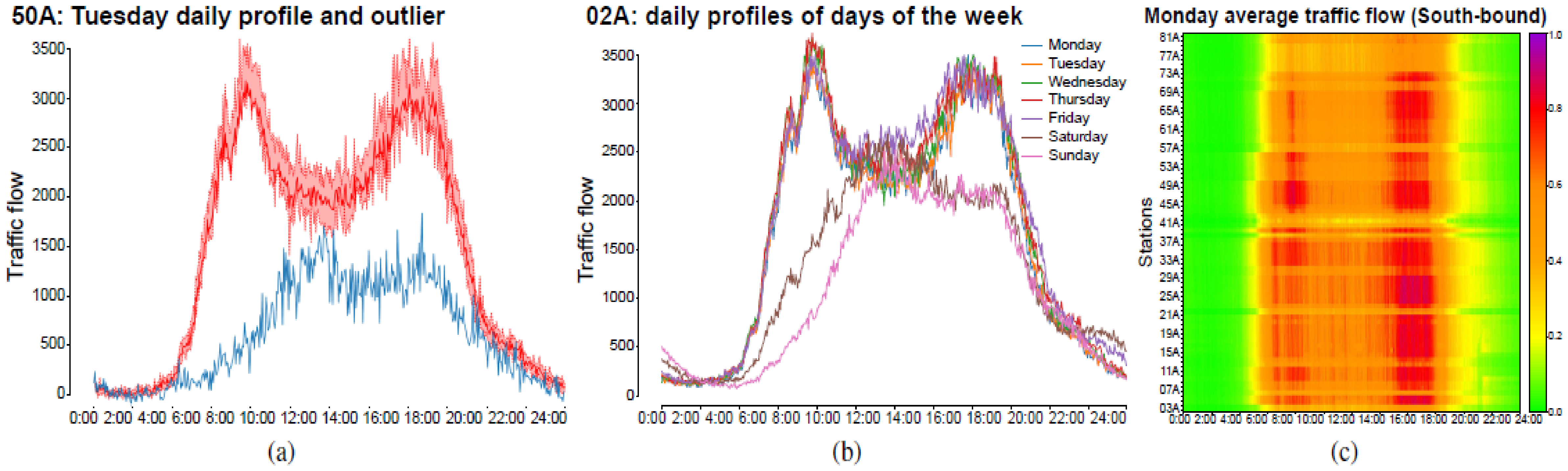


Fig. 5: (a) **Constructing the daily profile.** Mean (solid line) and the 20% – 80% percentiles (red area) for the traffic flow series for the station 50A, computed on the period 2017-02-01 to 04-30; (b) **Daily profiles for days of the week.** The daily profiles for station 02A for each of the days of the week, computed for the same period of time. (c) **Daily profiles for all stations – the Traffic Flow Congestion Map.** The colormap of the Monday traffic flow for all 104 south-bound (A) stations is calculated based on a Flow/Capacity ratio and ranges between 0 and 1.

Summary

Experimental Set-up

1) Prediction setup - 36.34 million data points

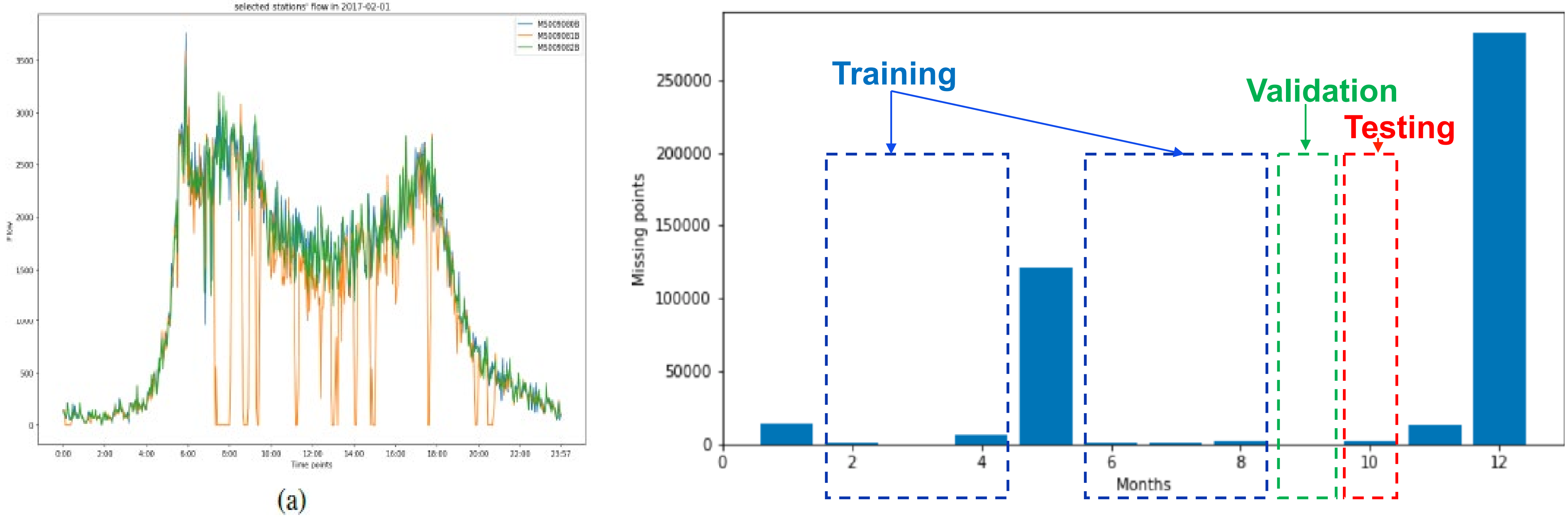


Fig. 9: Missing data in the Sydney Motorway traffic flow dataset. (a) The traffic flow for 1st of February 2017, for three contiguous stations (80B, 81B and 82B) with no entries and exits in between. 81B is showing missing data. (b) The total number of missing data points, aggregated per month.

Experimental Set-up

2) Other baseline models:

1. **Daily Profile Prediction (DPP)**

2. **BPNN for separate station prediction** (**Sep-BPNN**) – applied separately for each station; each model has 10 Hidden layers

3. **ARIMA**(p = 2; d = 1; q = 0) – after selection from p = {1..5}, d in{1..5}, q in {0..3}
   1. p is the parameter of the autoregression
   2. D is for the degree of differencing (the number of times the data have had past values subtracted) and
   3. q controls the moving average.

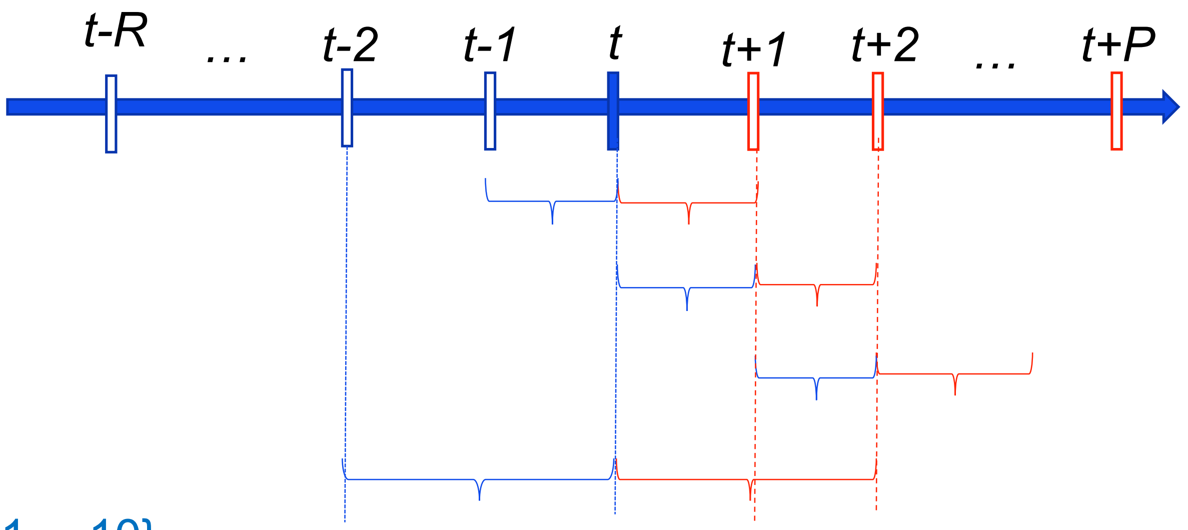
Total: 7 comparisons: DPP, ARIMA, Sep-BPNN, BPNN, CNN, LSTM, CNN-LSTM

Experimental Set-up

3) DL implementation and hyper-parameter selection:

By varying *t* on a dataset with n time points, we obtain n-R-P+1 pairs of inputs and outputs.

Ex1. R=2, P=1



Ex2. R=3, P=2

Our experimental range : R = {1…30}, P = {1,…10}

Total:

- Training : (42,721-R-P) + (44,161-R-P) combinations (2 contiguous training periods),
- Validation set: (14,401-R-P) pairs
- Test set: (14,881-R-P) pairs.

Experimental Set-up

c) DL implementation and hyper-parameter selection:

- Hyper parameters are tuned on the validation data set:
- we vary the batch size in the range [20;30;40; …75;100] and we obtain a value of 50.

- learning rate is 0:0003 and the weight of the L2 regularisation term is 10-8.
- implementation in PyTorch [17], using the Adam optimiser which provided a better performance than SGD or AdaGrad.

Given a value of the prediction time horizon P, we train the model 5 times and we calculate the average accuracy on the validation dataset. We select as the best R the value that achieves the highest average accuracy for the current P.

Open questions:

1. how much should we learn from the past to achieve best prediction results?
2. how long in the future should we predict?
3. is the size of the past horizon affecting the prediction results?
4. what is the relation between R, P and the performances of the advanced DL models?

Experimental Set-up

4) Performance evaluation:

- Root Mean Square Error (**RMSE**),
- Mean Absolute Error (**MAE**) and
- Symmetric Mean Absolute Percentage Error (**SMAPE**).

TABLE II: The time spent on training our models [sec]

	BPNN	CNN	LSTM	CNN-LSTM
Mean	101.190	219.452	302.105	382.538
Std	28.304	61.610	100.960	107.722

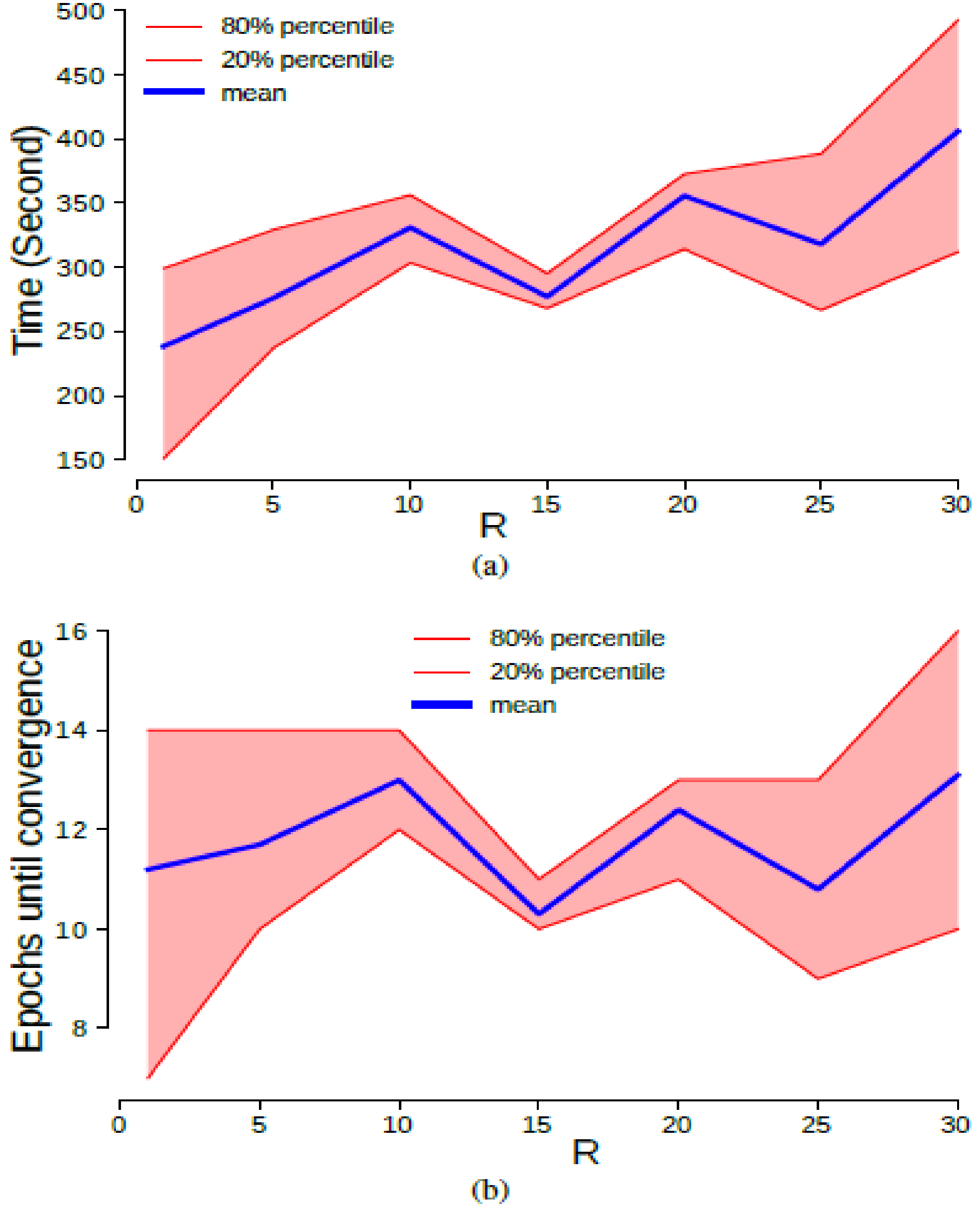


Fig. 10: Training time (a) and epochs to convergence (b) required by LSTM, with multiple values of R. The shaded area indicates the $20\% - 80\%$ percentiles interval.

UTS

Summary

1. Introduction

2. Methodology
    1. Network identification and data set preparat
    2. Data profiling and Outlier Identification
    3. Feature Construction
    4. Deep Learning Model Development

3. Case Study
    1. Sydney M7 motorway traffic flow
    2. Daily Profile and traffic flow map

4. Experimental setup

5. **Results**
    1. Model performances
    2. Past vs future time horizon analysis

6. Conclusions

Results – Model performance

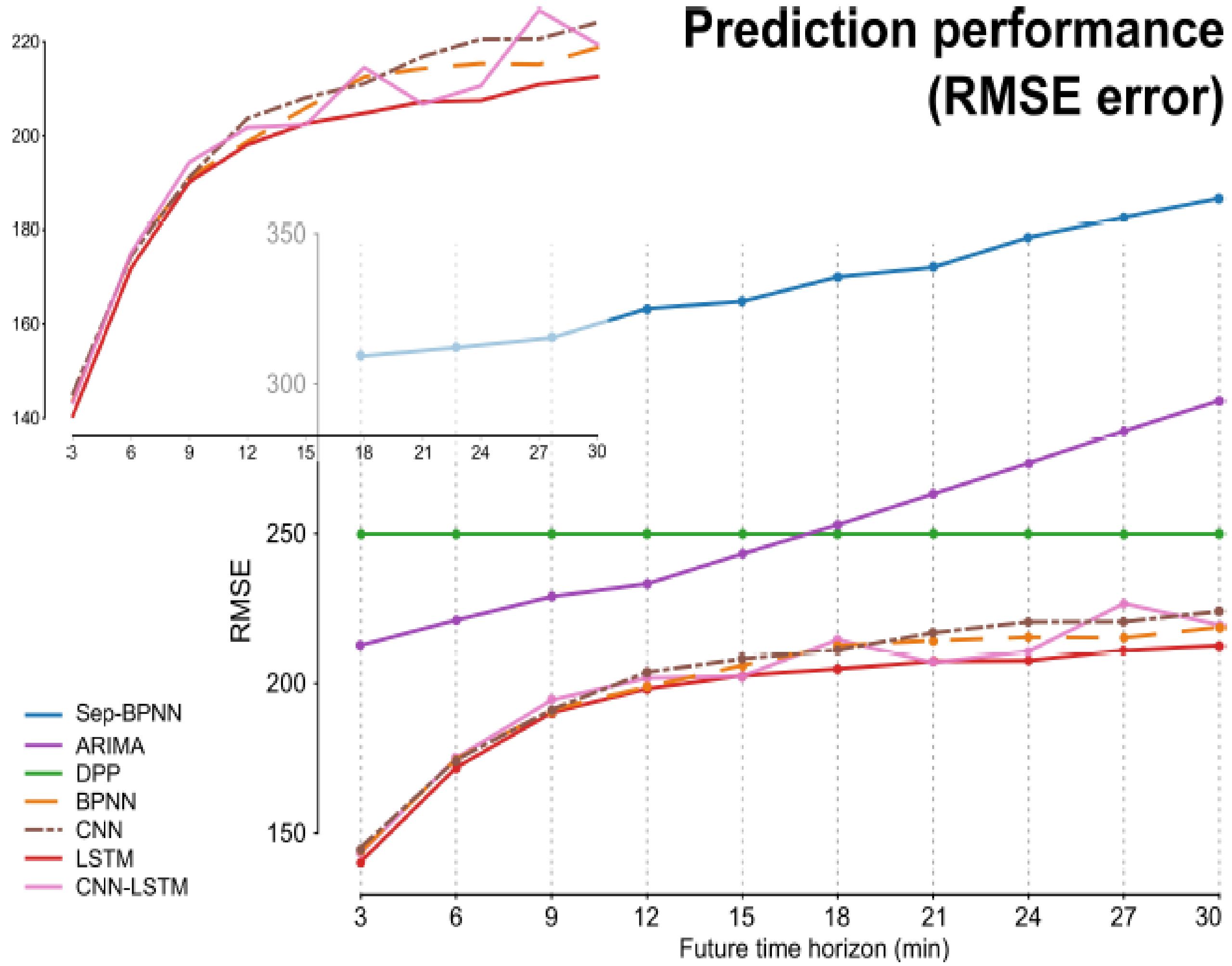
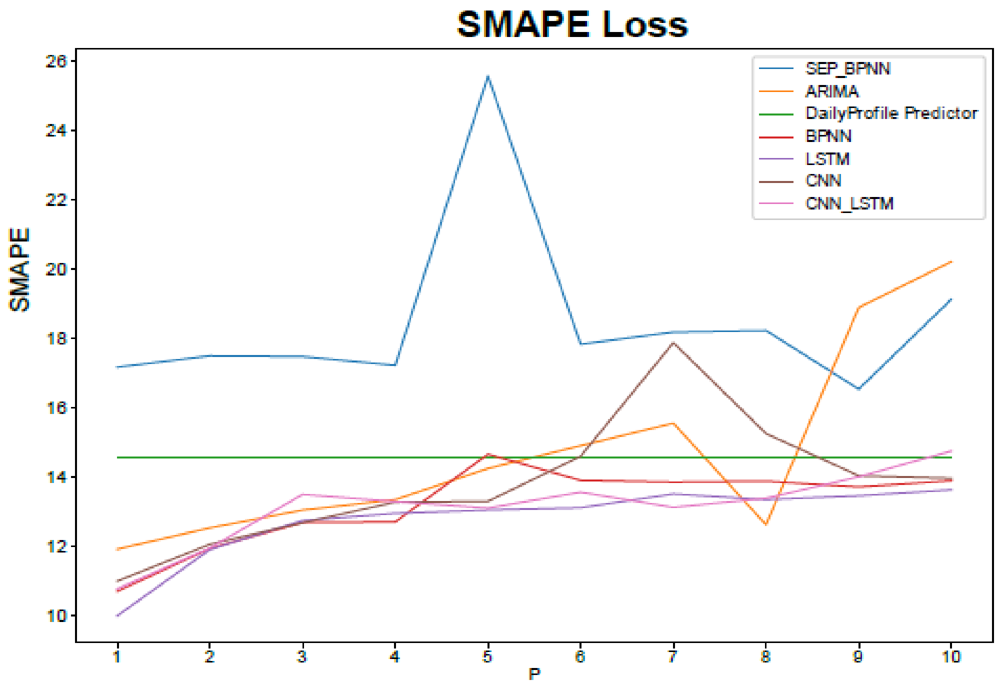
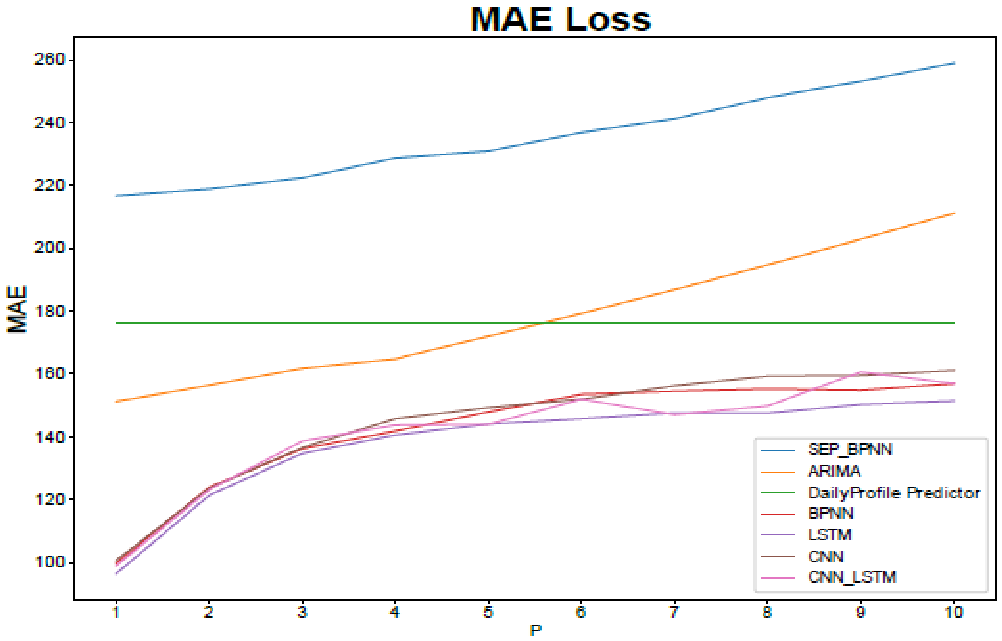


Fig. 7: Prediction performance for all models (Oy axis), for increasing future time-horizons P (Ox axis). The zoom expands the performance of DL models. The y-axes show the RMSE of each model (lower is better).





Results – Residual analysis

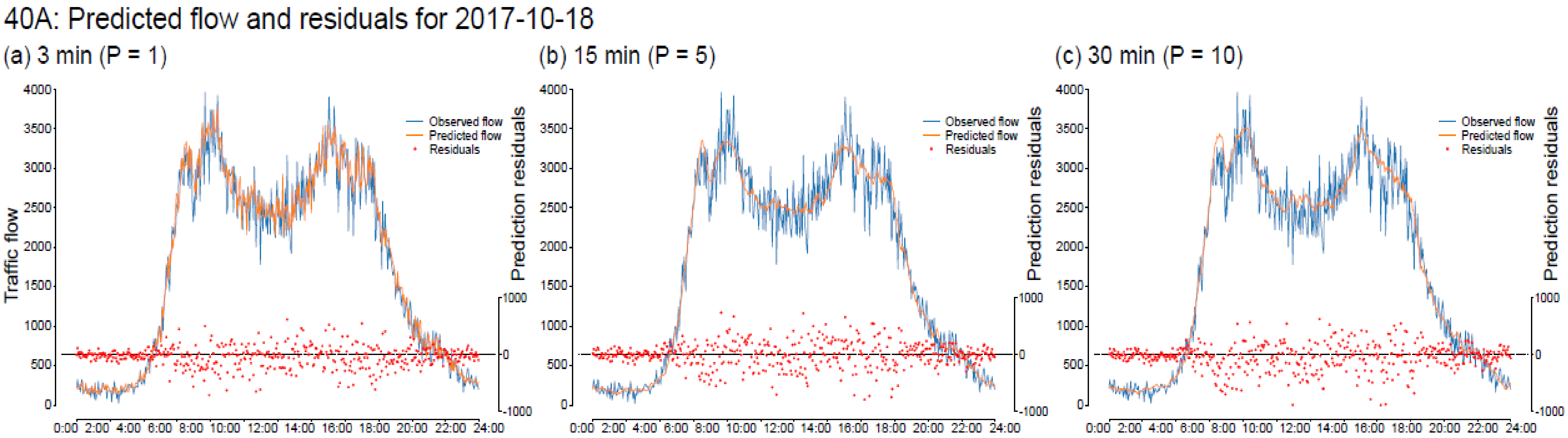


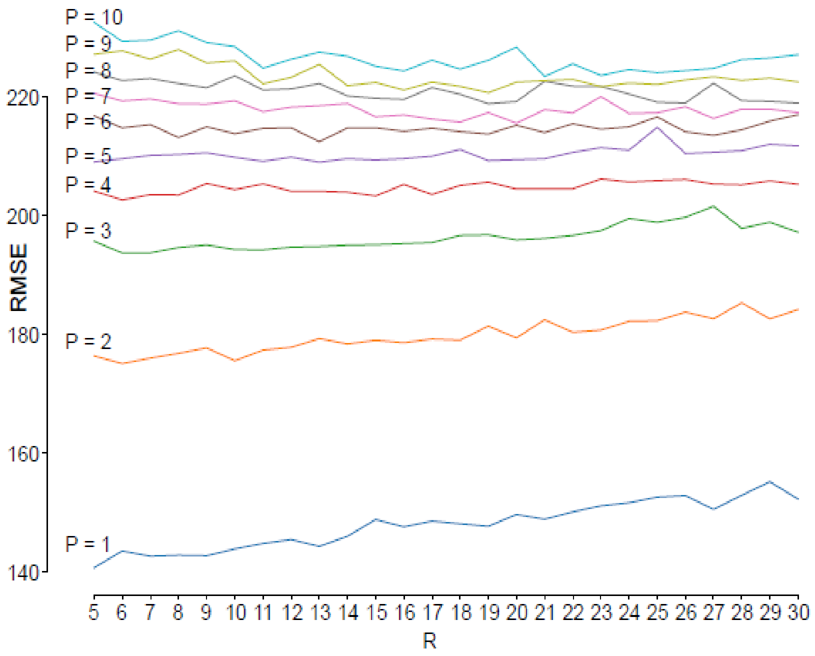
Fig. 6: Observed and predicted traffic flow, and residuals for 3 min **(a)**, 15 min **(b)** and 30 min **(c)** for station 40A on a weekday.

Max error of 10.8% in AM/P peaks

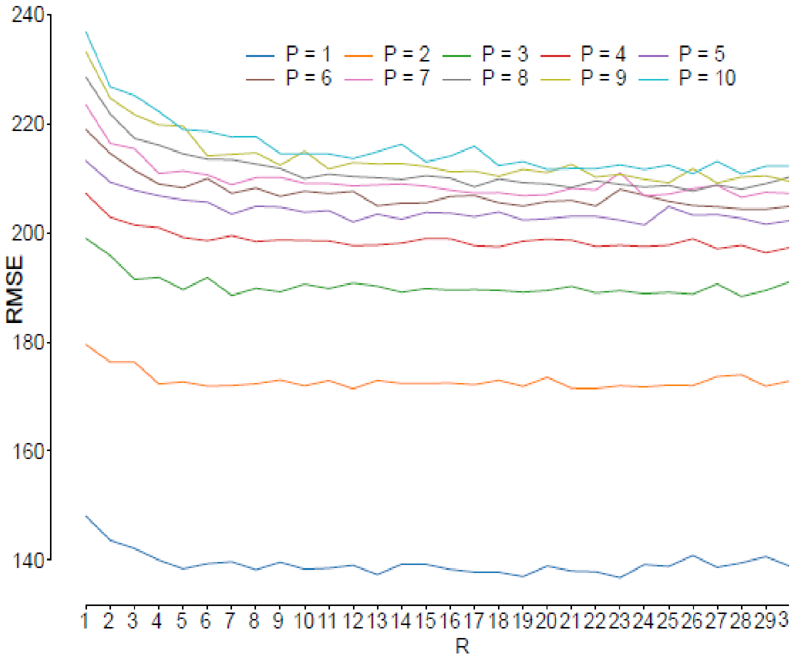
Results – Residual analysis

Prediction performance (multiple time horizons)

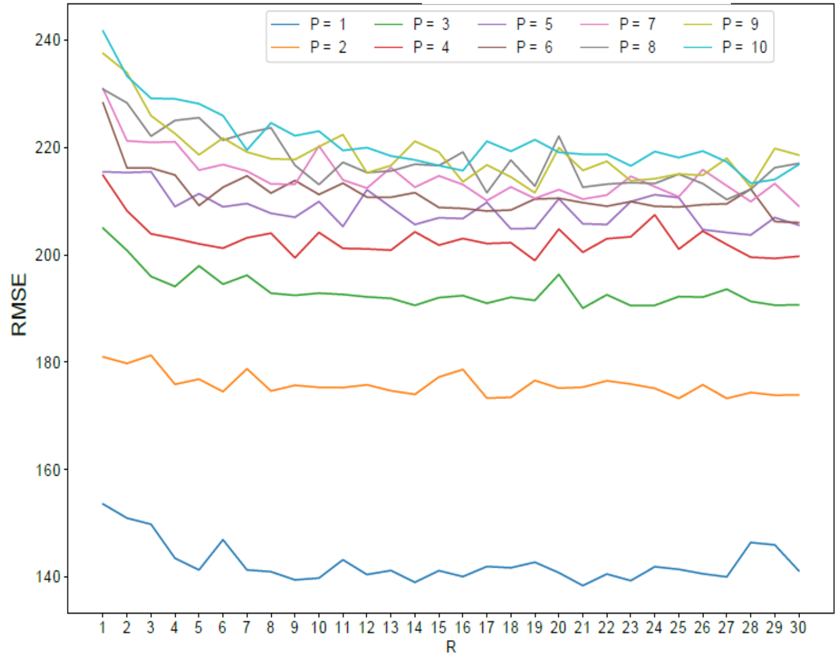
(a) CNN



(b) LSTM



(c) CNN_LSTM



(c)

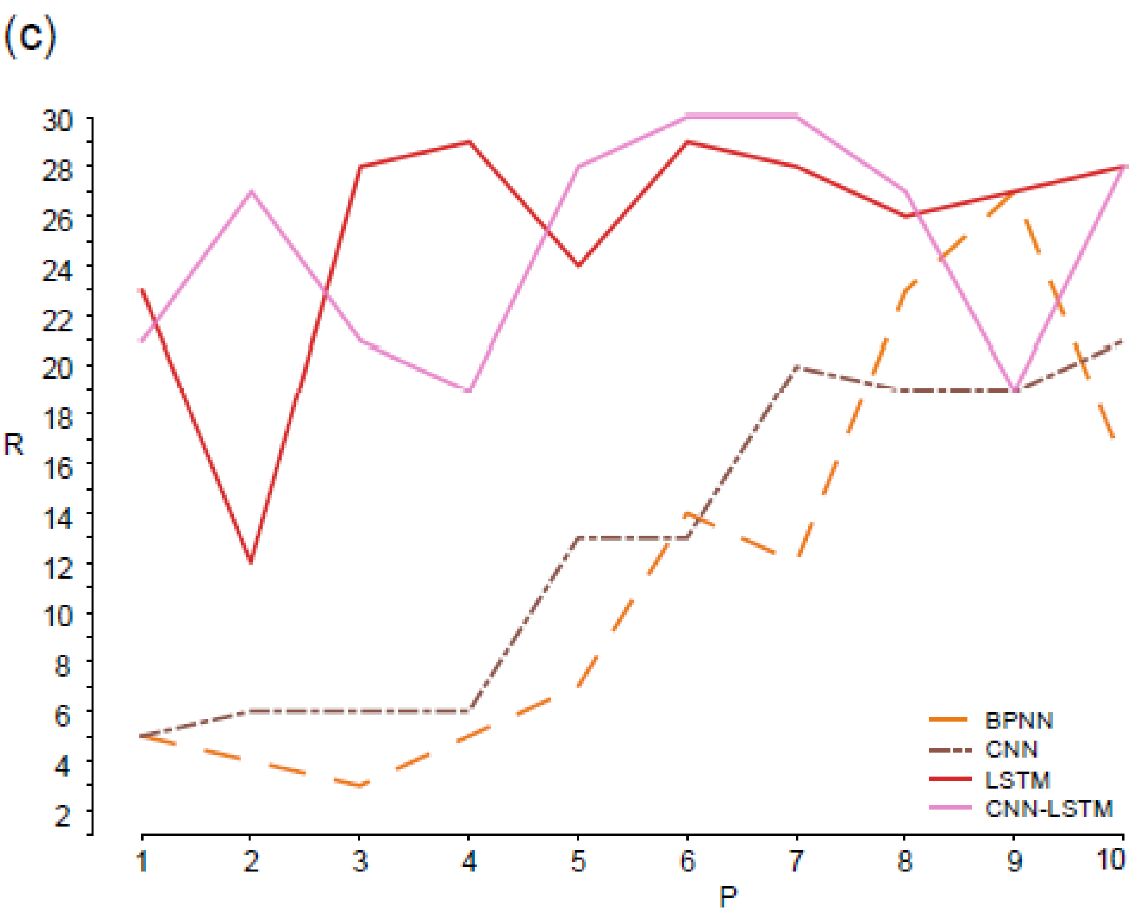
Results – Best R for each P

LSTM and the hybrid CNN-LSTM make use of larger past time horizons even when making short-term predictions.

P=3:
The best LSTM uses 69min in the past (R = 23), while CNN only uses 18min in the past (R = 6)!

This may prove problematic when long historical data is not available, in which case CNN and BPNN might provide better results.

Best R for each P



Conclusions

- DL provides overall good prediction accuracy for a large number of traffic flow counting stations

- Model usage should be based on past learning horizon and future prediction horizon (adapted to its functionality):
  - LSTM and its variants learn long-term trends and require longer histories, while
  - CNN learns spatial correlations from short histories.

- LSTM has the best predictive performance, despite having competed against a hybrid model combining CNN and LSTM

- The more complex deep learning models do not improve the prediction accuracy for our motorway flow prediction study.

Future work

- Designing traffic flow-based detection methods for stochastic events which can massively disrupt the traffic flow along motorways
- Early anomaly detection
- Graph-based prediction approaches

UTS

QUESTIONS?

Adriana-simona.Mihaita@uts.edu.au

2 PhD Opportunities:

- Modelling traffic disruptions using machine learning and simulation modelling
- Distributed traffic control for connected and autonomous vehicles in mixed traffic environments

FM LAB

www.fmlab.org