Machine learning approaches for transport predictions problems: are we making progress in the right direction?

Filipe Rodrigues, Associate Professor

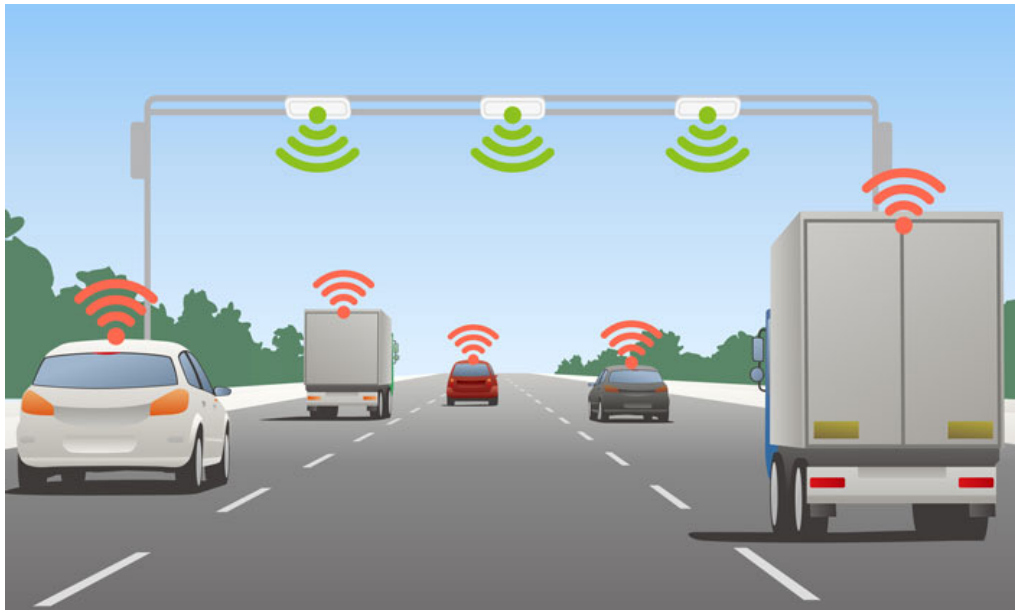
Technical University of Denmark DTU

MLSM
Machine Learning for Smart Mobility group
http://mlsm.man.dtu.dk

Trafikdage 2022

Emerging technologies and new sensors

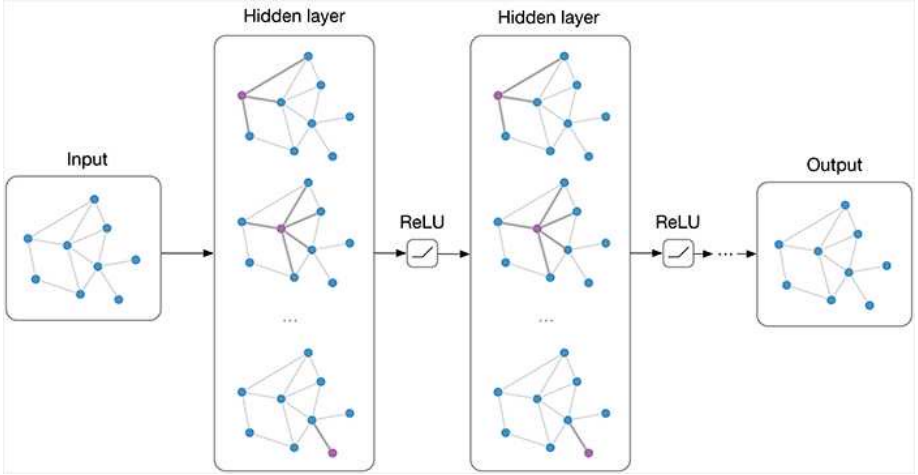
- Inductive loop detectors, probe vehicles, smart-card technologies, ride-sharing apps, etc.



- Provide unique opportunities for machine learning in transportation

Spatio-temporal forecasting in transportation

- Applications:
  - Traffic speeds, flows and volumes
  - Travel demand (taxi, bike-sharing, PT, micro-mobility, etc.)
  - Railway delays

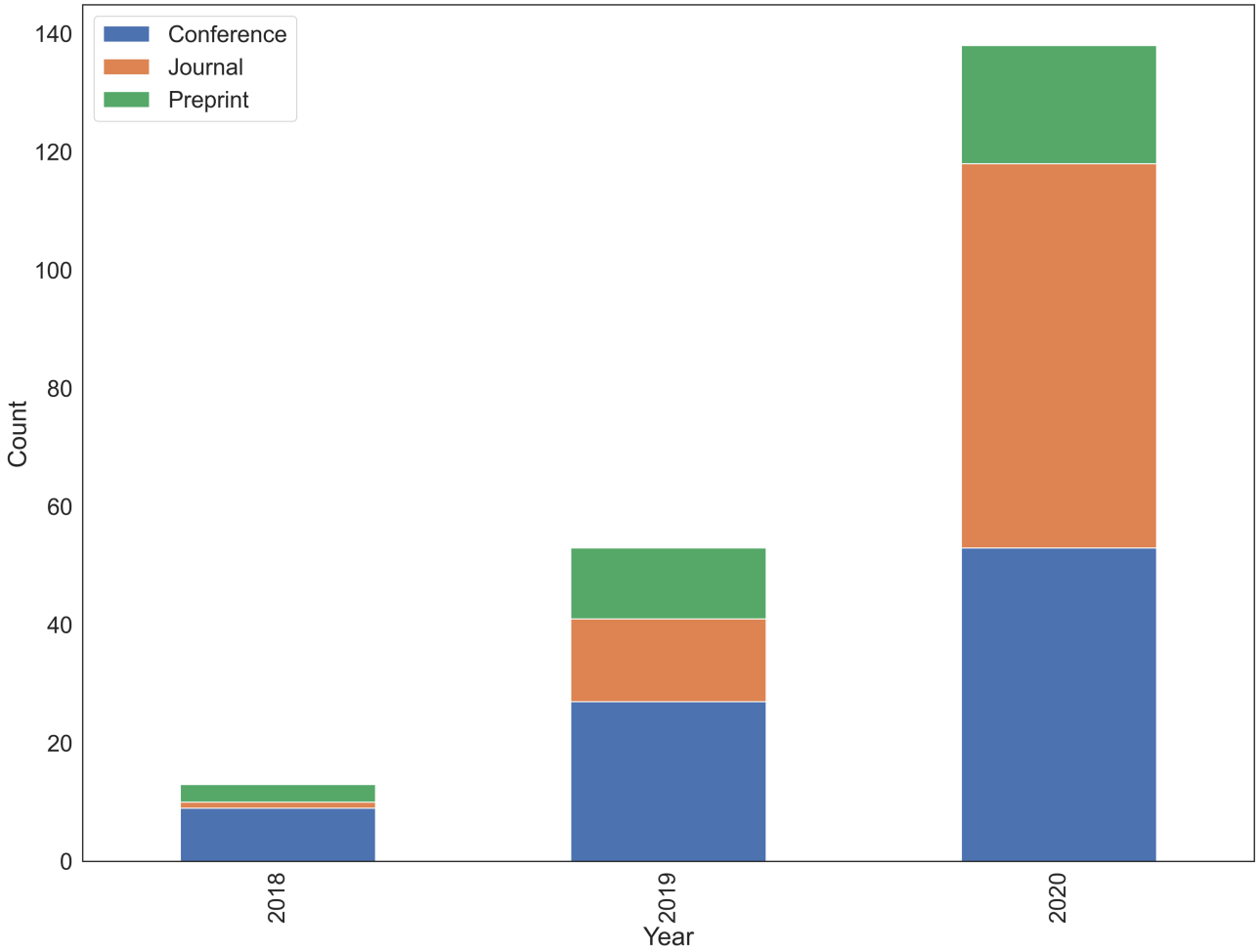


- Strong correlations in both time and space
- Huge focus on modelling those correlations
- Currently dominated by Graph Neural Networks

Hope or hype?

- Tremendous volume of new contributions in top-tier journals and conferences

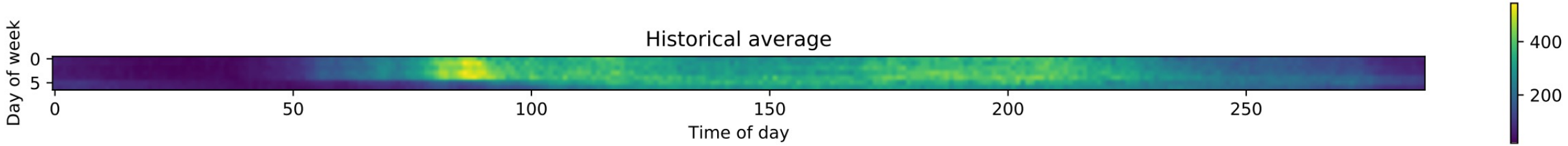
- Check out: https://github.com/jwwthu/GNN4Traffic

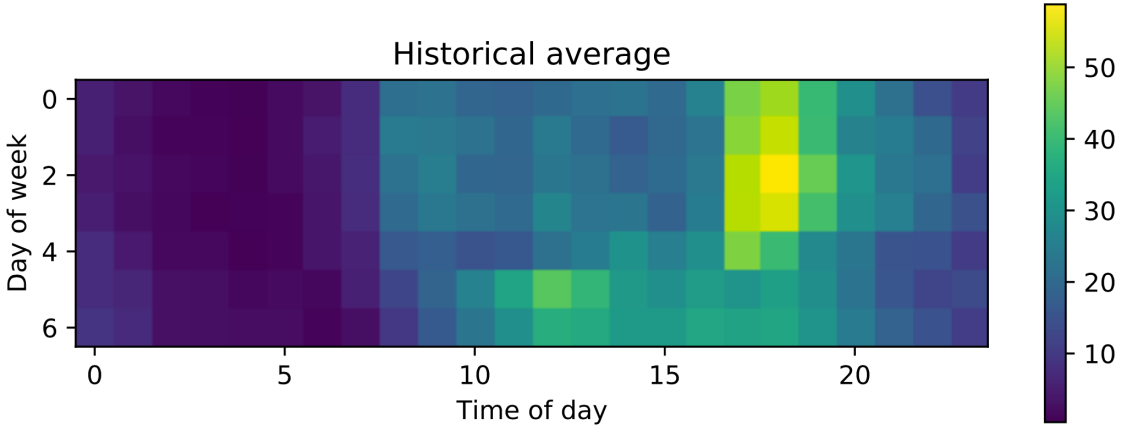


DTU MLSM
Machine Learning for Smart Mobility

Stable and recurrent mobility patterns

- Human mobility patterns are fairly stable and recurrent (with the exception of accidents, road works and special events)

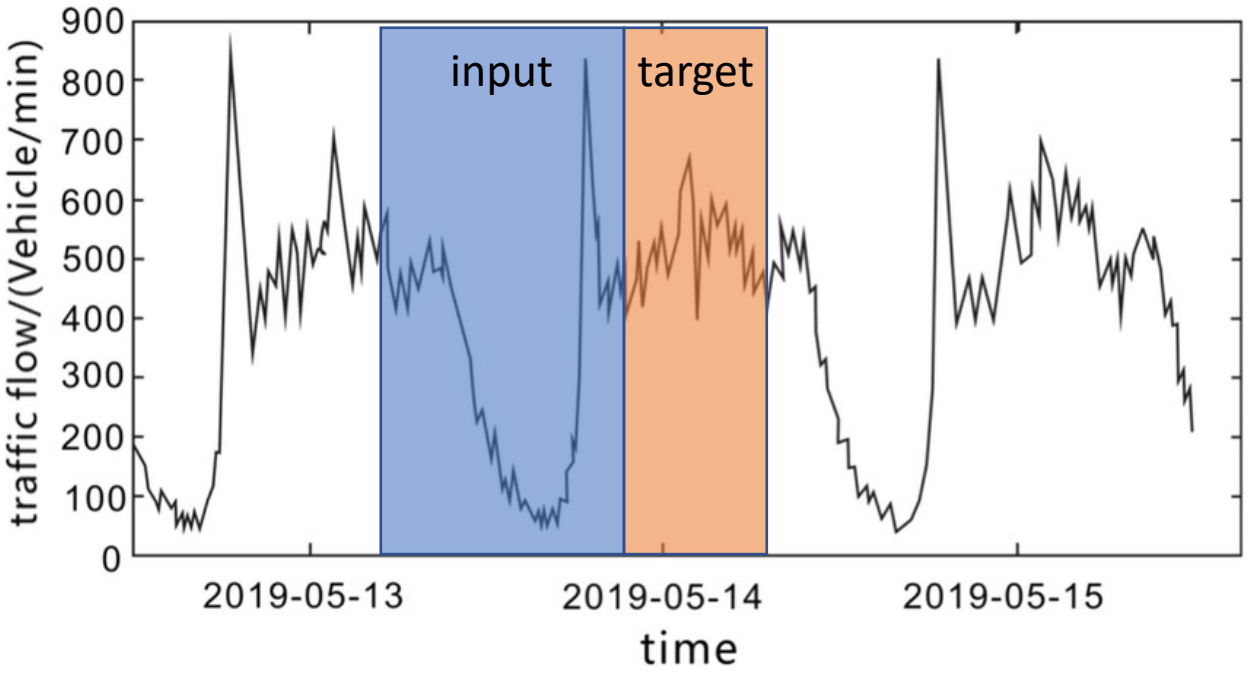




DTU MLSM Machine Learning for Smart Mobility

The typical experimental setup

- Work directly with highly non-stationary time series (ignore seasonality!)
- Use last m observations $(y_{t-m}, \dots, y_t)$ as input to forecast the next k time-steps $(y_{t+1}, \dots, y_{t+k})$



Deep-diving on the literature

- Downloaded 9 datasets and respective source code provided by 8 recent publications in top-tier conferences and journals

Name	Type	Timespan	Time granularity	Train/val/test split	Source
PeMSD7(M) - California	traffic speeds	01/04/2016 - 30/06/2016	5 minutes	34/5/5 days	Yu et al., 2018 [1]
Urban1 - South Korea	traffic speeds	01/04/2018 - 30/04/2018	5 minutes	70/10/20 %	Lee and Rhee, 2022 [2]
NYC Citi Bike - New York	pickups and dropoffs	01/04/2016 - 01/04/2016	30 minutes	63/14/14 days	Ye et al., 2021 [8]
PeMSD4 - California	traffic volumes	01/01/2018 - 28/02/2018	5 minutes	60/20/20 %	Choi et al., 2022 [6]
SZ-taxi - Shenzhen	traffic speeds	01/01/2015 - 31/01/2015	15 minutes	80/-/20 %	Zhao et al., 2021 [3]
METR-LA - Los Angeles	traffic speeds	01/03/2012 - 30/06/2012	5 minutes	70/10/20 %	Li et al., 2018 [4]
PEMS-BAY - California	traffic speeds	01/01/2017 - 31/05/2017	5 minutes	70/10/20 %	Li et al., 2018 [4]
NYC Citi Bike - New York	in- and out-flows	01/07/2017 - 30/09/2017	1 hour	80/10/10 %	Xia et al., 2021 [9]
Seattle loop data - Seattle	traffic speeds	01/11/2015 - 31/12/2015	5 minutes	56/-/5 days	Yang et al., 2021 [5]

Deep-diving on the literature

- [1] B. Yu, H. Yin, and Z. Zhu, “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting,” in *27th* ***International Joint Conference on Artificial Intelligence (IJCAI-18)***, 2018.
- [2] K. Lee and W. Rhee, “Ddp-gcn: Multi-graph convolutional network for spatiotemporal traffic forecasting,” ***Transportation Research Part C: Emerging Technologies***, 2022.
- [3] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, “T-gcn: A temporal graph convolutional network for traffic prediction,” ***IEEE Transactions on Intelligent Transportation Systems***, 2019.
- [4] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” in ***International Conference on Learning Representations (ICLR-18)***, 2018.
- [5] J.-M. Yang, Z.-R. Peng, and L. Lin, “Real-time spatio temporal prediction and imputation of traffic status based on lstm and graph laplacian regularized matrix factorization,” ***Transportation Research Part C: Emerging Technologies***, 2021.
- [6] J. Choi, H. Choi, J. Hwang, and N. Park, “Graph neural controlled differential equations for traffic forecasting,” in *Thirty-Sixth* ***AAAI Conference on Artificial Intelligence (AAAI-22)***, 2022.
- [8] J. Ye, L. Sun, B. Du, Y. Fu, and H. Xiong, “Coupled layer-wise graph convolution for transportation demand prediction,” in *Thirty-Fifth* ***AAAI Conference on Artificial Intelligence (AAAI-21)***, 2021.
- [9] T. Xia, J. Lin, Y. Li, J. Feng, P. Hui, F. Sun, D. Guo, and D. Jin, “3dgcn: 3-dimensional dynamic graph convolutional network for citywide crowd flow prediction,” ***ACM Transactions on Knowledge Discovery from Data (TKDD)***, 2021.

Dealing with non-stationarity of time-series

- A very simple approach:
  - Compute average weekly pattern based on trainset (careful with holidays!)
  - Subtract the weekly pattern from the time-series and fit a model to forecast the residuals
  - At test time, add back weekly pattern to the predictions

- Which model to use? Considered 2 simple baselines:
  - Empty model (always forecasts zero) – i.e., the forecast is the weekly average according to the historical data (referred to as “HA”)
  - Simple linear regression on the *m* previous observation/lags (“HA-LR”)

PEMSD7(M) – Traffic speeds in California

TABLE II
RESULTS FOR PEMSD7(M) - TRAFFIC SPEEDS IN CALIFORNIA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [1].

Model	MAE 15/ 30/ 45 min	MAPE 15/ 30/ 45 min	RMSE 15/ 30/ 45 min
LSVR*	2.50/ 3.63/ 4.54	5.81/ 8.88/ 11.50	4.55/ 6.67/ 8.28
FNN*	2.74/ 4.02/ 5.04	6.38/ 9.72/ 12.38	4.75/ 6.98/ 8.58
FC-LSTM*	3.57/ 3.94/ 4.16	8.60/ 9.55/ 10.10	6.20/ 7.03/ 7.51
GCGRU*	2.37/ 3.31/ 4.01	5.54/ 8.06/ 9.99	4.21/ 5.96/ 7.13
STGCN(Cheb)	**2.25**/ **3.03**/ 3.57	5.26/ **7.33**/ 8.69	**4.04**/ 5.70/ 6.77
STGCN(1st)	2.26/ 3.09/ 3.79	**5.24**/ 7.39/ 9.12	4.07/ 5.77/ 7.03
HA	3.90	10.14	7.09
HA+LR	2.48/ 3.13/ **3.45**	5.81/ 7.65/ **8.57**	4.22/ **5.50**/ **6.10**

[1] B. Yu, H. Yin, and Z. Zhu, “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting,” in *27th* ***International Joint Conference on Artificial Intelligence (IJCAI-18)***, 2018.

URBAN1 – Traffic speeds in South Korea

TABLE III
RESULTS FOR URBAN1 - TRAFFIC SPEEDS IN SOUTH KOREA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [2]. DCRNN AND STGCN DENOTE THE APPROACHES PROPOSED IN [4] AND [1], RESPECTIVELY.

Model	MAE 30/ 45/ 60 min	MAPE 30/ 45/ 60 min	RMSE 30/ 45/ 60 min
VAR*	5.06/ 4.99/ 4.97	23.10/ 22.82/ 22.73	7.04/ 6.92/ 6.88
LSVR*	3.82/ 3.89/ 3.93	15.35/ 17.99/ 17.39	5.64/ 5.74/ 5.84
ARIMA*	3.49/ 3.79/ 4.04	15.40/ 16.85/ 18.09	5.28/ 5.65/ 5.94
FC-LSTM*	3.91/ 3.92/ 3.92	17.29/ 17.32/ 17.31	6.38/ 6.39/ 6.39
DCRNN*	3.17/ 3.46/ 3.73	13.52/ 14.83/ 15.95	4.94/ 5.30/ 5.61
STGCN*	3.07/ 3.42/ 3.80	14.38/ 16.72/ 19.37	4.57/ 4.83/ 5.04
DDP-GCN	**3.00/ 3.00/ 2.99**	13.57/ **13.56/ 13.51**	**4.45/ 4.45/ 4.47**
HA	3.18	14.19	4.79
HA+LR	3.04/ 3.10/ 3.13	**13.39**/ 13.73/ 13.87	4.60/ 4.67/ 4.71

[2] K. Lee and W. Rhee, “Ddp-gcn: Multi-graph convolutional network for spatiotemporal traffic forecasting,” ***Transportation Research Part C: Emerging Technologies***, 2022.

NYC CITI BIKE – Pickups and drop-offs in NYC

TABLE IV
RESULTS FOR NYC CITI BIKE - PICKUPS AND DROPOFFS IN NYC. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [8].

Model	MAE	RMSE
XGBoost*	2.469	4.050
FC-LSTM*	2.303	3.814
DCRNN*	1.895	3.209
STGCN*	2.761	3.604
STG2Seq*	2.498	3.984
Graph WaveNet*	1.991	3.294
CCRNN	1.740	2.838
HA	**1.726**	2.871
HA+LR	1.738	**2.758**

[8] J. Ye, L. Sun, B. Du, Y. Fu, and H. Xiong, "Coupled layer-wise graph convolution for transportation demand prediction," in *Thirty-Fifth* ***AAAI Conference on Artificial Intelligence (AAAI-21)***, 2021.

PEMSD4 – Traffic volumes in California

TABLE V
RESULTS FOR PEMSD4 - TRAFFIC VOLUMES IN CALIFORNIA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [6].

Model	MAE	MAPE	RMSE
ARIMA*	33.73	24.18	48.80
VAR*	24.54	17.24	38.61
FC-LSTM*	26.77	18.23	40.65
TCN*	23.22	15.59	37.26
GRU-ED*	23.68	16.44	39.27
DSANet*	22.79	16.03	35.77
STGCN*	21.16	13.83	34.89
DCRNN*	21.22	14.17	33.44
GraphWaveNet*	24.89	17.29	39.66
ASTGCN(r)*	22.93	16.56	35.22
MSTGCN*	23.96	14.33	37.21
STG2Seq*	25.20	18.77	38.48
LSGCN*	21.53	13.18	33.86
STSGCN*	21.19	13.90	33.65
AGCRN*	19.83	12.97	32.26
STFGNN*	20.48	16.77	32.51
STGODE*	20.84	13.77	32.82
Z-GCNETs*	19.50	12.78	31.61
STG-NCDE	**19.21**	**12.76**	**31.09**
HA	26.26	17.07	42.87
HA+LR	20.03	13.39	32.73

[6] J. Choi, H. Choi, J. Hwang, and N. Park, "Graph neural controlled differential equations for traffic forecasting," in *Thirty-Sixth* ***AAAI Conference on Artificial Intelligence (AAAI-22)***, 2022.

SZ-Taxi – Traffic speeds in Shenzhen

TABLE VI
RESULTS FOR SZ-TAXI - TRAFFIC SPEEDS IN SHENZHEN FROM TAXI TRAJECTORIES. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [3].

Model	MAE 15/ 30/ 45/ 60 min	RMSE 15/ 30/ 45/ 60 min
T-GCN	4.517/ 4.572/ 4.621/ 4.671	5.997/ 6.034/ 6.064/ 6.100
HA	4.630	6.463
HA+LR	**3.464/ 3.507/ 3.534/ 3.554**	**4.998/ 5.057/ 5.091/ 5.115**

[3] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," ***IEEE Transactions on Intelligent Transportation Systems***, 2019.

METR-LA – Traffic speeds in Los Angeles

TABLE VII
RESULTS FOR METR-LA - TRAFFIC SPEEDS IN LA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [4].

Model	MAE 15/ 30/ 60 min	MAPE 15/ 30/ 60 min	RMSE 15/ 30/ 60 min
ARIMA*	3.99/ 5.15/ 6.90	9.6/ 12.7/ 17.4	8.21/ 10.45/ 13.23
VAR*	4.42/ 5.41/ 6.52	10.2/ 12.7/ 15.8	7.89/ 9.13/ 10.11
SVR*	3.99/ 5.05/ 6.72	9.3/ 12.1/ 16.7	8.45/ 10.87/ 13.76
FNN*	3.99/ 4.23/ 4.49	9.9/ 12.9/ 14.0	7.94/ 8.17/ 8.69
FC-LSTM*	3.44/ 3.77/ 4.37	9.6/ 10.9/ 13.2	6.30/ 7.23/ 8.69
DCRNN	**2.77/ 3.15/ 3.60**	**7.3/ 8.8/ 10.5**	**5.38/ 6.45**/ 7.59
HA	4.19	13.0	7.84
HA+LR	3.28/ 3.68/ 4.02	8.8/ 10.4/ 11.9	5.71/ 6.60/ **7.32**

[4] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” in ***International Conference on Learning Representations (ICLR-18)***, 2018.

PEMS-BAY – Traffic speeds in California

TABLE VIII
RESULTS FOR PEMS-BAY - TRAFFIC SPEEDS IN CALIFORNIA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [4].

Model	MAE 15/ 30/ 60 min	MAPE 15/ 30/ 60 min	RMSE 15/ 30/ 60 min
ARIMA*	1.62/ 2.33/ 3.38	3.5/ 5.4/ 8.3	3.30/ 4.76/ 6.50
VAR*	1.74/ 2.32/ 2.93	3.6/ 5.0/ 6.5	3.16/ 4.25/ 5.44
SVR*	1.85/ 2.48/ 3.28	3.8/ 5.5/ 8.0	3.59/ 5.18/ 7.08
FNN*	2.20/ 2.30/ 2.46	5.2/ 5.4/ 5.9	4.42/ 4.63/ 4.98
FC-LSTM*	2.05/ 2.20/ 2.37	4.8/ 5.2/ 5.7	4.19/ 4.55/ 4.96
DCRNN	**1.38/ 1.74/ 2.07**	**2.9/ 3.9/ 4.9**	2.95/ 3.97/ 4.74
HA	2.58	6.1	5.04
HA+LR	1.54/ 1.91/ 2.22	3.2/ 4.3/ 5.1	**2.93/ 3.83/ 4.45**

[4] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” in ***International Conference on Learning Representations (ICLR-18)***, 2018.

NYC Bike In- and Out-flows

TABLE IX
RESULTS FOR NYC BIKE IN- AND OUT-FLOWS. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [9].

Model	MAE 1h/ 2h/ 3h	RMSE 1h/ 2h/ 3h
ARIMA*	10.41/ 11.84/ 13.00	19.14/ 21.76/ 23.90
STGCN*	6.49/ 7.06/ 7.94	11.73/ 12.93/ 15.37
DCRNN*	5.88/ 6.19/ 7.72	9.85/ 10.39/ 12.37
STGNN*	5.79/ 6.00/ 7.56	9.80/ 9.98/ 11.91
MVGCN*	5.65/ 7.72/ 8.00	9.64/ 13.53/ 13.93
3DGCN	**4.81**/ 5.61/ 6.99	**7.76**/ **9.49**/ 11.74
HA	5.97	11.04
HA+LR	5.10/ **5.45**/ **5.56**	8.72/ 9.69/ **10.04**

[9] T. Xia, J. Lin, Y. Li, J. Feng, P. Hui, F. Sun, D. Guo, and D. Jin, "3dgcn: 3-dimensional dynamic graph convolutional network for citywide crowd flow prediction," ***ACM Transactions on Knowledge Discovery from Data (TKDD)***, 2021.

Seattle traffic speed data

TABLE X
RESULTS FOR SEATTLE TRAFFIC SPEED DATA. THE * INDICATES RESULTS TAKEN DIRECTLY FROM [5].

	MAPE	RMSE
BTMF*	7.70	4.59
TRMF-GRMF*	8.20	4.83
TRMF-ALS*	8.36	4.96
Linear-LSTM-ReMF*	8.01	4.59
BiLSTM-GL-ReMF	7.83	4.50
LSTM-ReMF	7.64	4.42
LSTM-GL-ReMF	7.64	4.43
HA	12.5	9.82
HA+LR	**5.5**	**3.95**

[5] J.-M. Yang, Z.-R. Peng, and L. Lin, “Real-time spatio temporal prediction and imputation of traffic status based on lstm and graph laplacian regularized matrix factorization,” ***Transportation Research Part C: Emerging Technologies***, 2021.

Discussion: stationarity and recurrent weekly patterns

- Must account for the recurrent weekly patterns
- Working with the original non-stationary time-series artificially (and unnecessarily!) increases the complexity of the forecasting problem
- Working with residuals to the ”historical average” is a good step towards stationarity
  - We have know this for decades, but somehow we decided to ignore it! Why??
- One may also wish to consider detrending approaches

Discussion: baselines

- We need strong baselines that ideally are also simple to implement
- The “HA-LR” baseline is trivial to implement and has no hyper-parameters to tune
- When comparing with other state-of-the-art approaches, spend enough time on carefully tuning them so that the comparison is fair
  - Consider using well-known benchmark datasets – avoid hyper-parameter tuning altogether

Discussion: benchmarks

- Inexistence of standard reference benchmarks aggravates the problem
- Most papers tend to use their own datasets with their own experimental setup (train/val/test split, forecasting horizons, evaluation metrics, etc.)
- Majority of papers don't make their code and datasets publicly available
- Must consider multiple datasets ("no free lunch" theorem)
- Created a repository with 9 benchmark datasets: https://github.com/fmpr/mobility-baselines

Discussion: sharing code and datasets

- Good practices such as the sharing of code and datasets has massively accelerated the pace of progress in other fields such as computer vision and natural language processing
- The transportation community has yet to widely embrace these good practices
- Facilitates the comparison of approaches and speeds up scientific progress

Discussion: relevancy

- Are these improvements relevant in practice?
- Is the field making progress, or are we stuck in a loop of small incremental contributions and repeating the same mistakes?
- Does an improvement of 0.5 in MAE when forecasting traffic speeds has a relevant impact on traffic management that justifies a complex model that requires several hours to train and multiple days to tune its hyper-parameters?
- Is the CO2 fingerprint that we are causing due to all this computation worth it?

Discussion: evaluation metrics

- We are overly focused on error metrics like MAE and RMSE
- We should try to progressively move away from these error metrics towards other KPIs that are better proxies of societal impact
  - E.g.: number of hours spent in traffic, CO2 emissions and quality of service
- When the predictions of a model are used as inputs for another downstream task, the empirical evaluation should include the improvements on that task
- Emphasis of evaluation should be on abnormal conditions, such as the ones caused by incidents, special events, extreme weather, etc.

Discussion: spatial correlations

- Spatial correlations can have a significant impact on forecasting error, especially when shorter forecasting horizons
- Particularly relevant during abnormal conditions – e.g. for anticipating propagation of delays and other “domino-type” effects
- As longer forecasting horizons are considered, the impact of modeling spatial correlations on forecasting error gradually decreases

Conclusion

- Showed that a naive baseline method can achieve comparable results to many state-of-the-art deep learning approaches for spatio-temporal forecasting in transportation
- Contrasted the importance of stationarity and recurrent patterns in the data with the importance of spatial correlations
- Highlighted several issues with current “modus operandi”
- Discussed best practices and the direction that the field is taking
- Read more: https://arxiv.org/abs/2203.02954
- Check out: https://github.com/fmpr/mobility-baselines