Supplement of Atmos. Chem. Phys., 22, 1939–1950, 2022
https://doi.org/10.5194/acp-22-1939-2022-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

High-resolution mapping of regional traffic emissions using land-use machine learning models

Xiaomeng Wu et al.

Correspondence to: Ye Wu (ywu@tsinghua.edu.cn)

Supplementary Figures

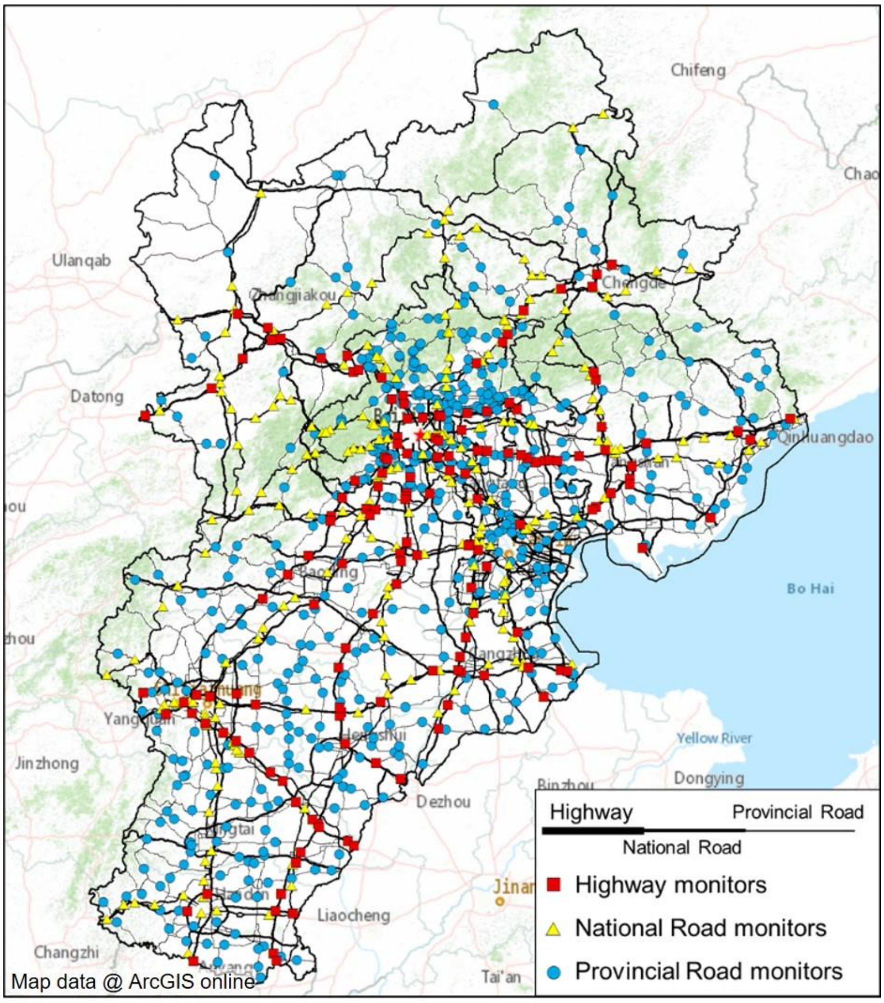

Figure S1. Map of traffic monitoring sites with traffic mix data available.

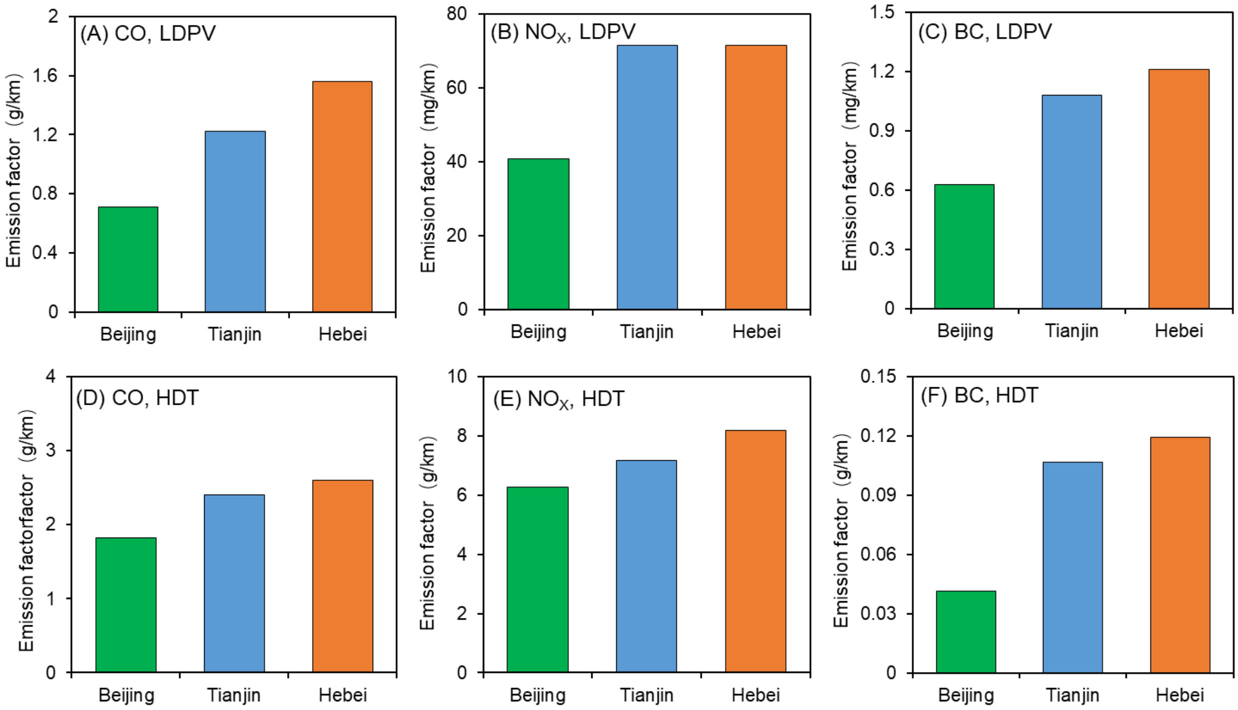

Figure S2. Fleet-average emission factors for LDPVs and HDTs estimated by the updated EMBEV model.

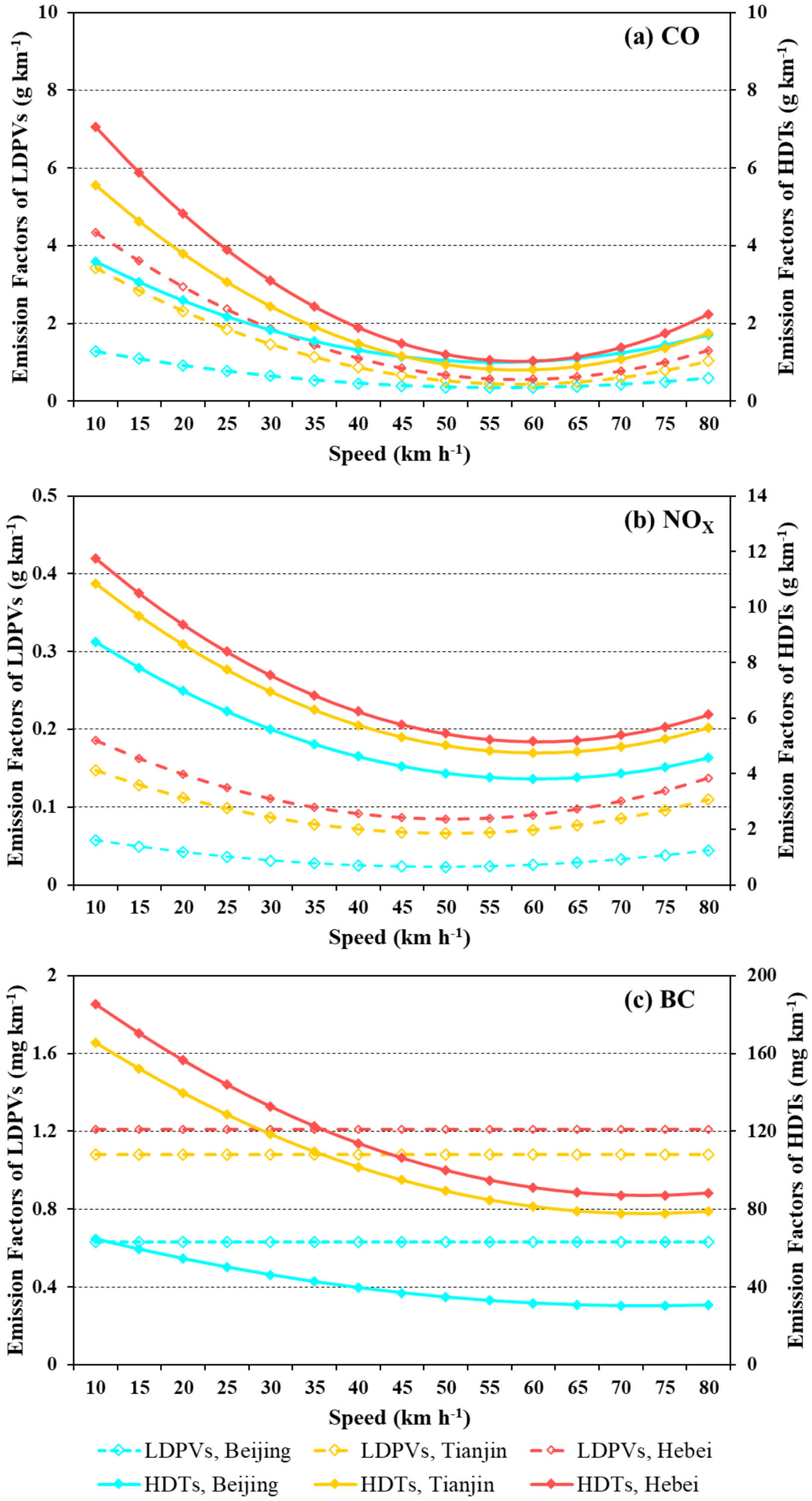

Figure S3. Speed-dependent fleet-average emission factors for LDPVs and HDTs estimated by the EMBEV model.

Note: Speed correction is not applicable to BC emissions from LDPVs due to the lack of testing data.

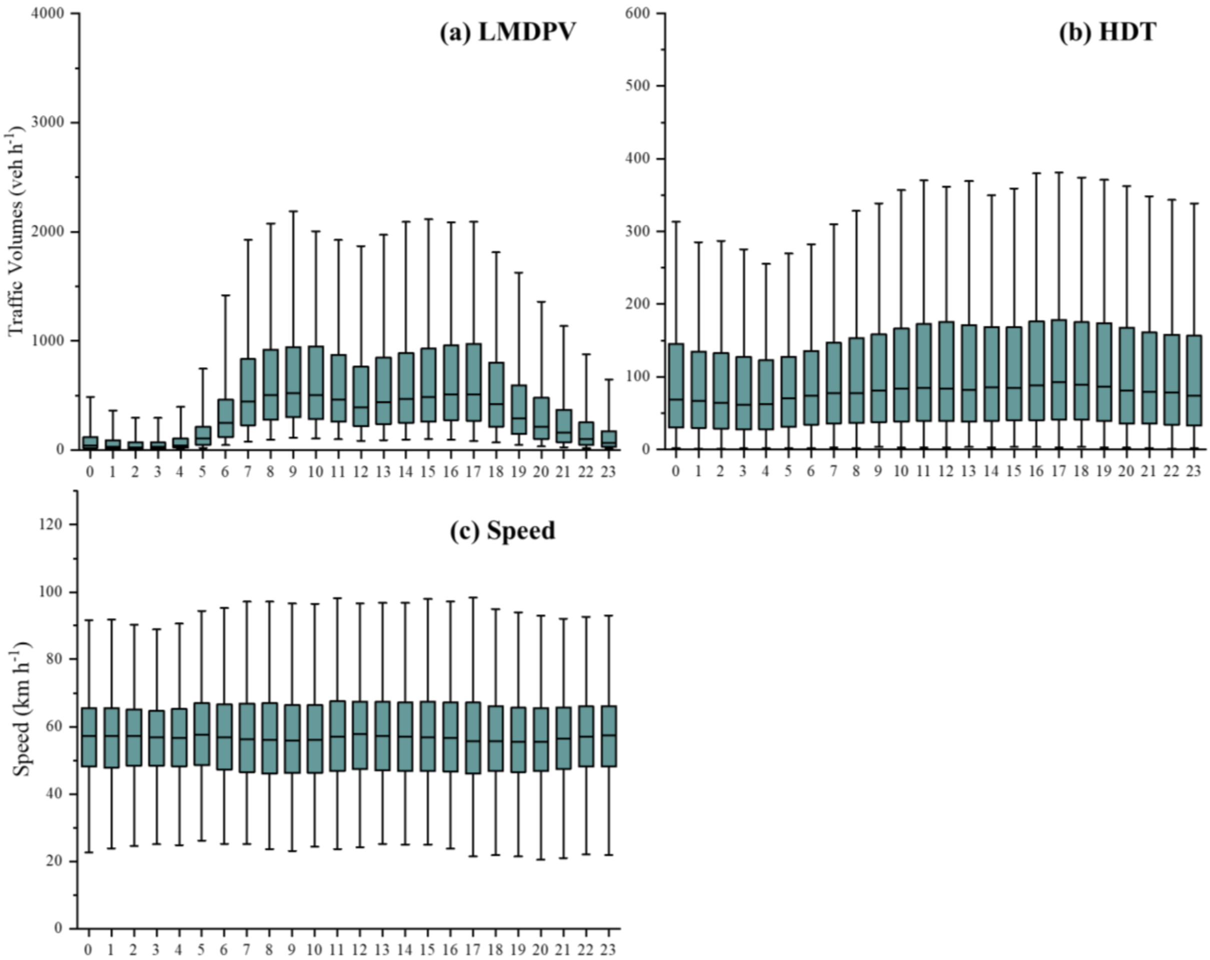

Figure S4. Box plot of traffic volumes by vehicle category and speed used to train the land use models.

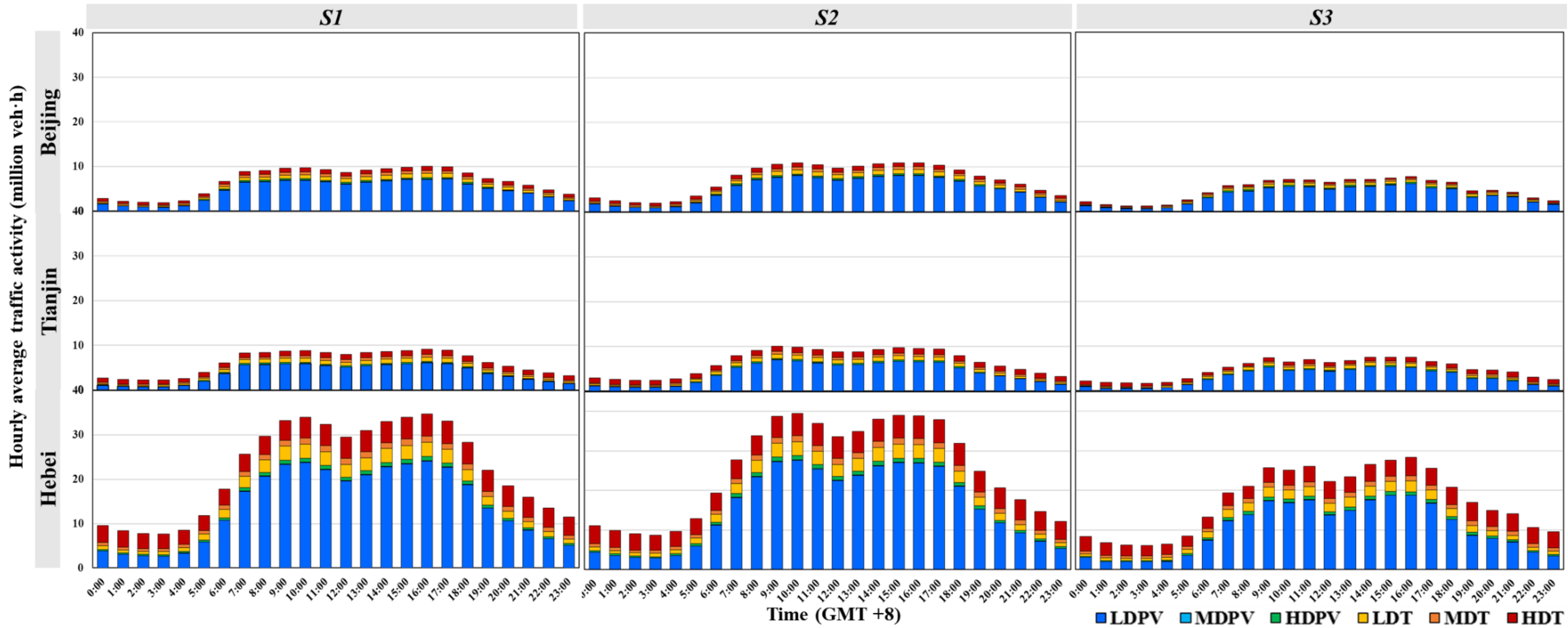

Figure S5. Average diurnal fluctuations in hourly traffic activity by vehicle category of the BTH region during various traffic scenarios *S1* to *S3*

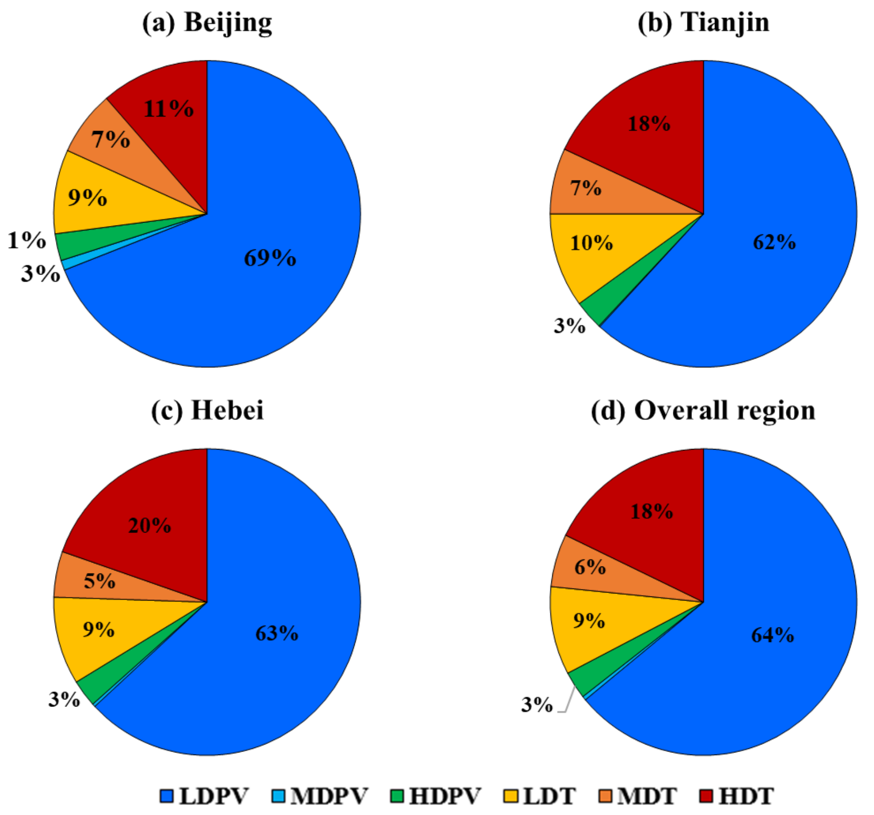

Figure S6. The proportion of each vehicle category accounting for the total traffic activity of (a) Beijing, (b) Tianjin, (c) Hebei, and (d) the overall region.

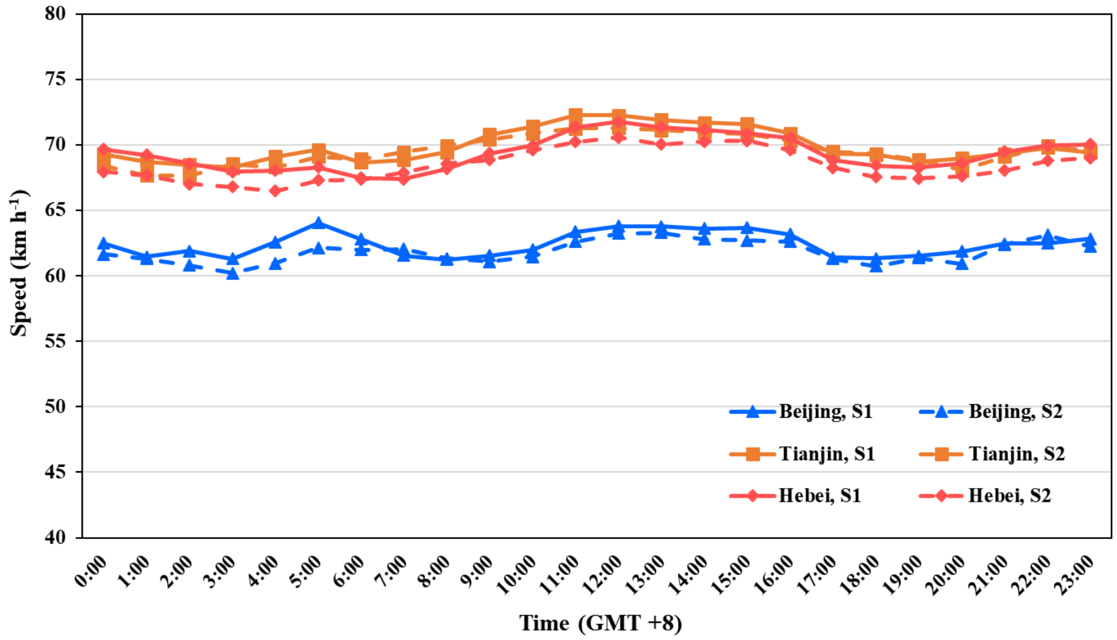

Figure S7. Average hourly speed by region under various traffic scenarios.

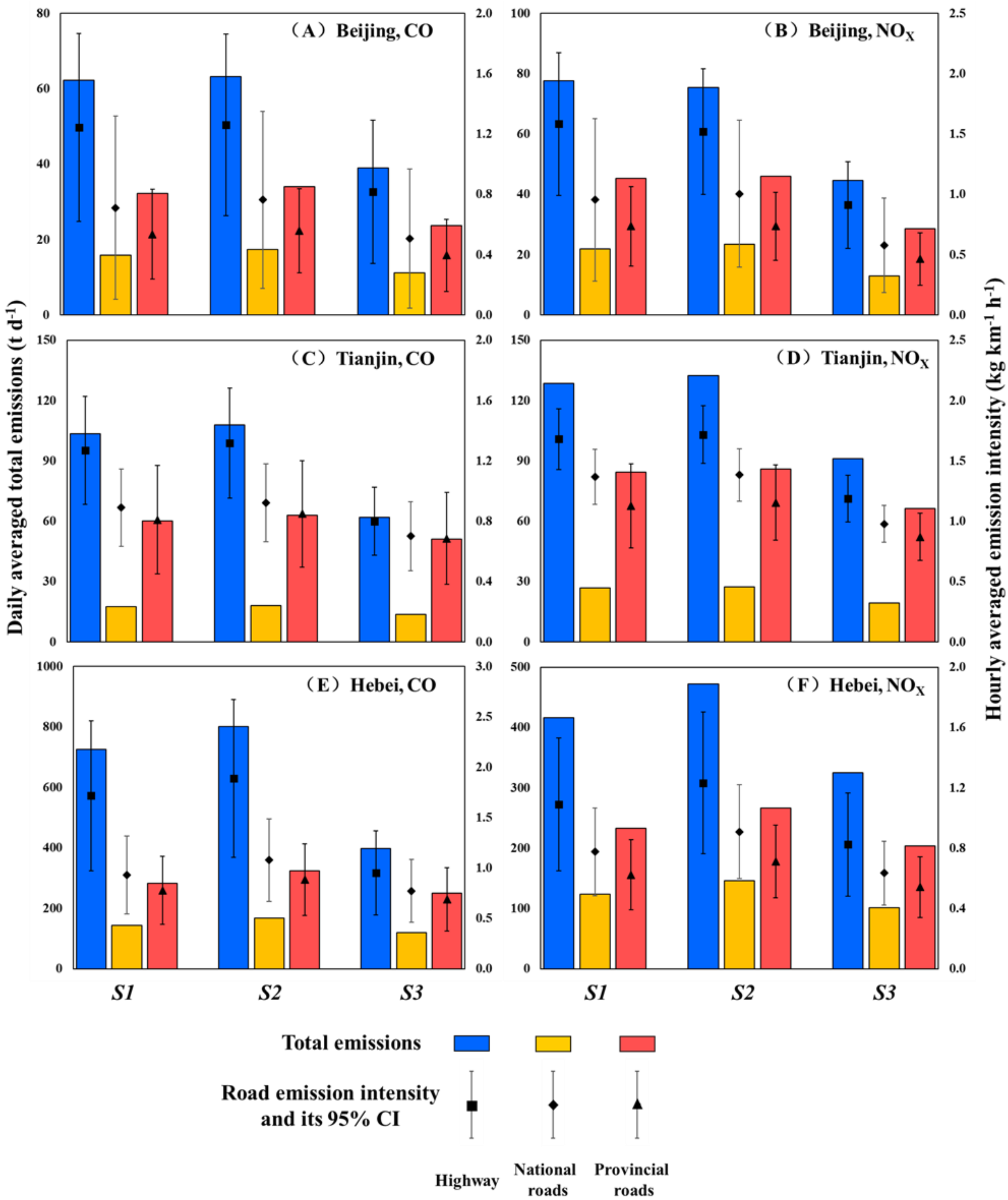

Figure S8. Estimated total emissions and emission intensity of CO and NO_X by region and road type under various traffic scenarios, S1 to S3.

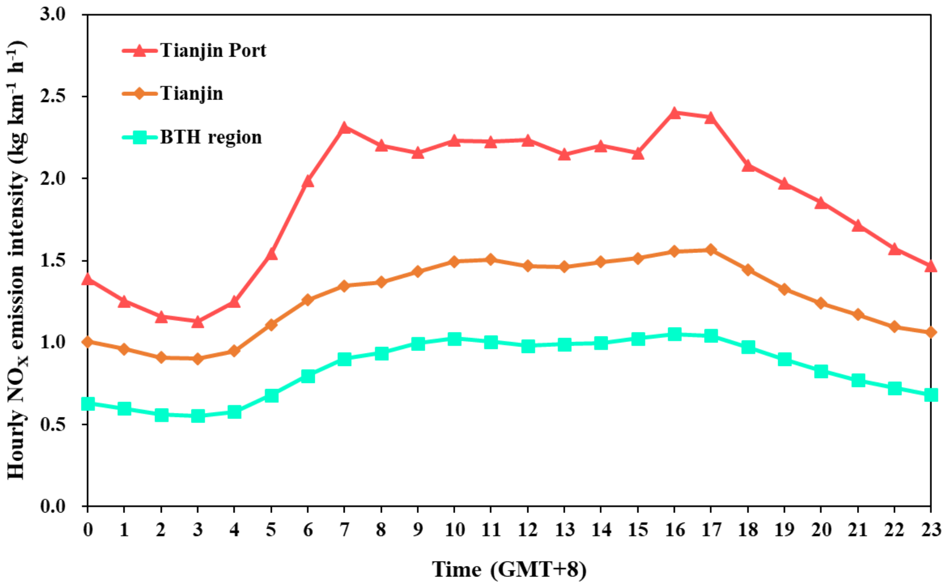

Figure S9. Hourly emission intensity of NO_X in the region of Tianjin Port, Tianjin and the BTH region.

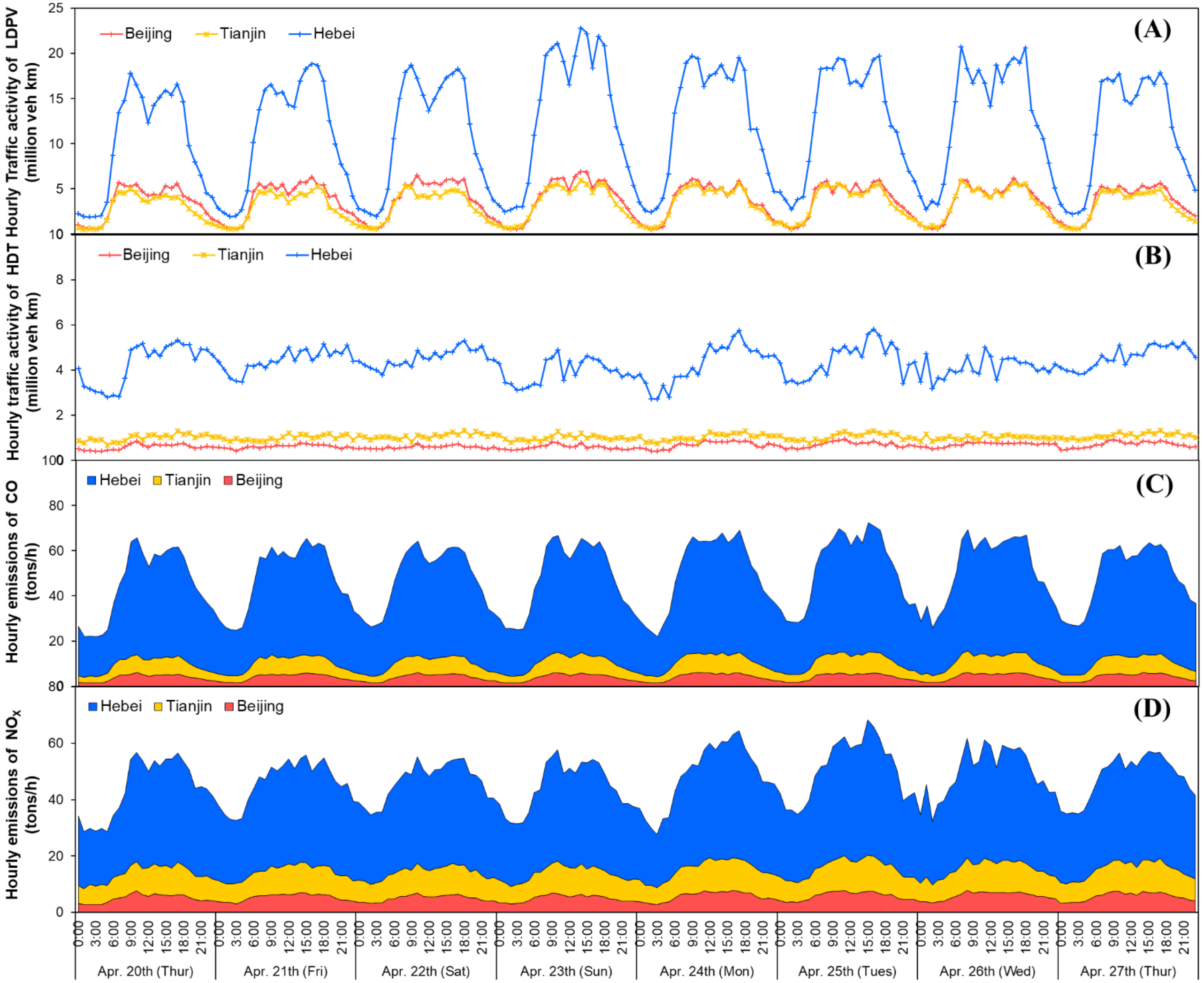

Figure S10. Hourly traffic activity of LDPV (A) and HDT (B) and vehicle emissions of CO (C) and BC (D) by region from April 20th to April 27th,2017

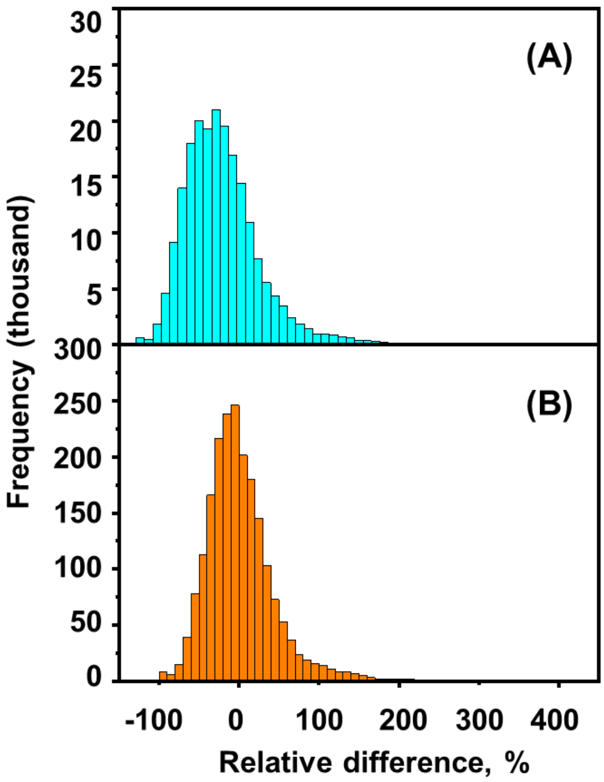

Figure S11. Distribution of relative differences of CO and NO_X of M2, compared to M1.

Supplementary Tables

Table S1. Definition of road types

Road type	Description	Designed speed
Expressways	Inter-provincial roads, often constructed by the national highway administration	two thirds of the roads above 100 km/h
National highways	Inter-provincial roads, often constructed by the national highway administration	more than half of the roads below 80 km/h
Provincial highways	Inner-provincial roads, often constructed by the provincial highway administration	more than half of the roads below 80 km/h

Table S2. Definition and abbreviation of vehicle categories

Vehicle classification	Abbreviation	Description
Light-duty passenger vehicle	LDPV	Length $\leq$ 3.5 m, PC [a] $\leq$ 9
Medium-duty passenger vehicle	MDPV	Length $<$ 6 m, 9 $<$ PC $\leq$ 20
Heavy-duty passenger vehicle	HDPV	Length $\geq$ 6 m, PC $>$ 20
Light-Duty Truck	LDT	Length $<$ 6 m, GVW [b] $\leq$4500 kg
Medium-Duty Truck	MDT	Length $\geq$ 6 m, 4500 $<$ GVW $\leq$12000 kg
Heavy-Duty Truck [c]	HDT	GVW $>$ 12000 kg [d]

Notes: [a] Passenger capacity; [b] Gross vehicle weight; [c] The HDTs are further classified into local HDTs and non-local HDTs according to the registration place; [d] Emission factor for local HDTs are weighted by HDT2 and HDT3 according to their registration number and annual VKT (Zhang et al., 2014).

Table S3. Summary of data used to train machine learning models

Category	Potential variables	Variable code
Land-use Data		
Land use (total area [km^2] / buffer area)	Urban land	urbanland
	Crop land	cropland
	Grass land	grassland
	Bare lands	bareland
Further	Transit	POI_transit
	Restaurant	POI_restaurant
	Office	POI_office
	Mall	POI_mall
	Hotel	POI_hotel
	Education	POI_education
	Bank	POI_bank
	Recreation	POI_recreation
	Touristic	POI_touristic
Distance (Euclidean [m])	Airport	D_airport
	Port	D_port
	Freight	D_freight
	CBD	D_CBD
Population density (total population / buffer area)	Population density	pop
Road Information Data		
Road density (total length [km] / buffer area)	Highways	rd00
	National roads	rd01
	Province roads	rd03
Value extracted at point	Location	Lon/Lat
	Administration	Province/City/County
	Road type	rdtype
	Number of road lane	LaneNum
	Designed road speed	DeSpeed

Note: a Buffer radii 50 m, 100 m, 200 m, 300 m, 500 m, 1000 m, 2000 m, 5000 m

Table S4. Advantages and disadvantages of machine learning models used in this study

Methods	Advantages	Disadvantages	Application on predicting traffic
LR	Easy to be applied; Easy to interpret and to be understood	Poor results on non-linear problems due to the linear assumption	To interpret the relationship between traffic variables (Alam, Farid, and Rossetti 2019); Travel time prediction (Zhang and Rice 2003; Rice and Zwet 2004)
GPR	Flexible and suitable for a wide range of problems	Low efficiency when solving high-dimensional problems	Dynamic traffic congestion (Liu, Yue, and Krishnan 2013); Short-term traffic volume forecast (Xie et al. 2010)
SVR	Works well on non-linear and high-dimensional problems; Perform well on small sample problems	Difficult to choose the optimal kernel; Need to complete feature scaling in advance; Difficult to interpret	Short-term traffic flow prediction (Li and Xu 2021)
GBDT	Ensemble learning methods; Able to improve model performance continuously based on the result and the error of last iteration	Easy overfitting; Parameters such as the number of decision trees need to be decided	Traffic volume prediction over a certain time period (Xia and Chen 2017; Yang et al. 2017); Traffic flow prediction considering spatial-temporal relationship (Yang, Zheng, and Sun 2019); Travel time prediction (Li and Bai 2016)
LURF	Ensemble learning methods; High computational capacity and high accuracy; Great performance on non-linear and high-dimensional problems; Easy to evaluate the contribution of each independent variable	Easy overfitting; Parameters such as the number of decision trees need to be decided	Road traffic congestion forecast (Liu and Wu 2017) Traffic flow prediction (Gokul L Rajeev et al. 2021)

Table S5. Simulation performance of the machine learning models in predicting traffic profiles in this study

	Traffic profiles	LURF	GBDT	SVR	GPR	LR
Pearson's R	LMDPV	0.79	0.81	0.65	0.62	0.48
	HDPV	0.61	0.54	0.51	0.46	0.3
	LDT	0.62	0.55	0.44	0.49	0.17
	MDT	0.64	0.6	0.48	0.47	0.26
	HDT	0.65	0.58	0.56	0.58	0.5
	Speed	0.75	0.74	0.7	0.71	0.55
MAPE	LMDPV	1.37	1.37	1.25	1.57	2.06
	HDPV	2.92	2.85	2.64	3.1	3.05
	LDT	1.26	1.41	1.07	1.41	1.59
	MDT	4.23	4.04	4.35	6.67	9.11
	HDT	2.08	2.24	1.81	2.45	2.71
	Speed	0.16	0.16	0.17	0.17	0.2
RMSE	LMDPV	5360	7917	10715	219458	13382
	HDPV	226	536	561	276419	739
	LDT	1205	1679	1741	30745	2382
	MDT	380	1024	1162	16504	1546
	HDT	2706	2207	2242	49899	2596
	Speed	5.68	10.77	11.26	0.36	15.56

Note: The units of RMSE for the traffic volumes and speed are veh d^{-1} and km h^{-1}, respectively.

Table S6. Top 10 important variables for the LURF predicting the traffic characteristic

LMDPV	HDPV	LDT	MDT	HDT	Speed
City# (***5.2***)	County# (***2.7***)	pop_5000m* (***5.8***)	County# (***2.6***)	County# (***1***)	rdtype# (***1.0***)
POI_office_5000m* (***6.2***)	City# (***5.5***)	pop_2000m* (***10.7***)	Province# (***2.7***)	rdtype# (***2.0***)	City# (***2.4***)
urbanland_5000m *(***7.5***)	LaneNum# (***9.3***)	Admin# (***12.8***)	City# (***3.7***)	LaneNum# (***3.7***)	County# (***3.5***)
pop_5000m* (***9.8***)	rdtype# (***12***)	pop_1000m* (***14.7***)	POI_office_5000m* (***4.5***)	City# (***5.9***)	rd00_50m# (***3.6***)
rdtype# (***10.9***)	urbanland_5000m* (***20.7***)	LaneNum# (***18.8***)	rdtype# (***8.1***)	Lat# (***7.7***)	rd00_100m# (***4.5***)
County# (***11.5***)	DeSpeed# (***25.6***)	urbanland_5000m* (***24.4***)	urbanland_5000m* (***9.1***)	DeSpeed# (***8.3***)	rd00_200m# (***7.3***)
pop_2000m* (***13***)	Province# (***25.9***)	City# (***26.9***)	urbanland_2000m* (***9.7***)	urbanladn_5000m* (***8.6***)	Province# (***9***)
LaneNum# (***13.7***)	pop_5000m* (***27.3***)	POI_mall_5000m* (***27.9***)	LaneNum# (***14.1***)	rd00_5000m# (***13.5***)	DeSpeed# (***9.5***)
POI_transit_5000m* (***14.8***)	rd00_50m# (***31.9***)	POI_office_5000m* (***28.0***)	pop_5000m* (***15***)	cropland_1000m* (***16.8***)	LaneNum# (***10.3***)
pop_1000m* (***15.9***)	Lat# (***33.4***)	POI_restaurant_5000m * (***28.6***)	POI_transit_5000m* (***15.6***)	rd00_2000m# (***20.1***)	rd00_300m# (***12.9***)

Note: The number in the bracket is the average hourly importance ranks of the variables. * variables representing the land-use information; # variables representing the road information.

Table S7. The VKT allocation weights by region and road type

		LMDPV	HDPV	LDT	MDT	HDT
Beijing	Expressway	56%	51%	53%	55%	55%
	National-level highways	15%	15%	16%	15%	15%
	Provincial-level highways	29%	34%	31%	30%	30%
Tianjin	Expressway	48%	49%	43%	48%	53%
	National-level highways	11%	12%	11%	12%	12%
	Provincial-level highways	41%	39%	45%	40%	35%
Hebei	Expressway	48%	46%	39%	46%	51%
	National-level highways	17%	18%	19%	18%	18%
	Provincial-level highways	35%	36%	42%	36%	31%

Reference

Alam, Ishteaque, Dewan Md. Farid, and Rosaldo J. F. Rossetti. 2019. "The Prediction of Traffic Flow with Regression Analysis." In, 661-71. Singapore: Springer Singapore.

Gokul L Rajeev, Revathy Nancy, Megha S, Jeena Mary John, and Nithya Elsa John. 2021. 'Traffic Flow Prediction using Random Forest and Bellman Ford for Best Route Detection', *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, NCREIS – 2021 (Volume 09 – Issue 13).

Li, Cong, and Pei Xu. 2021. 'Application on traffic flow prediction of machine learning in intelligent transportation', *Neural Computing and Applications*, 33: 613-24.

Li, X., and R. Bai. 2016. "Freight Vehicle Travel Time Prediction Using Gradient Boosting Regression Tree." In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1010-15.

Liu, Siyuan, Yisong Yue, and Ramayya Krishnan. 2013. "Adaptive collective routing using gaussian process dynamic congestion models." In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 704–12. Chicago, Illinois, USA: Association for Computing Machinery.

Liu, Y., and H. Wu. 2017. "Prediction of Road Traffic Congestion Based on Random Forest." In *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, 361-64.

Rice, J., and E. van Zwet. 2004. 'A simple and effective method for predicting travel times on freeways', *IEEE Transactions on Intelligent Transportation Systems*, 5: 200-07.

Xia, Ying, and Jungang Chen. 2017. "Traffic Flow Forecasting Method based on Gradient Boosting Decision Tree." In *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*, 413-16. Atlantis Press.

Xie, Yuanchang, Kaiguang Zhao, Ying Sun, and Dawei Chen. 2010. 'Gaussian Processes for Short-Term Traffic Volume Forecasting', *Transportation Research Record*, 2165: 69-78.

Yang, Jie, Linjiang Zheng, and Dihua Sun. 2019. "Urban Traffic Flow Prediction Using a Gradient-Boosted Method Considering Dynamic Spatio-Temporal Correlations." In *Knowledge Science, Engineering and Management*, edited by Christos Douligeris, Dimitris Karagiannis and Dimitris Apostolou, 271-83. Cham: Springer International Publishing.

Yang, Senyan, Jianping Wu, Yiman Du, Yingqi He, and Xu Chen. 2017. 'Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine', *Journal of Sensors*, 2017: 7074143.

Zhang, Xiaoyan, and John A. Rice. 2003. 'Short-term travel time prediction', *Transportation Research Part C: Emerging Technologies*, 11: 187-210.