SUPPLEMENTAL INFORMATION

DEVELOPMENT AND COMPARISON OF AIR POLLUTION EXPOSURE SURFACES DERIVED FROM ON-ROAD MOBILE MONITORING AND SHORT-TERM STATIONARY SIDEWALK MEASUREMENTS

Laura Minet[1], Rick Liu[1], Marie-France Valois[2], Junshi Xu[1], Scott Weichenthal[3], Marianne Hatzopoulou[1]*

[1]Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada

[2]Division of Clinical Epidemiology, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

[3]Epidemiology, Biostatistics & Occupational Health, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

*Corresponding author, Department of Civil Engineering, University of Toronto

35 St George Street, Toronto, ON M5S 1A4

Telephone: 416-978-0864. Email: marianne.hatzopoulou@utoronto.ca

NUMBER OF PAGES: 41

NUMBER OF TABLES: 21

NUMBER OF FIGURES: 21

The supportive information contains information about 1) the design of the campaigns, 2) the LUR predictors, 3) the correction of the monitors, 4) the data collected, 5) the choice of the models, 6) the comparison of the exposures, 7) the characteristics of various UFP and BC monitoring campaigns and LUR models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 METHODOLOGY

1.1 Campaign design

FIGURE S1 presents the fixed points and biking routes designed for the campaigns.

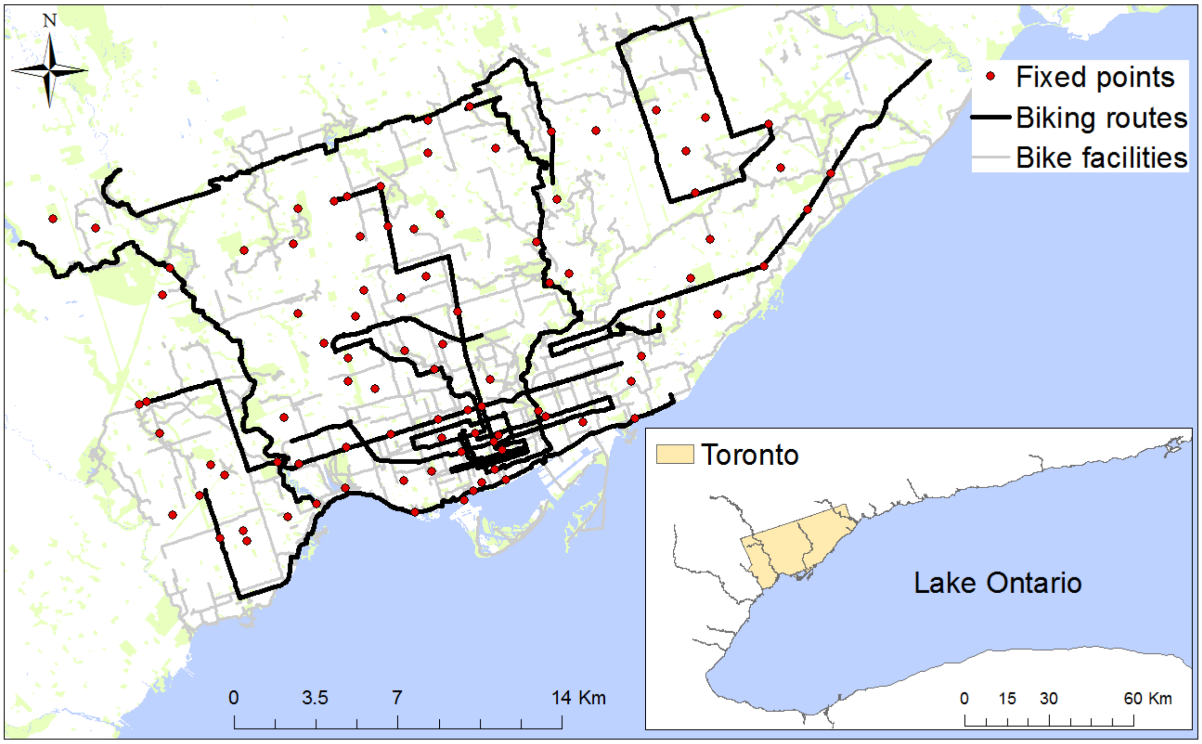


FIGURE S1 Fixed points and biking routes, overlaying the bicycle facilities across the Toronto bike network.

Using knowledge of the proportions of various land-uses in Toronto (land-use data in GIS format) and knowledge of the road network (GIS format), the routes were designed iteratively by comparing the chosen sample with the land-use composition of the City of Toronto. The aim was to go through different types of land uses while comparing with the land use of Toronto. The final selected routes in fact, represent the land-use distribution in the City. TABLE S1 illustrates the proportions of the areas of the various land-uses within a 50m buffer around the road segments, compared with the corresponding proportions within Toronto. We do not observe significant bias.

As for the fixed points, we equally distributed them between intersections and mid-blocks, and were typically located on the curbside, i.e. a couple of meters away from traffic, which is why we qualify them as near-road fixed points. We placed 40 fixed points (i.e. almost half of them) along the biking routes. We also looked at the traffic counts estimated at the fixed point locations by the city of Toronto [1] and compared them with the traffic counts determined for the City of Toronto as a whole. We computed TABLE S2 from this data.

TABLE S1 Proportions of land use areas within a 50m buffer around the biking routes designed and around the fixed points selected, as well as within the City of Toronto

	Residential	Open Area	Parks Recreational	Resource Industrial	Governmental Institutional	Commercial
Within 50m of the routes sampled	49.4%	9.7%	22.1%	10.2%	5.8%	3.0%
Within 50m around the fixed points sampled	41.6%	17.5%	13.7%	9.1%	6.7%	11.3%
In Toronto	50.2%	5.5%	18.7%	16.6%	6.7%	2.3%

TABLE S2: Summary of the 8-h traffic volumes at the fixed points and at the various locations of measurement in the city of Toronto [1]

	Min	Mean	Median	Max
Fixed points	6,011	19,833	18,204	45,185
City of Toronto	1,081	15,929	14,475	53,678

1.2 Frequency of visits

TABLE S3 presents the frequency distribution of the number of visits and the time spent on each fixed point and each road segment. When looking at the results from the road segments, it is clear that the relationship between the number of visits and the number of seconds per road segment is not linear, reflecting differences in the length of the road segments as well as differences in traffic conditions and travel speed.

TABLE S3 Frequency distribution of the number of visits and time spent at each fixed point and on each road segment

	Number of visits		Time spent	
	Fixed points	**Road segments**	**Fixed points (minutes)**	**Road segments (seconds)**
Minimum	3	1	60	1
Quartile 1	5	2	100	19
Median	5	5	100	52
Quartile 3	6	6	120	122
Maximum	6	26	120	3675
Average	5.1	4.3	102	106.5

TABLE S4 presents the proportion of sampling that occurred during the different days of the week and time blocks. We do not observe significant bias between the two different campaigns. Weekends were on purpose less covered. Time block 2 was slightly more covered as a response to a greater availability of the research assistants during that time block.

TABLE S4 Proportions of sampling occurring during the different days and time blocks

		Cyclists	Fixed points
Day of the week	**Monday**	0.10	0.13
	Tuesday	0.20	0.14
	Wednesday	0.22	0.18
	Thursday	0.20	0.23
	Friday	0.19	0.17
	Saturday	0.04	0.08
	Sunday	0.04	0.07
Time block	**Block 1**	0.30	0.33
	Block 2	0.39	0.34
	Block 3	0.30	0.33

1.3 Land-use and built environment characterisation

TABLE S5 summarises the predictors computed for each fixed point or road segment visited during the campaign and later used for the development of the land-use regression (LUR) models.

TABLE S5 Land-use and built environment predictors computed

Description	Unit	Layer	Source	Type	Buffers (m)
Area of the buildings	m^2	Ontario Building footprint	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Average height of the buildings	m	3D Massing Toronto	Open Data Toronto, City Planning (2014)	Polygon	25, 50, 100
Number of bus stops	Count	TTC schedule	Open Data Toronto, TTC Geographic Data (June, 2016)	Point	50, 100, 200, 300, 500, 750, 1000
Length of the bus roads	m	TTC Bus Routes	Open Data Toronto, TTC Geographic Data (June, 2016)	Polyline	50, 100, 200, 300, 500, 750, 1000
Area of the commercial land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Distance from the closest airport	m	Aerodromes Ontario	DMTI Spatial Inc. (2014)	Point	/
Distance from the closest highway (types 1, 2 and 3)	m	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	/
Distance from the closest major road (type 4)	m	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	/
Distance to the closest NOx emitting chimney	m	NPRI NOx	Environment and Climate Change Canada (2015)	Points	/
Distance to the closest PM emitting chimney	m	NPRI NOx	Environment and Climate Change Canada (2015)	Points	/
Distance from the closest railline	m	Railroads Ontario	DMTI Spatial Inc. (2014)	Polyline	/
Distance from the shore	m	Waterbody Area Ontario	DMTI Spatial Inc. (2014)	Polyline	/

Road segments unique ID	/	Bikeways Toronto	Open Data Toronto, Transportation Services, Cycling Infrastructure & Programs (June, 2016)	Polyline	/
Area of the governmental and institutional land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Length of the highways (types 1, 2 and 3)	m	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	50, 100, 200, 300, 500, 750, 1000
Area of the resource and industrial land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Number of intersections	Count	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	50, 100, 200, 300, 500, 750, 1000
Length of the road segment	m	Bikeways Toronto	Open Data Toronto, Transportation Services, Cycling Infrastructure & Programs (June, 2016)	Polyline	/
Length of the major roads (type 4)	m	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	50, 100, 200, 300, 500, 750, 1000
Maximum height of the buildings	m	3D Massing Toronto	Open Data Toronto, City Planning (2014)	Polygon	25, 50, 100
Number of NO_x emitting chimneys	Count	NPRI NOx	Environment and Climate Change Canada (2015)	Points	50, 100, 200, 300, 500, 750, 1000
Number of PM emitting chimneys	Count	NPRI PM	Environment and Climate Change Canada (2015)	Points	50, 100, 200, 300, 500, 750, 1000
Area of the open area land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Area of the parks land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Population	Count	Neighbourhoods Toronto	Open Data Toronto, Social Development, Finance & Administration (2011)	Polygon	500, 750, 1000
Length of rail lines	m	Railroads Ontario	DMTI Spatial Inc. (2014)	Polyline	50, 100, 200, 300, 500, 750, 1000
Area of the residential land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000
Type of road	/	Bikeways Toronto and Roads Ontario	Open Data Toronto, Transportation Services, Cycling Infrastructure & Programs (June, 2016) and DMTI Spatial Inc. (2014)	Polyline	/
Length of the roads (types 1, 2, 3, 4, 5 and 6)	m	Roads Ontario	DMTI Spatial Inc. (2014)	Polyline	50, 100, 200, 300, 500, 750, 1000
Hourly traffic (average from 6am to 7pm)	Count			Polyline	50, 100, 200, 300, 500, 750, 1000
Number of trees	Count	Trees Toronto	Open Data Toronto, Parks, Forestry & Recreation - Urban Forestry (June, 2016)	Point	50, 100, 200, 300, 500, 750, 1000
Area of the waterbody land use	m^2	Ontario land use	DMTI Spatial Inc. (2014)	Polygon	50, 100, 200, 300, 500, 750, 1000

1.4 Exposure surface

FIGURE S2 presents the area of Toronto considered for the development of the exposure surfaces.

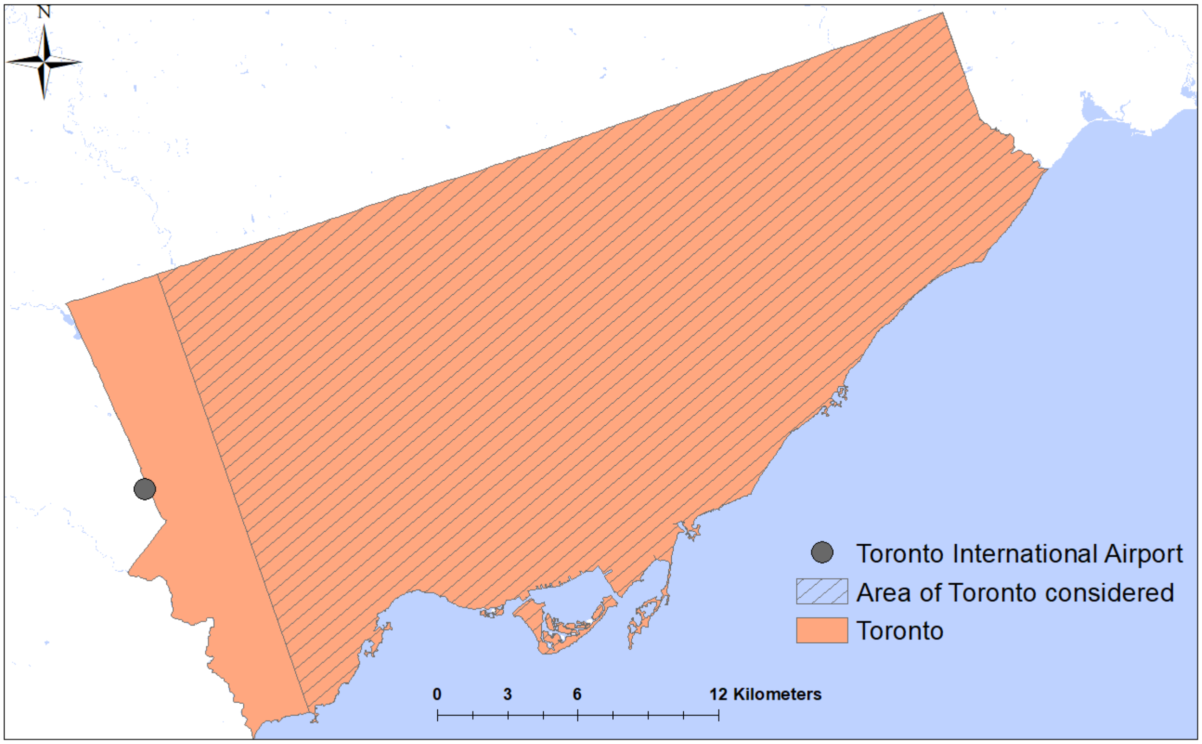


FIGURE S2 Comparison between the official boundaries of Toronto and the area considered for the development of the exposure surfaces.

2 RESULTS

2.1 Collocation of portable monitors

TABLE S6 presents the Pearson correlation coefficients between the different DiSCMinis for the combination of the measurements done during the two collocations. The pink highlighting indicates that the DiSCMinis concerned could not be used as the reference. Indeed, DiSCMinis 557 and 558 were deployed for the first and the second collocation only respectively, and DiSCMini 647 turned off very quickly during the second collocation due to a charging issue. Therefore, it is the DiSCMini 551 that was chosen as the reference to correct the other devices.

TABLE S6 Pearson correlation coefficients between the different DiSCMinis (data from the two collocations aggregated)

DiSCMini ID	551	552	556	557	558	608	647
551	1						
552	0.93	1					
556	0.95	0.92	1				
557	0.98	0.96	0.94	1			
558	0.75	0.68	0.71	NA	1		
608	0.78	0.75	0.76	0.70	0.89	1	
647	0.93	0.91	0.89	0.94	0.36	0.68	1
Average per DiSCMini	0.89	0.86	0.86	0.90	0.68	0.76	0.79

The correction equations applied to the other instruments are the following, where C means Concentration, and the number into brackets and corr. refer to the ID of the DiSCMinis and the measurement corrected respectively:

$$C(552\ corr.) = 1.379 * C(552) + 555.2$$

$$C(556\ corr.) = 0.9411 * C(556) - 1{,}203$$

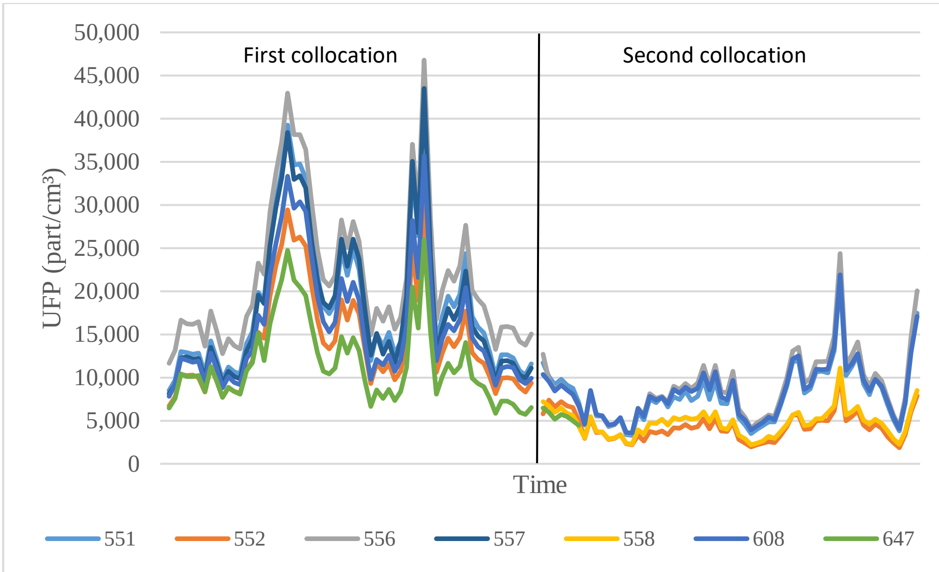
$$C(557\ corr.) = 1.019 * C(557) + 149.4$$

$$C(558\ corr.) = 1.524 * C(558) - 3{,}807$$

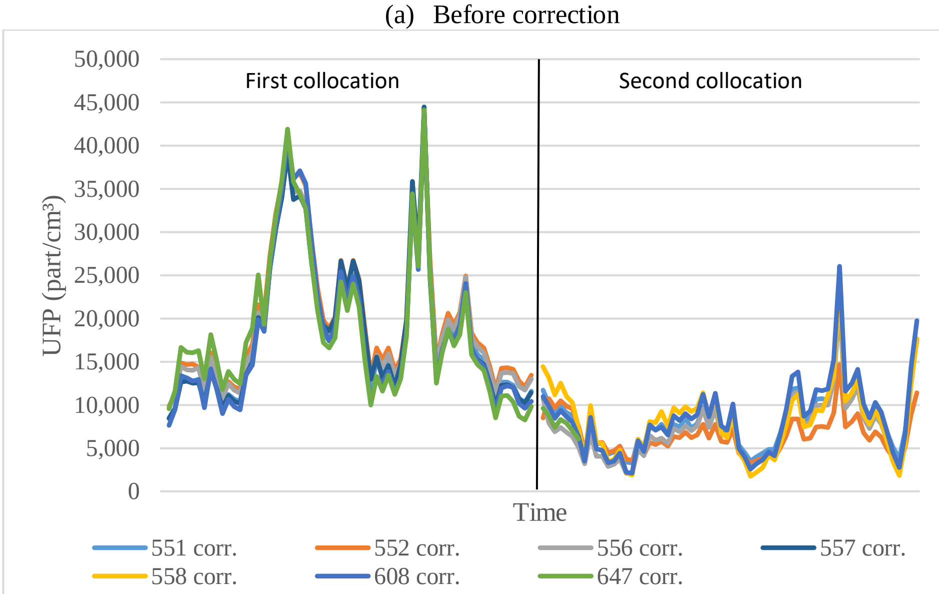
$$C(608\ corr.) = 1.304 * C(608) - 2{,}534$$

$$C(647\ corr.) = 1.768 * C(647) - 1{,}812$$

FIGURE S3 presents the measurements during the collocations before and after the corrections were applied.



(a) Before correction



(b) After correction

FIGURE S3 Graphs of the non-corrected and corrected UFP hourly averages of DiSCMinis, recorded during the two colocations (the legend indicates the ID of the DiSCMinis).

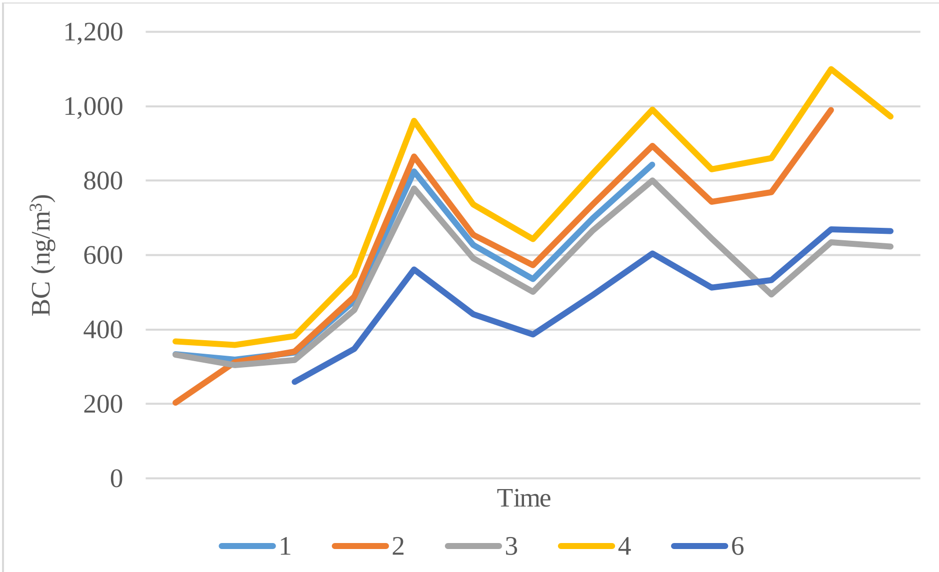
FIGURE S4 shows the hourly averaged BC measurements during the first and the second collocation. We can see how similar they are, particularly during the second colocation.

TABLE S7 presents the Pearson correlation coefficients between the different Microaeths for the combination of the measurements done during the two collocations. The pink highlighting indicates that the Microaeth concerned was collocated only once.

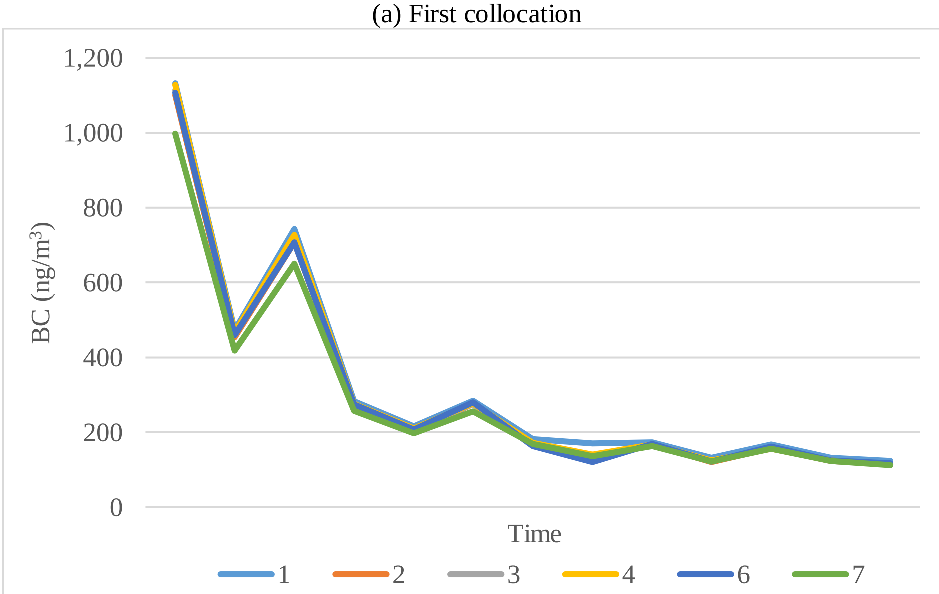
TABLE S7 Pearson correlation coefficients between the different Microaeths (data from the two collocations aggregated)

Microaeth ID	1	2	3	4	6	7
1	1					
2	0.97	1				
3	0.98	0.97	1			
4	0.97	0.96	0.99	1		
6	0.72	0.60	0.72	0.71	1	
7	0.98	0.99	0.99	0.98	0.99	1
Average per Microaeth	0.92	0.90	0.93	0.92	0.75	0.99

FIGURE S4 shows the hourly averaged BC measurements during the first and the second collocation. We can see how similar they are, particularly during the second colocation.



(a) First collocation



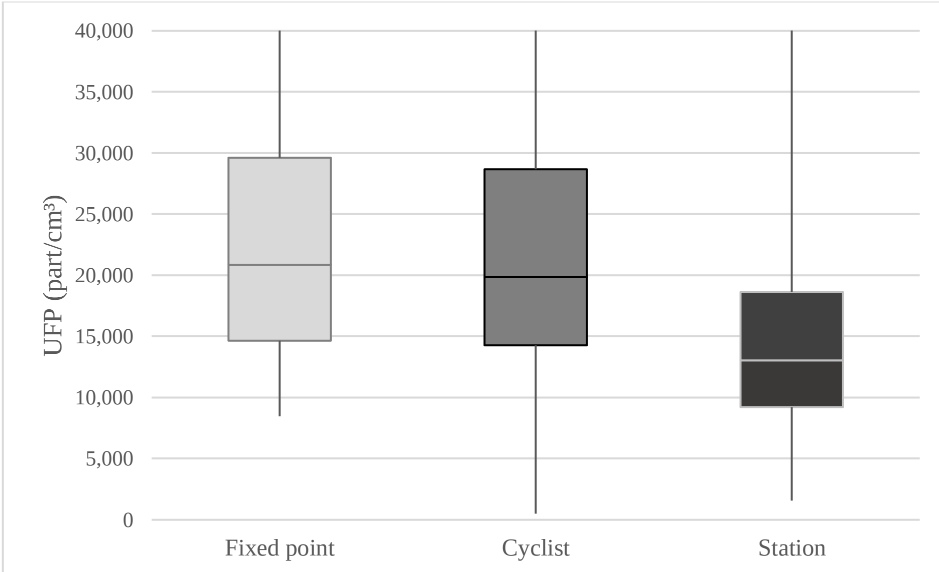
(b) Second collocation

FIGURE S4 Graphs of the hourly averages of the Microaeths measurements during the first (a) and the second (b) collocation.

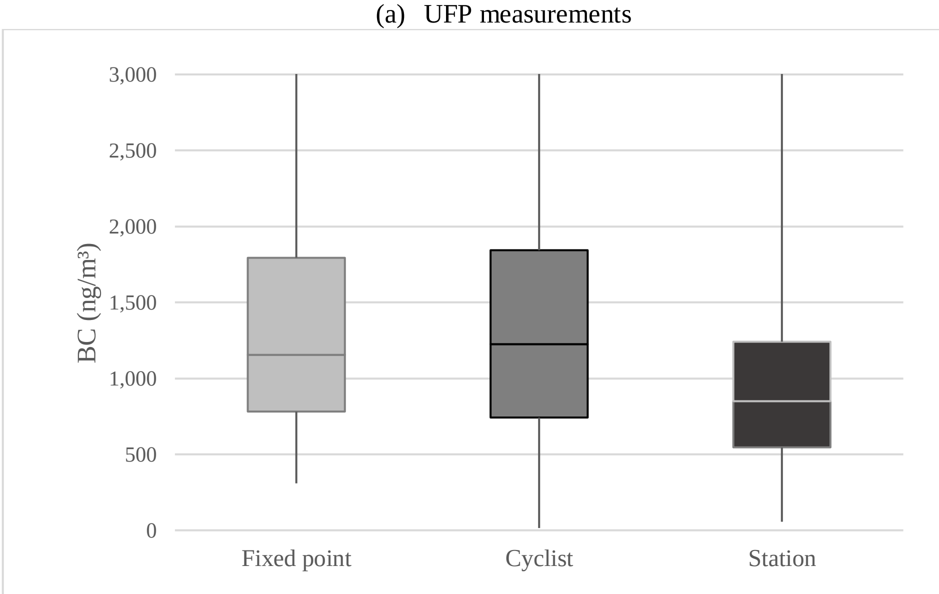
2.2 Descriptive statistics

FIGURE S5 presents boxplots of the measurements carried out during the two campaigns. The station boxplot corresponds to the reference station located at the street level.

FIGURE S6 and FIGURE S7 show the comparison of the average UFP and BC concentrations at fixed points and on the portions of road segments within buffers of 50 and 100m around the fixed points. The scatterplots show the influence of very high measurements done on road segments.

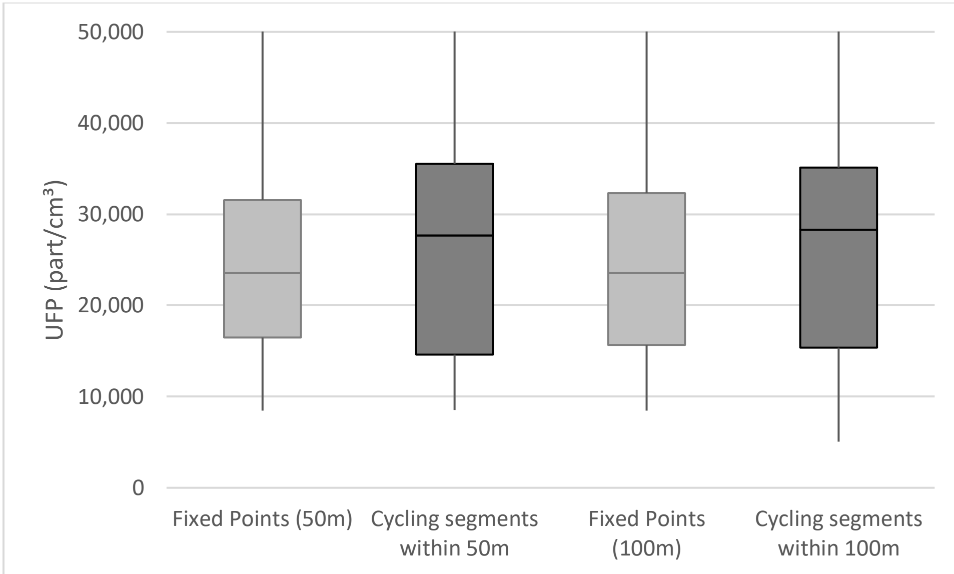


(a) UFP measurements

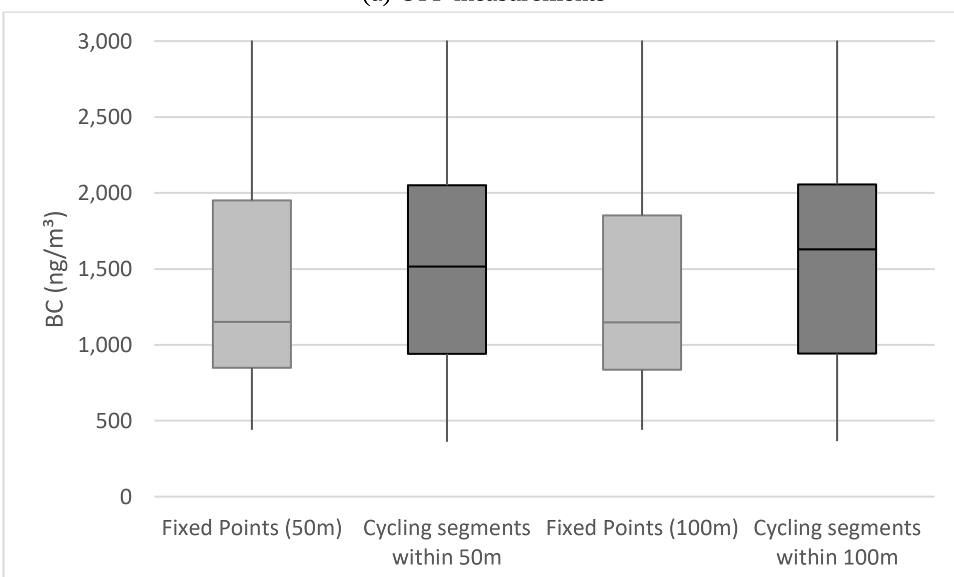


(b) BC measurements

FIGURE S5 Boxplots of the UFP and BC measurements conducted during the short-term fixed points and cycling campaigns as well as recorded by a reference station during the same period (the lower and the upper whiskers correspond to the minimum and maximum respectively, and the box is drawn between the first and the third quartile with the middle line corresponding to the median. Axes are truncated).

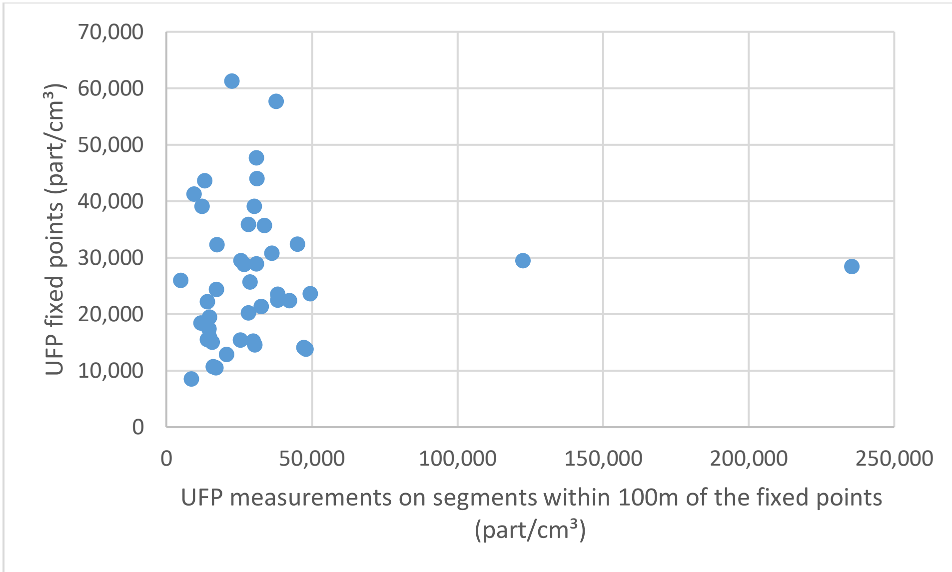


(a) UFP measurements

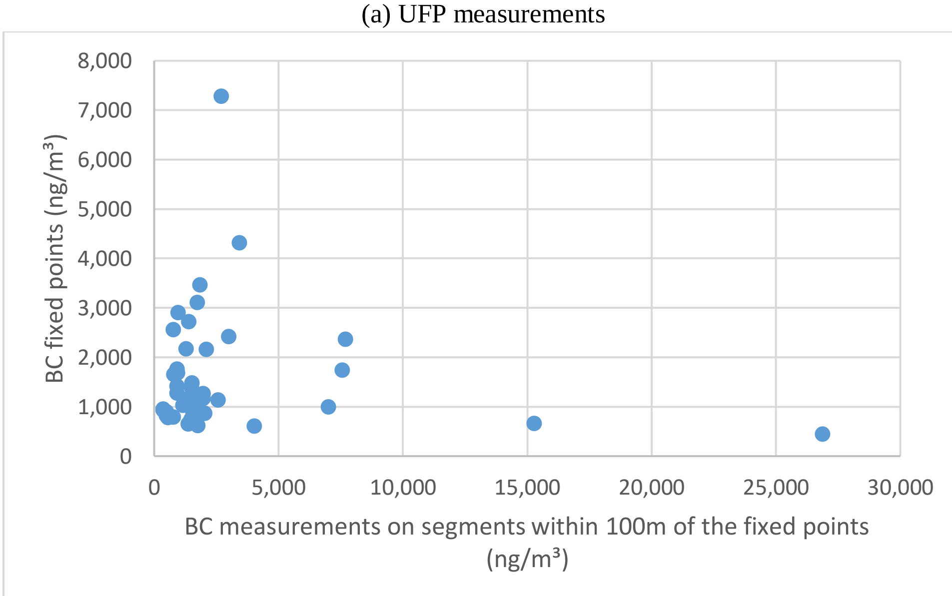


(b) BC measurements

FIGURE S6 Comparison between UFP and BC levels at fixed points and on the surrounding road segments (the lower and the upper whiskers correspond to the minimum and maximum respectively, and the box is drawn between the first and the third quartile with the middle line corresponding to the median. Axes are truncated).



(a) UFP measurements



(b) BC measurements

FIGURE S7 Scatterplots of the comparison between measurements at fixed points and on the road segments within 100m of the segments.

3 LUR MODELS

3.1 Fixed points

3.1.1 UFP

TABLE S8 presents the results of the different ln(UFP) models developed based on the fixed points data. The subscript "t" indicates that the concentrations used were temporally adjusted and that meteorological variables were not included in the model.

TABLE S8 Models developed based on the ln(UFP) fixed points data (unstandardized coefficients)

Variables	Forward	Backward	Forward_t	Backward_t
Adjusted R^2	**0.405**	**0.300**	**0.323**	**0.176**
Number of variables	**10**	**4**	**9**	**2**
Intercept	7.08E+00***	6.87E+00***	8.35E+00***	9.32E+00***
Wind Speed	-2.41E-02*			
Temperature	8.76E-02***	1.01E-01***		
Number of bus stops within 300m	4.96E+03*			
Commercial area within 200m				1.38E-01*
Distance to the CBD	2.57E-05**		3.79E-05***	
Length of highways within 1000m	1.01E+02*			
Length of highways within 750m		8.75E+01**	1.15E+02***	
Number of intersections within 500m			7.10E+03**	
Number of intersections within 750m	7.36E+03**			
Length of major roads within 1000m			1.30E+02**	
Length of major roads within 500m	8.43E+01*	1.33E+02***		
Maximum building height within100m			1.54E-03*	
Open area within 1000m	-1.49E+00*			
Open area within 300m			-8.99E-01*	
Park area within 50m			3.10E-01.	
Population within 500m	4.34E+01.	4.16E+01.	9.63E+01***	
Length of roads within 500m				6.12E+01***
Area of water within 100m			4.64E-01	
Area of water within 50m	6.28E-01.			

Significance codes: 0 '***', 0.001 '**',0.01 '*', 0.05 '.'

TABLE S9, FIGURE S8 Cross validation of the UFP fixed point modelsand FIGURE S9 present the results of the cross validation of the ln(UFP) fixed point models.

TABLE S9 Results of the cross-validation of the UFP fixed point models

	100% points	**90% points**	**10% points**	**10% points**
Model	**Adjusted R^2**	**Adjusted R^2 (median)**	**Correlation (median)**	**RMSE (median)**
Forward	0.405	0.399	0.597	0.329
Backward	0.300	0.298	0.550	0.332
Forward_t	0.323	0.323	0.553	0.342
Backward_t	0.176	0.172	0.502	0.356

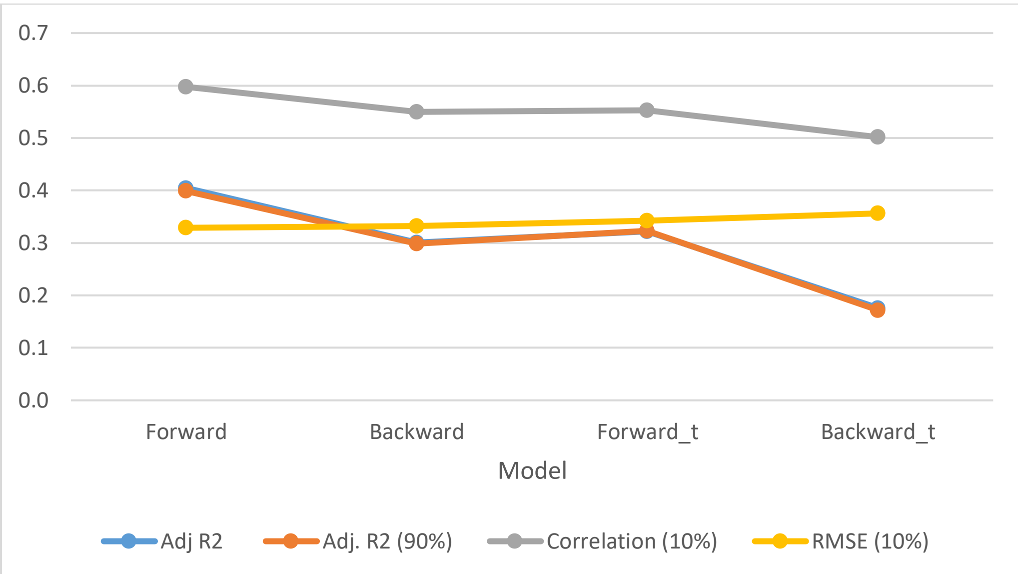
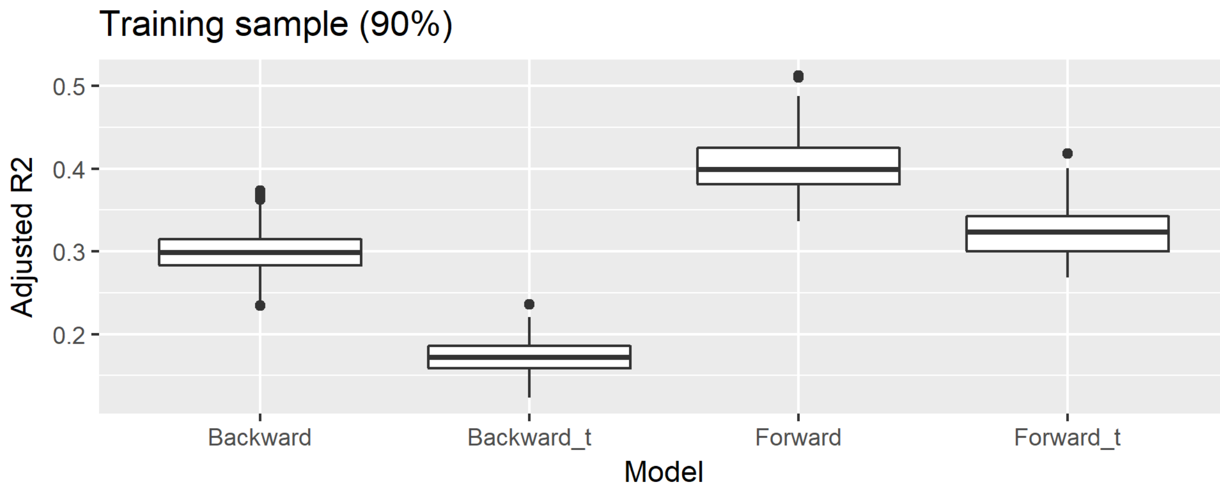
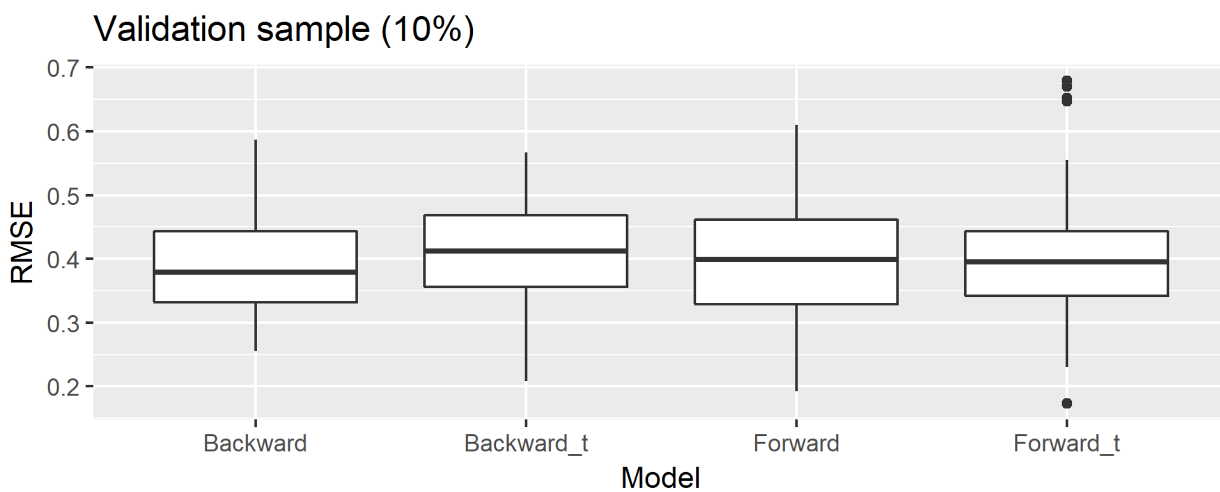


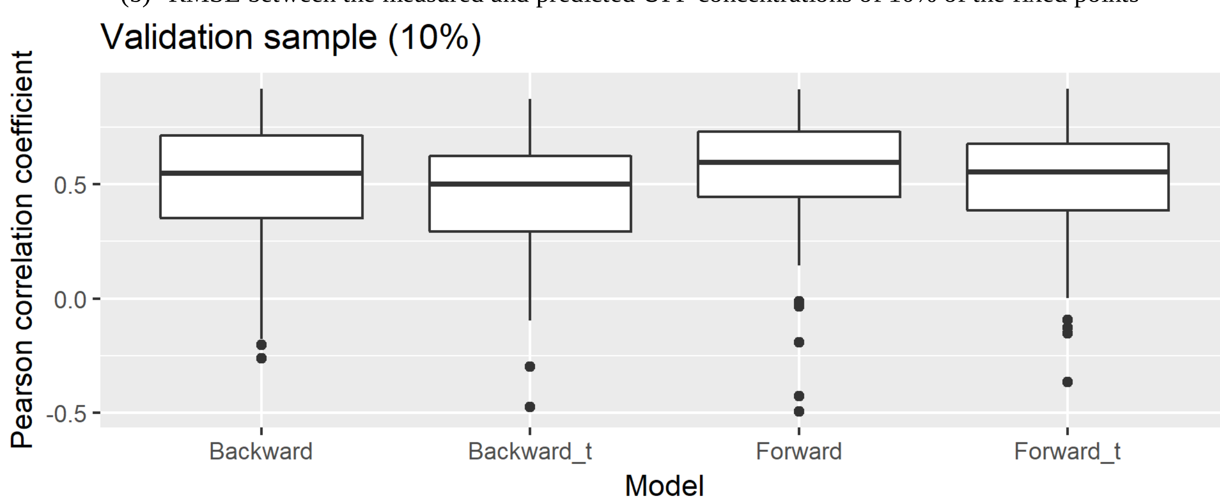
FIGURE S8 Cross validation of the UFP fixed point models.



(a) Adjusted R^2 of the model developed based on 90% of the fixed points



(b) RMSE between the measured and predicted UFP concentrations of 10% of the fixed points



(c) Pearson correlation coefficient between the measured and predicted UFP concentrations of 10% of the fixed points

FIGURE S9 Boxplots of the 100-fold cross validation of the UFP fixed point models (the lower and upper whiskers correspond to the minimum and maximum no further than 1.5*IQR, the box corresponds to the interquartile range (IQR) with the middle line corresponding to the median).

When using concentrations adjusted temporally, the LUR models developed show lower adjusted R^2. The forward model provides the highest adjusted R^2, i.e. 0.405, and the cross-validation confirms its better performances: the average Pearson correlation coefficient between the predicted and measured concentrations of the hold-out sample is the highest, while the average RMSE is similar to the other models.

For these reasons, we choose the forward model for the fixed points UFP LUR. FIGURE S10 presents the correlation matrix between the variables included in the forward model. The highest absolute Pearson correlation coefficient is between the variables inter_750m (number of intersections within a buffer of 750m) and dCBD (distance to the Central Business District), and is -0.60.

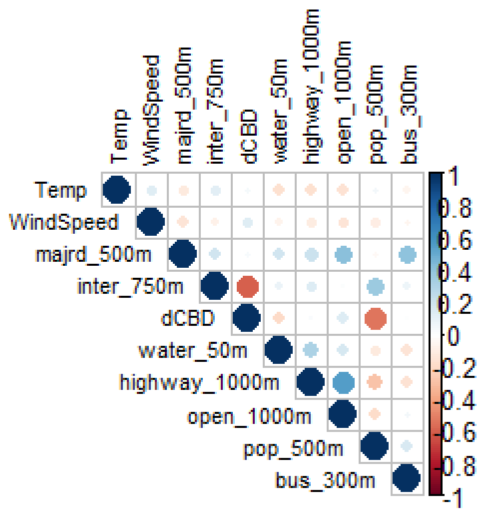


FIGURE S10 Correlation matrix between the predictors included in the fixed points UFP forward model.

3.1.2 BC

TABLE S10 presents the results of the different BC models developed based on the fixed points data.

TABLE S10 Models developed based on the BC fixed points data (unstandardized coefficients)

Variables	Forward	Backward	Forward_t	Backward_t
Adjusted R^2	**0.525**	**0.491**	**0.503**	**0.430**
Number of variables	**7**	**9**	**8**	**5**
Intercept	3.13E+00**	6.13E+00***	6.09E+00***	6.24E+00***
Relative Humidity	2.76E-02**			
Temperature	7.10E-02*			
Building height within 100m		2.58E-03.		
Number of bus stops within 100m	1.62E+03*			
Number of bus stops within 300m			1.34E+04***	1.12E+04***
Length of bus line within 100m	5.00E+00*	7.01E+00**		
Commercial area within 500m	9.87E-01*		1.44E+00***	
Commercial area within 750m		1.65E+00*		2.08E+00***
Distance to the shore		-2.36E-05*		
Governmental area within 500m			1.10E+00*	
Length of highway within 1000m		1.24E+02**	1.60E+02***	1.08E+02*
Industrial area within 500m			1.24E+00***	8.46E-01***
Length of major road within 100m				
Length of major road within 300m		8.09E+01**		
Open area within 500m			-1.21E+00*	
Park area within 1000m		-1.27E+00*		
Population within 500m	9.32E+01***	8.92E+01**	1.20E+02***	1.37E+02***
Residential area within 200m		1.53E-01.		
Residential area within 300m				
Traffic within 1000m				
Traffic within 750m	4.55E+01***			
Number of trees within 100m			1.36E+02.	

Significance codes: 0 ‘***’, 0.001 ‘**’,0.01 ‘*’, 0.05 ‘.’

TABLE S11, FIGURE S11 and FIGURE S12 present the results of the cross validation of the BC fixed point models.

TABLE S11 Results of the cross-validation of the BC fixed point models

	100% points	90% points	10% points	10% points
Model	**Adjusted R^2**	**Adjusted R^2 (median)**	**Correlation (median)**	**RMSE (median)**
Forward	0.525	0.522	0.732	0.420
Backward	0.491	0.486	0.718	0.463
Forward_t	0.503	0.498	0.729	0.409
Backward_t	0.430	0.421	0.715	0.420

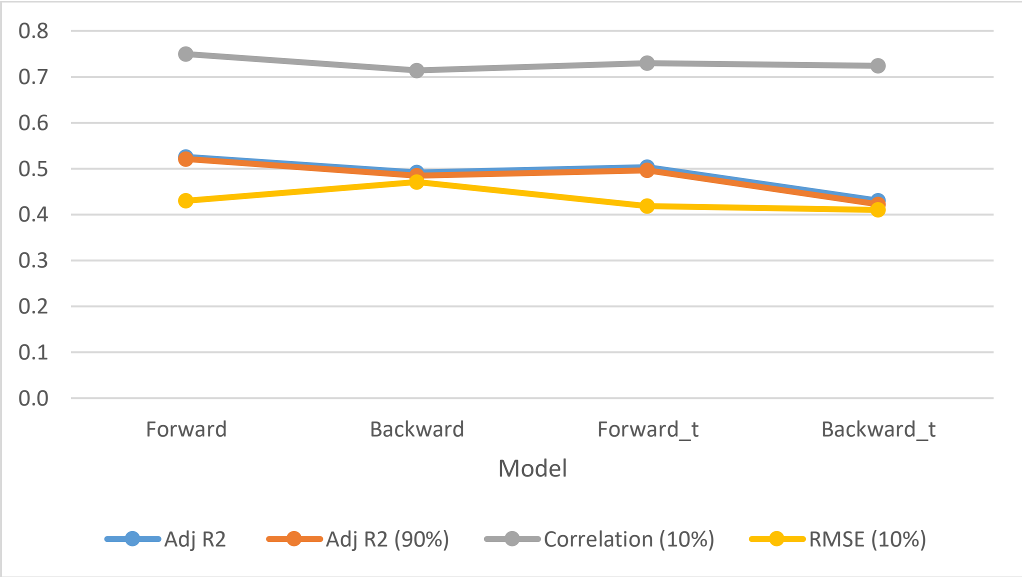
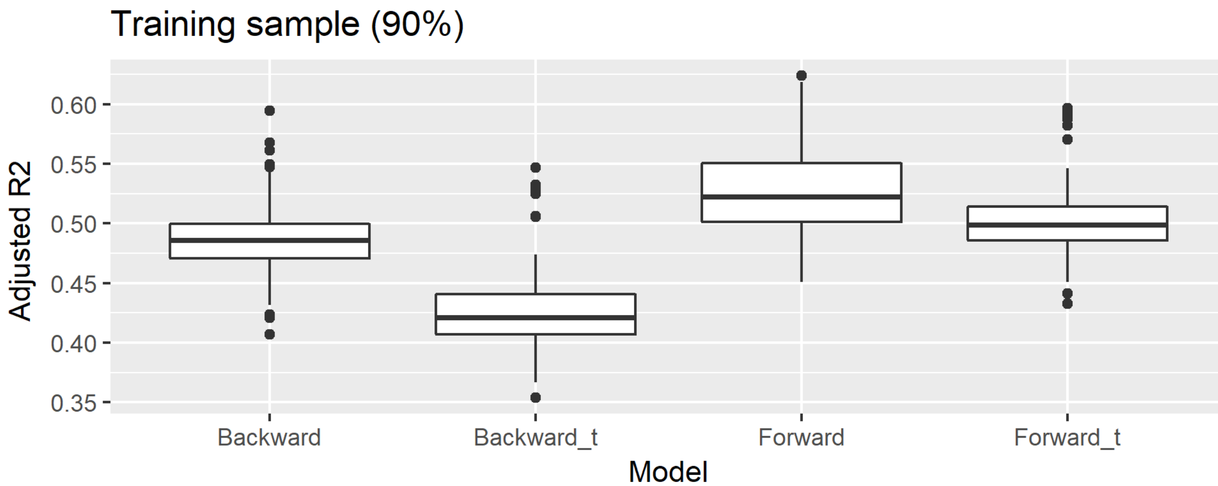
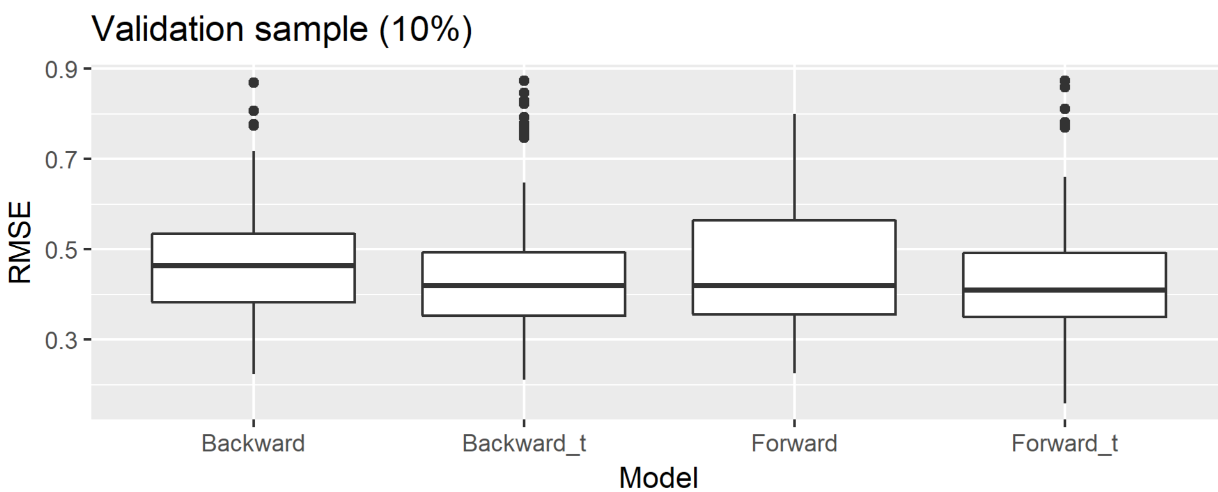


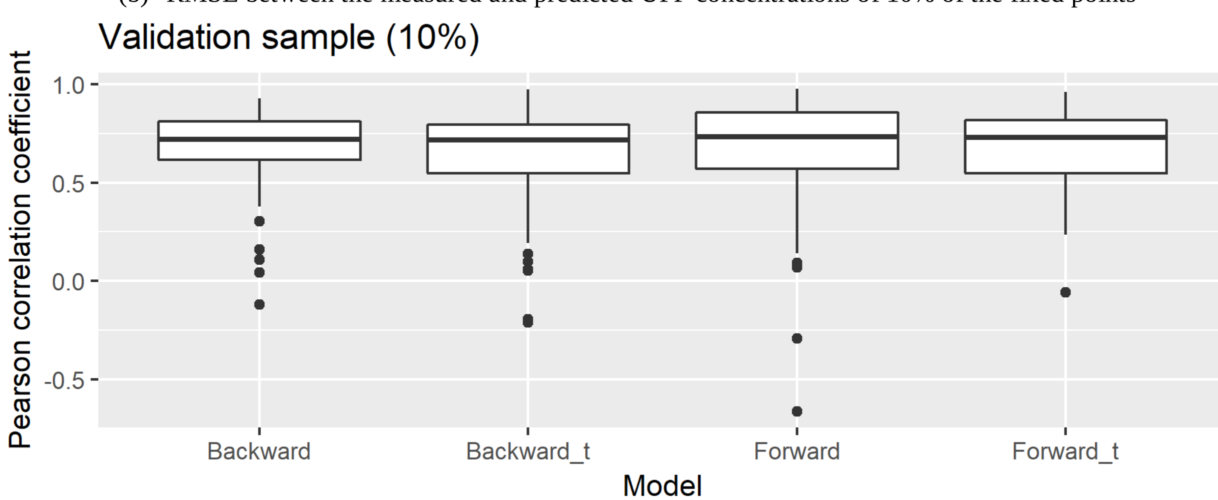
FIGURE S11 Cross validation of the BC fixed point models.



(a) Adjusted R^2 of the model developed based on 90% of the fixed points



(b) RMSE between the measured and predicted UFP concentrations of 10% of the fixed points



(c) Pearson correlation coefficient between the measured and predicted UFP concentrations of 10% of the fixed points

FIGURE S12 Boxplots of the 100-fold cross validation of the BC fixed point models (the lower and upper whiskers correspond to the minimum and maximum no further than 1.5*IQR, the box corresponds to the interquartile range (IQR) with the middle line corresponding to the median).

The results with the temporally adjusted BC concentrations are closer to the results with non-temporally adjusted concentrations than it was the case with UFP, but they still present lower performances. The forward model gives a higher adjusted R^2 similarly to the case of UFP. As for the cross validation results, they are better for the forward model than for the backward model, with a lower RMSE of the hold-out sample and a higher adjusted R^2 of the model with the 90% points. We therefore decide to take the forward model.

FIGURE S13 presents the correlation matrix between the variables included in the forward model. The highest absolute Pearson correlation coefficient is between bus_100m (number of bus stops within a buffer of 100m) and busline_100m (length of bus lines within 100m) and is 0.52.

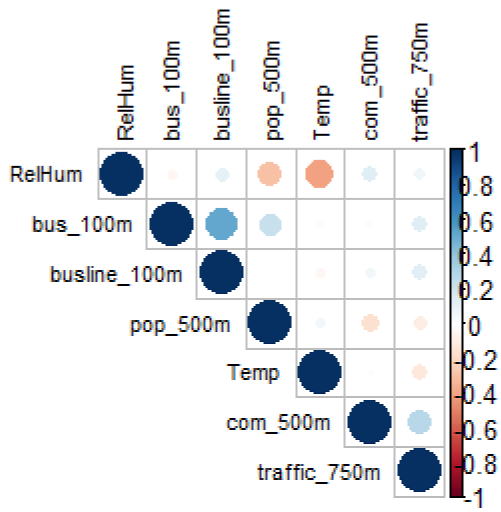


FIGURE S13 Correlation matrix between the predictors included in the fixed points BC forward model.

3.2 Cycling

3.2.1 UFP

TABLE S12 presents the results of the different UFP models developed based on the cycling data.

TABLE S12 Models developed based on the UFP cycling data (unstandardized coefficients)

Variables	Forward	Backward	Forward_t	Backward_t
Adjusted R^2	**0.430**	**0.418**	**0.283**	**0.262**
Number of variables	**8**	**15**	**8**	**13**
Intercept	1.07E+01***	1.06E+01	1.01E+01	1.00E+01
Relative Humidity				
Temperature				
Wind Speed	-3.03E-02***	-3.22E-02		
Building area within 1000m				
Building area within 300m		3.93E-01		4.46E-01
Building height within 50m			6.80E-04	
Number of bus stops within200m			1.53E+03	
Commercial area within 1000m				-5.19E-01
Distance to the closest airport	-3.98E-05***	-2.58E-05	-2.61E-05	-1.58E-05
Distance to the closest major road		-2.13E-04		-2.60E-04
Distance to the closest NPRI NO_x chimney		-9.62E-06		-9.39E-06
Distance to the shore	-1.06E-05***		-1.65E-05	-8.70E-06
Governmental area within 1000m		-3.20E-01		-2.62E-01
Length of highway within 1000m		1.78E+01		
Industrial area within 1000m		-1.46E-01		
Length of major roads within 200m	1.34E+01***			
Length of major roads within 50m			2.15E+00	
Maximum building height within 25m		2.98E-04		2.89E-04
Maximum building height within 50m	5.87E-04***			
Number of NPRI PM chimneys within 1000m		1.28E+05		
Open area within 1000m				
Park area within 1000m		-2.15E-01	-3.83E-01	-4.03E-01
Population within 1000m		9.66E+00		-7.92E+00
Residential area within 200m				
Traffic within 1000m		6.91E+00		5.76E+00
Traffic within 100m	3.13E+00***		2.94E+00	
Traffic within 300m				
Number of trees within 1000m		2.83E+02		1.96E+02
Number of trees within 750m	1.76E+02***			
Water area within 100m				-2.90E-01
Water area within 200m		-4.80E-01		
Water area within 750m	-8.28E-01***		-4.97E-01	

Significance codes: 0 '***', 0.001 '**',0.01 '*', 0.05 '.'

TABLE S13, FIGURE S14 and FIGURE S15 present the results of the cross validation of the UFP cycling models.

TABLE S13 Results of the cross-validation of the UFP cycling models

	100% points	**90% points**	**10% points**	**10% points**
Model	**Adjusted R^2**	**Adjusted R^2 (median)**	**Correlation (median)**	**RMSE (median)**
Forward	0.430	0.428	0.668	0.344
Backward	0.418	0.416	0.653	0.352
Forward_t	0.283	0.283	0.530	0.365
Backward_t	0.262	0.262	0.507	0.372

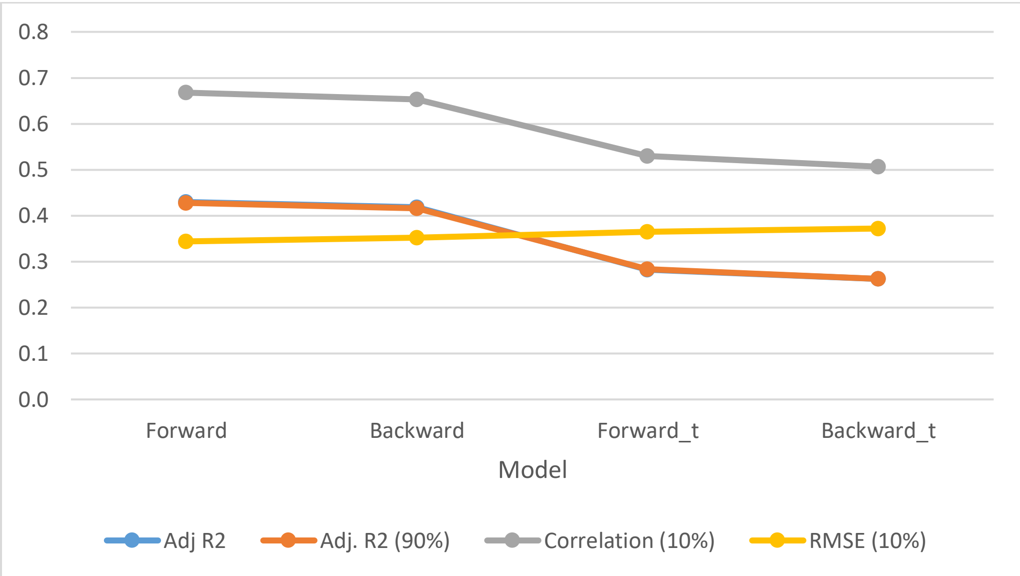
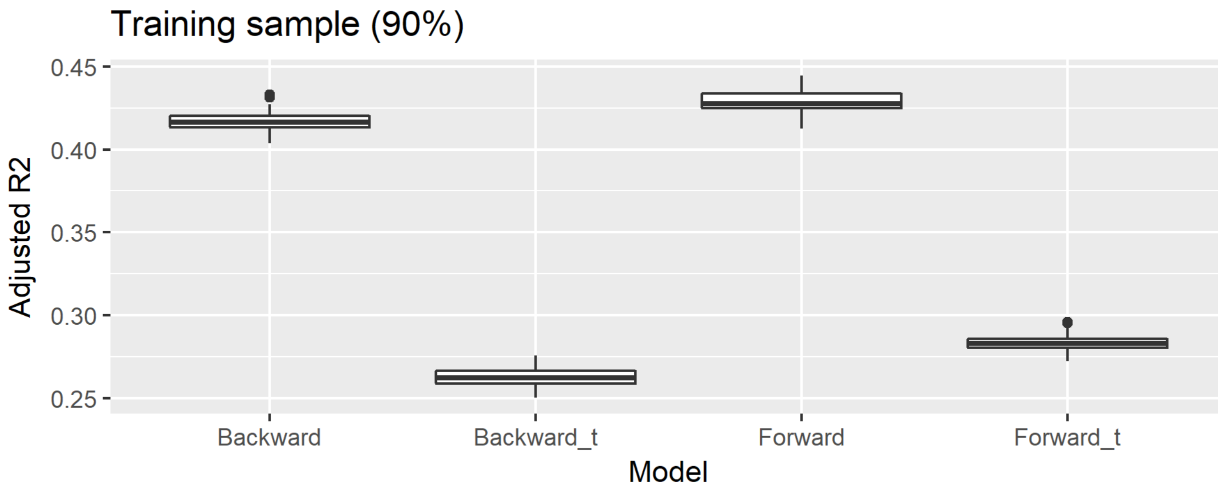
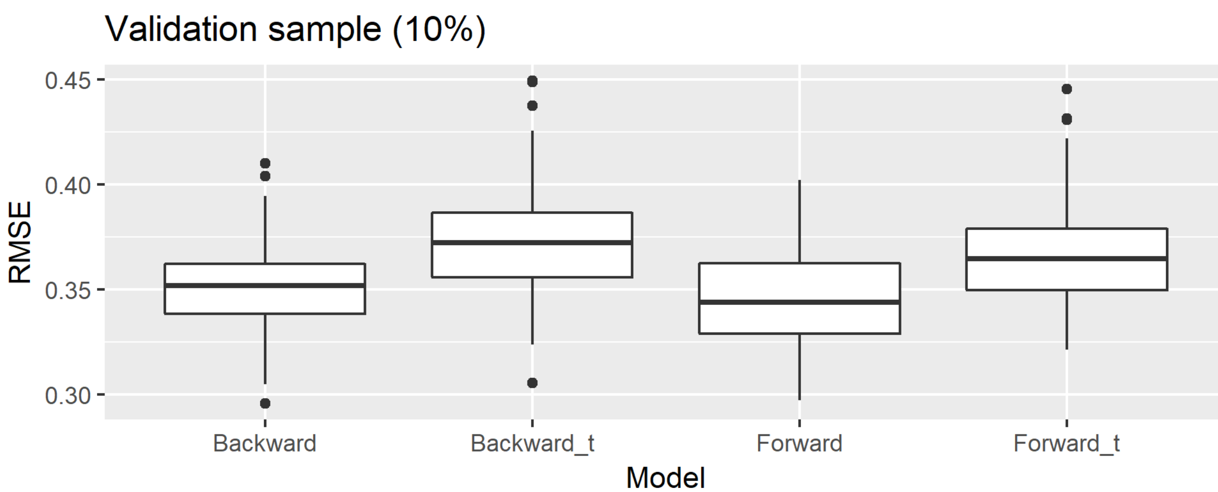


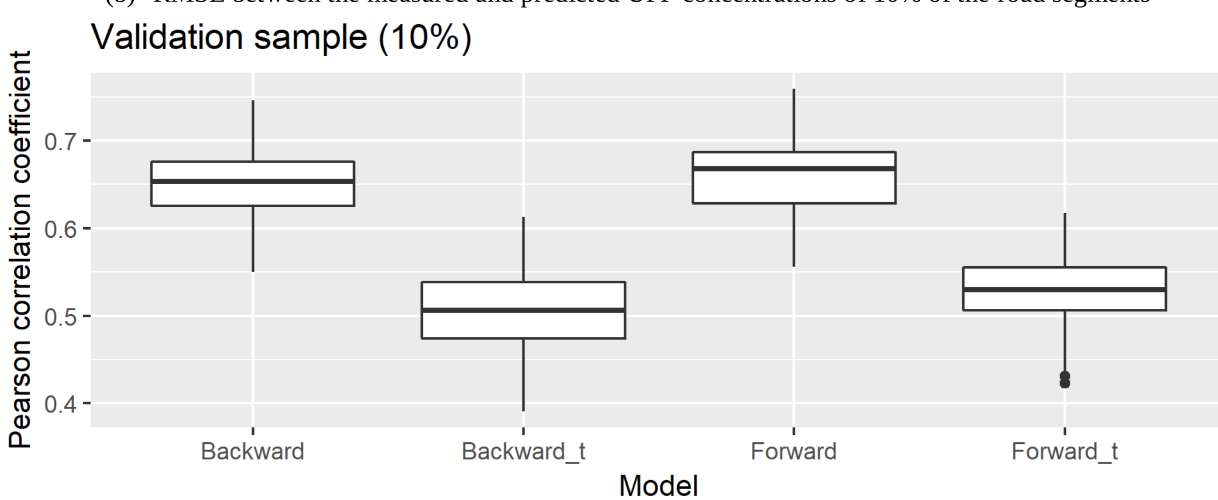
FIGURE S14 Cross validation of the UFP cycling models.



(a) Adjusted R^2 of the model developed based on 90% of the road segments



(b) RMSE between the measured and predicted UFP concentrations of 10% of the road segments



(c) Pearson correlation coefficient between the measured and predicted UFP concentrations of 10% of the road segments

FIGURE S15 Boxplots of the 100-fold cross validation of the UFP cycling models (the lower and upper whiskers correspond to the minimum and maximum no further than 1.5*IQR, the box corresponds to the interquartile range (IQR) with the middle line corresponding to the median).

Again, the temporal adjustment does not provide results as good as when using meteorological variables in the model. The backward model achieves an adjusted R^2 similar to the forward model but with more variables (15 variables versus 8 variables). Furthermore, the cross validation results are slightly better for the forward model, with a higher average correlation and a lower average RMSE for the 10% hold-out sample. For these reasons, we will keep the forward model.

FIGURE S16 presents the correlation matrix between the variables included in the forward model. The highest absolute Pearson correlation coefficient is between traffic_100m (traffic volume within 100m) and majrd_200m (length of major roads within 200m) and is equal to 0.418.

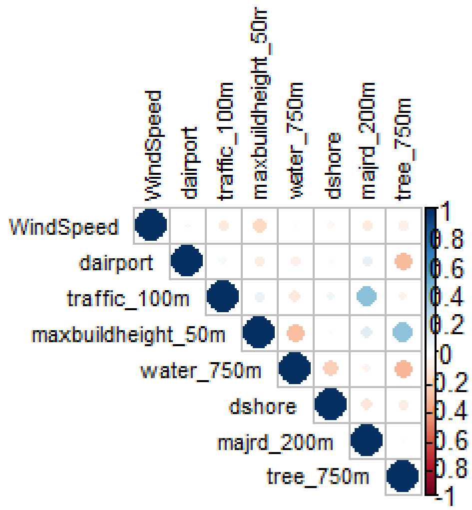


FIGURE S16 Correlation matrix between the predictors included in the cycling UFP forward model.

3.2.2 BC

TABLE S14 presents the results of the different BC models developed based on the cycling data.

TABLE S14 Models developed based on the BC cycling data (unstandardized coefficients)

Variables	Forward	Backward	Forward_t	Backward_t
Adjusted R^2	**0.434**	**0.436**	**0.302**	**0.316**
Number of variables	**11**	**15**	**7**	**10**
Intercept	6.23E+00***	6.18E+00***	6.78E+00***	6.72E+00***
Temperature	2.77E-02***	2.65E-02***		
Relative Humidity	1.33E-02***	1.40E-02***		
Wind speed	-3.89E-02***	-3.83E-02***		
Building area within 500m	4.81E-01***		5.70E-01***	
Number of bus stops within 100m				
Number of bus stops within 200m		2.64E+03***	2.79E+03***	3.52E+03***
Number of bus stops within 50m	4.25E+02***			
Commercial within 1000m		4.14E-01*		
Distance to the closest airport		-6.17E-06**		
Distance to the closest major road	-3.93E-04***	-3.54E-04***		-4.89E-04***
Distance to the closest NPRI PM chimney				
Distance to the shore				3.85E-06.
Length of highway within 50m		4.14E+00.		7.11E+00**
Industrial area within 1000m				4.91E-01***
Industrial area within 750m		1.36E-01*		
Number of intersections within 1000m			2.02E+03***	2.35E+03***
Length of major roads within 200m		7.23E+00.		
Length of major roads within 50m			4.95E+00***	
Number of NPRI NOx chimneys within 100m				
Number of NPRI NOx chimneys within 200m	8.19E+04***			
Number of NPRI PM chimneys within 1000m		1.15E+05.		
Park area within 300m				
Park area within 500m	-3.77E-01***	-3.74E-01***	-6.03E-01***	-5.29E-01***
Length of rail line within 300m				
Length of roads within 1000m		2.79E+01***		
Length of roads within 200m				
Length of roads within 500m	2.11E+01***			
Traffic within 100m	4.50E+00***	4.15E+00***	5.28E+00***	4.69E+00***
Number of trees within 750m				1.71E+02***
Water area within 500m				-9.23E-01***
Water area within 750m	-1.76E+00***	-1.72E+00***	-1.14E+00***	

Significance codes: 0 ‘***’, 0.001 ‘**’,0.01 ‘*’, 0.05 ‘.’

TABLE S15, FIGURE S17 and FIGURE S18 present the results of the cross validation of the BC cycling models.

TABLE S15 Results of the cross-validation of the BC cycling models

	100% points	90% points	10% points	10% points
Model	**Adjusted R^2**	**Adjusted R^2 (median)**	**Correlation (median)**	**RMSE (median)**
Forward	0.434	0.434	0.656	0.467
Backward	0.436	0.436	0.660	0.475
Forward_t	0.302	0.302	0.550	0.496
Backward_t	0.316	0.316	0.564	0.495

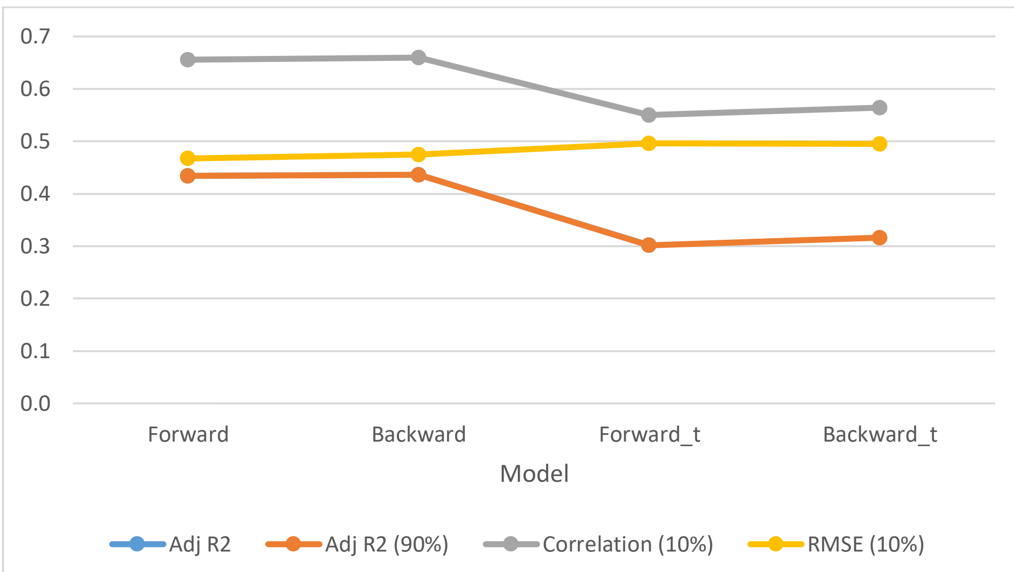
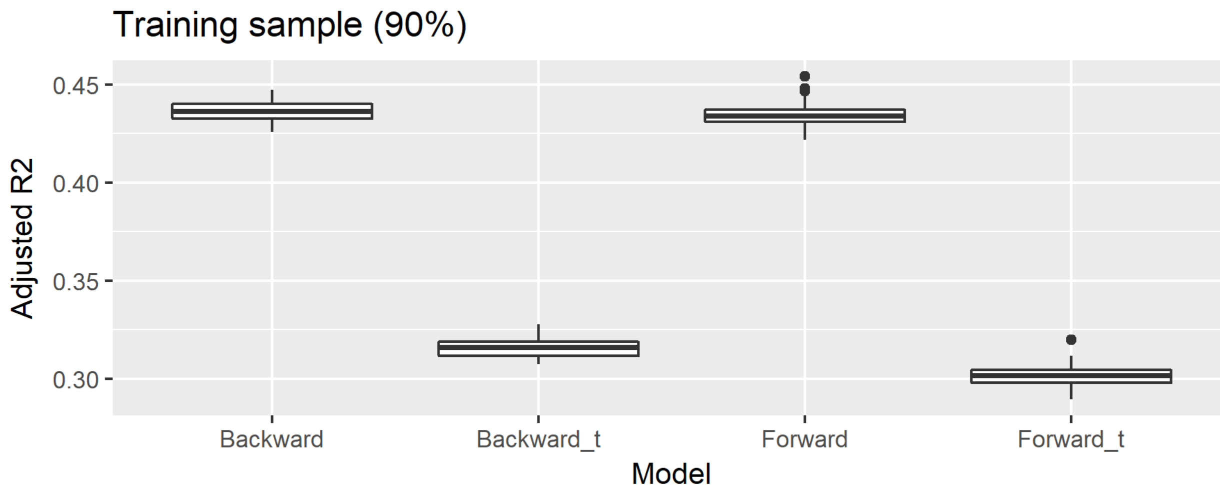
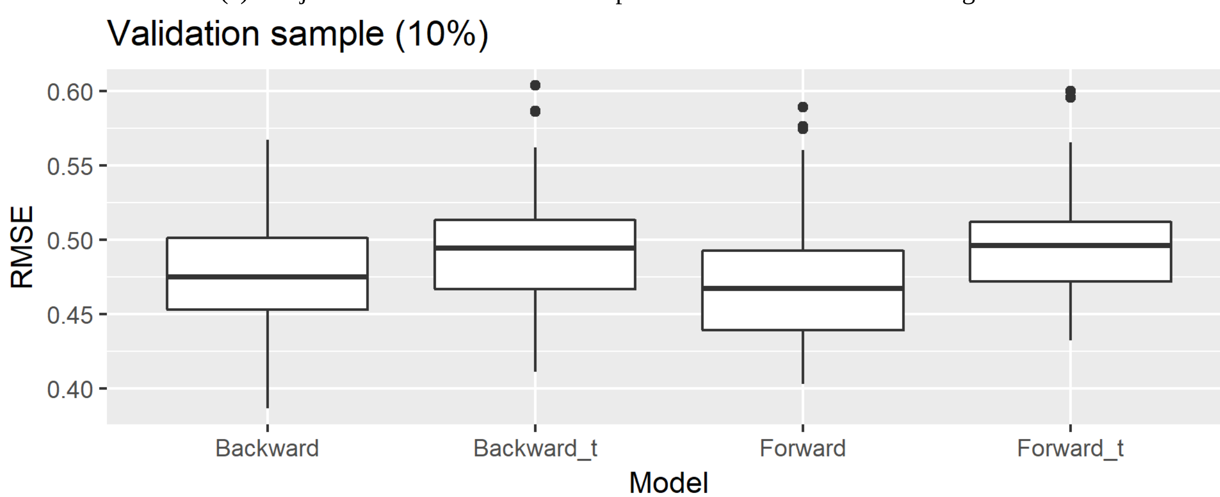


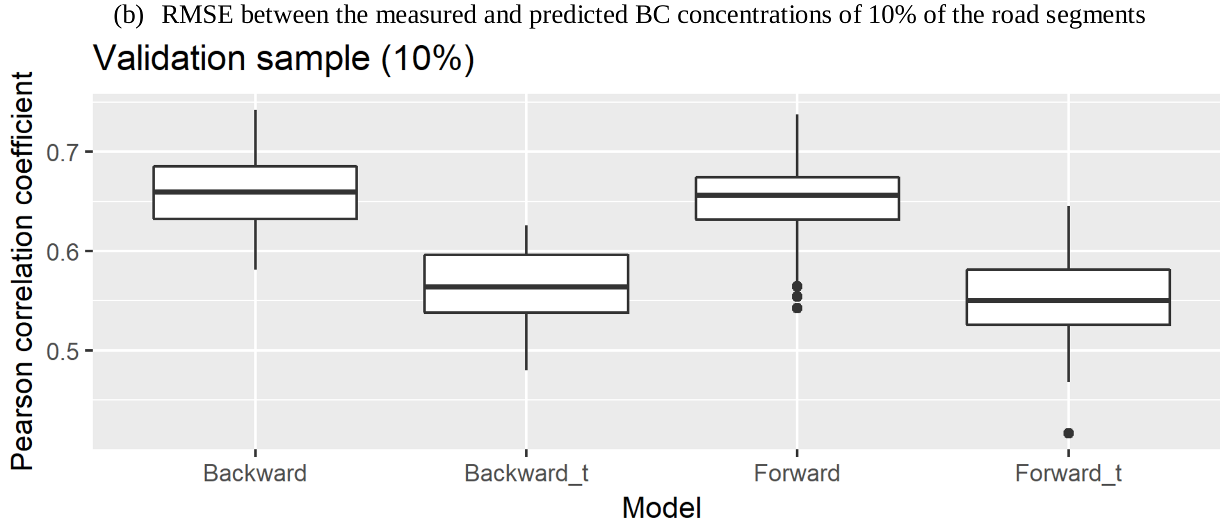
FIGURE S17 Cross validation of the BC cycling models.



(a) Adjusted R^2 of the model developed based on 90% of the road segments



(b) RMSE between the measured and predicted BC concentrations of 10% of the road segments



(c) Pearson correlation coefficient between the measured and predicted BC concentrations of 10% of the road segments

FIGURE S18 Boxplots of the 100-fold cross validation of the BC cycling models (the lower and upper whiskers correspond to the minimum and maximum no further than 1.5*IQR, the box corresponds to the interquartile range (IQR) with the middle line corresponding to the median).

The models based on the non-temporally adjusted data and including meteorological variables are all better than the others. The backward and the forward model show very similar characteristics in terms of adjusted R^2 and of the results of the cross-validation. However, less variables are included in the forward model, we therefore decide to keep this model.

FIGURE S19 presents the correlation matrix between the variables included in the forward model. The highest absolute Pearson correlation coefficient is between road_500m (length of roads within a 500m buffer) and park_500m (area of parks within 500m) and is equal to -0.440.

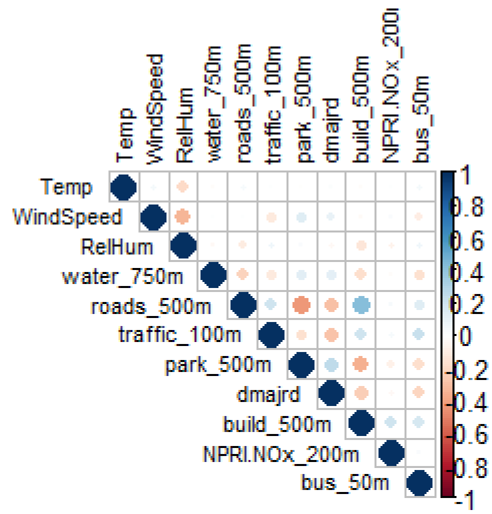


FIGURE S19 Correlation matrix between the predictors included in the cycling BC forward model.

3.2.3 Cycling models with segments within 100m of the fixed points

TABLE S16 presents the results of the forward model procedure applied to the cycling road segments within 100m around the fixed points. We had 122 road segments in the UFP model and 97 in the BC model. Variables included in these cycling models differ from those included in the fixed point models. This is not surprising as the measurements at fixed points and on the surrounding road segments were not correlated, as presented in FIGURE S6. Furthermore, although the fixed points and cycling models are in this case based on very similar number of input data (i.e. points or segments), the averages associated with them are not based on the same time scale. The fixed points are associated with average concentrations based on 100 minutes of measurement, while the road segments are associated with average concentrations based on 100 seconds of measurements. Therefore, very little value should be given to this comparison.

TABLE S16 Fixed points LUR models and cycling LUR models developed with the road segments within 100m around the fixed points

	ln(UFP)		ln(BC)	
	Fixed points	**Cycling (segments within 100m around fixed points)**	**Fixed points**	**Cycling (segments within 100m around fixed points)**
Adjusted R²	0.405	0.717	0.525	0.607
Number of points/segments	92	122	92	97
Intercept	7.08E+00***	1.180e+01 ***	3.13E+00**	8.631e+00 ***
Temperature	8.76E-02***	-5.698e-02 **	7.10E-02*	-7.834e-02 **
Relative Humidity			2.76E-02**	
Wind Speed	-2.41E-02*	-4.095e-02 ***		
Number of bus stops within 500m				3.433e+04 ***
Number of bus stops within 100m			1.62E+03*	
Number of bus stops within 300m	4.96E+03*			
Length of busline within 50m				4.656e+00 ***
Length of busline within 100m			5.00E+00*	
Distance to the CBD	2.57E-05**			
Distance to the closest airport		-4.365e-05 ***		
Length of highway within 1000m	1.01E+02*	5.527e+01 **		
Number of intersections within 300m				4.253e+03 **
Number of intersections within 750m	7.36E+03**			
Number of intersections within 1000m		4.036e+03 *		
Length of major roads within 200m		5.335e+01 ***		
Length of major roads within 500m	8.43E+01*			
Open area within 1000m	-1.49E+00*			
Population within 500m	4.34E+01.		9.32E+01***	
Water area within 50m	6.28E-01.	-9.052e-01 ***		
Water area within 750m				-1.256e+00 **
Commercial area within 500m			9.87E-01*	
Traffic within 750m			4.55E+01***	
Number of trees within 1000m		7.303e+02 ***		
Building area within 100m				-1.446e+00 ***
Area of park within 750m				-2.154e+00 ***
Governmental area within 1000m		-1.019e+00 **		

Significance codes: 0 ‘***’, 0.001 ‘**’,0.01 ‘*’, 0.05 ‘.’

3.3 Analysis with median concentrations

In this section, we considered the median concentrations for the road segments and fixed points instead of the average concentrations. The aim of this analysis was to investigate if this would have an impact on the levels predicted with the LUR models. Currently, the cycling models predict higher concentrations than the fixed point models.

3.3.1 Descriptive statistics

TABLE S17 presents a comparison of the minimum and maximum values obtained when computing the average and the median concentrations across the visits by road segment or fixed point. Median concentrations based on the cycling data, and especially in the case of UFP, are higher than median concentrations at fixed points. Therefore, considering median concentrations would not change the fact that segments with high UFP and BC values measured on bikes were higher than those measured at fixed points.

TABLE S17 Comparison of the extreme values obtained when computing the average and median concentrations (across visits) by road segment or fixed point

(a) UFP measurements

	Cycling data		Fixed point data	
	Average across visits	**Median across visits**	**Average across visits**	**Median across visits**
Minimum	2,563	598	8,461	3,664
Maximum	280,740	120,797	61,255	48,624

(b) BC measurements

	Cycling data		Fixed point data	
	Average across visits	**Median across visits**	**Average across visits**	**Median across visits**
Minimum	93	15	309	83
Maximum	37,281	33,836	7,268	3,725

3.3.2 LUR models

We ran the forward procedure based on median concentrations instead of average concentrations by road segment or fixed points. TABLE S18 summarizes the descriptive statistics of the predictions of concentrations for the city of Toronto (i.e. for all grid cells) with all models. The mean and median concentrations for the city predicted with the "median cycling models" are lower than those predicted with the "average cycling models", but it is also the case for the fixed point models.

It is true that considering median concentrations instead of average concentrations results in general in lower predicted concentrations, but this is the case for both cycling and fixed point models. This means that even if we had considered median concentrations instead of average concentrations, we would still have observed higher concentrations predicted by the cycling models.

TABLE S18 Descriptive statistics of the concentrations for the City of Toronto predicted with the models based on average and median concentrations by road segment and fixed point

(a) UFP concentrations

	Cycling model		Fixed point model	
	Average	**Median**	**Average**	**Median**
Minimum	6,931	6,141	3,555	4,356
Median	21,601	14,819	12,755	9,479
Mean	21,376	14,724	13,153	9,818
Maximum	62,887	44,645	44,740	44,337

(b) BC concentrations

	Cycling model		Fixed point model	
	Average	**Median**	**Average**	**Median**
Minimum	43	222	563	436
Median	1,058	871	764	665
Mean	1,078	913	857	793
Maximum	7,557	16,027	4,196	5,991

4 COMPARISON OF THE EXPOSURES

4.1 Panel study

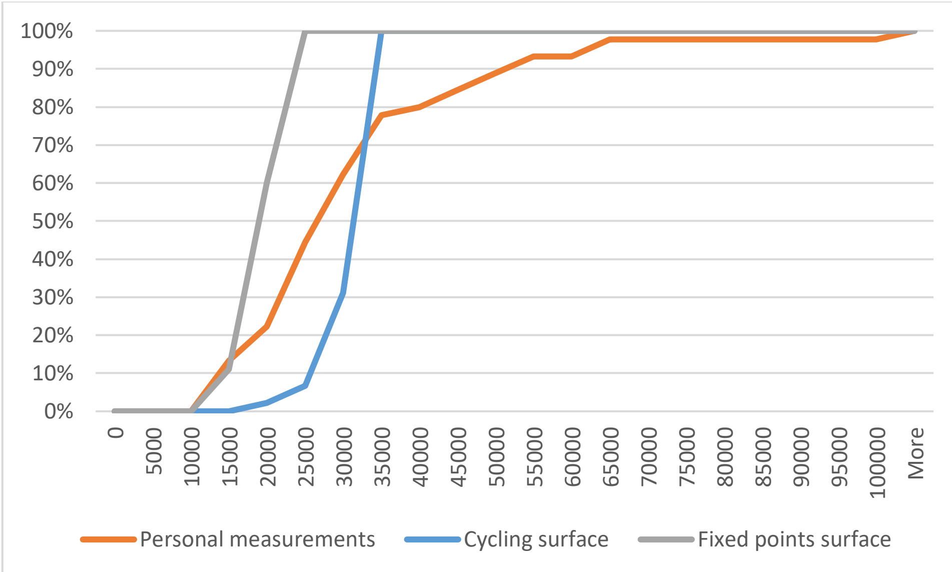
TABLE S19 presents the descriptive statistics of the time spent outdoors by the participants of the panel study. They were asked to spend at least 2 hours outdoors, but we saw that on average they spent more than 4 hours.

TABLE S19 Descriptive statistics of the time spent outdoor by the participants of the panel study

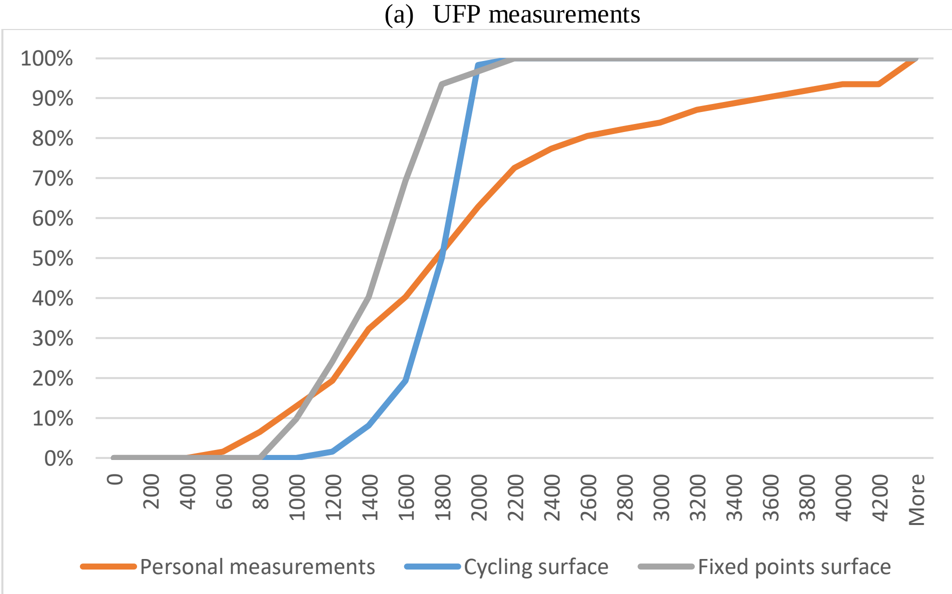
	Time spent outdoor (hour)
Minimum	1.4
Average	4.6
Median	4.8
Maximum	6.5

4.2 Personal measurements and mobility-based exposures

FIGURE S20 and FIGURE S21 present additional statistics for the comparison of measured outdoor exposures and estimated (mobility-based) exposures for the participants of the panel study.

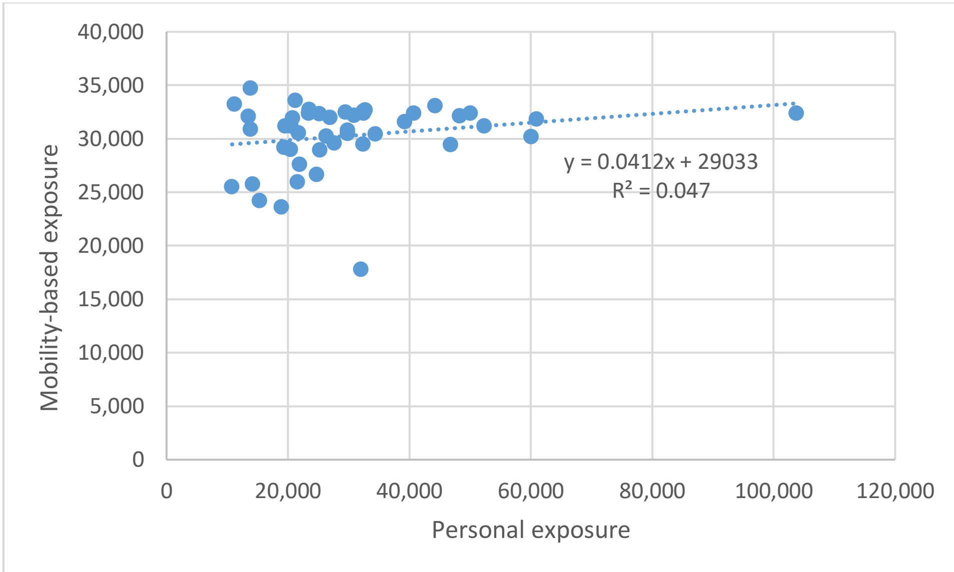


(a) UFP measurements

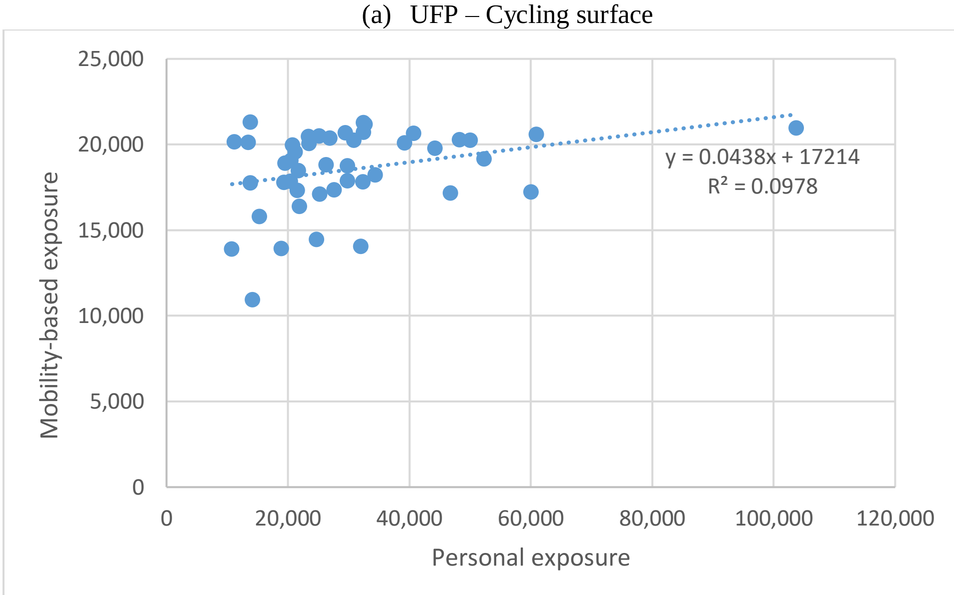


(b) BC measurements

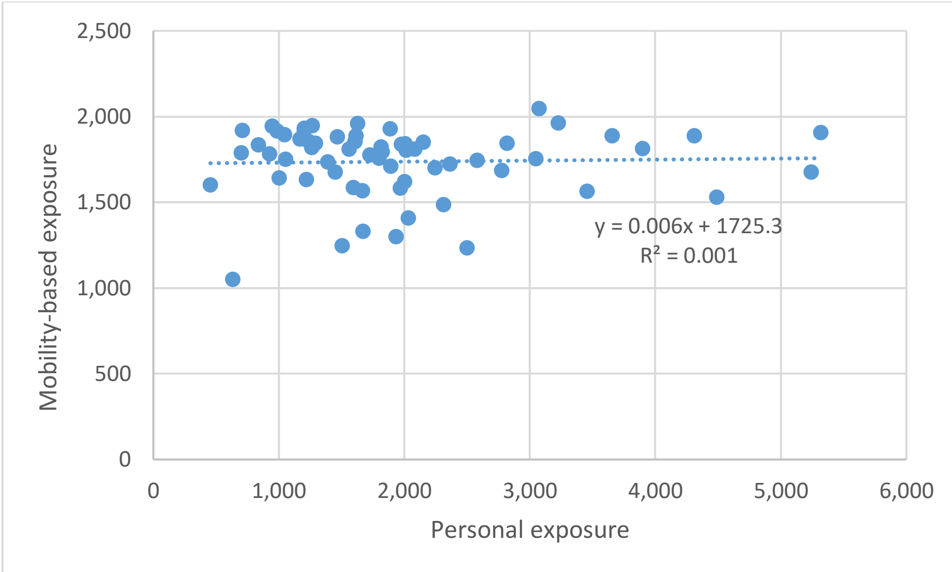
FIGURE S20 Cumulative distributions of the measured and estimated exposures of the participants of the panel study.



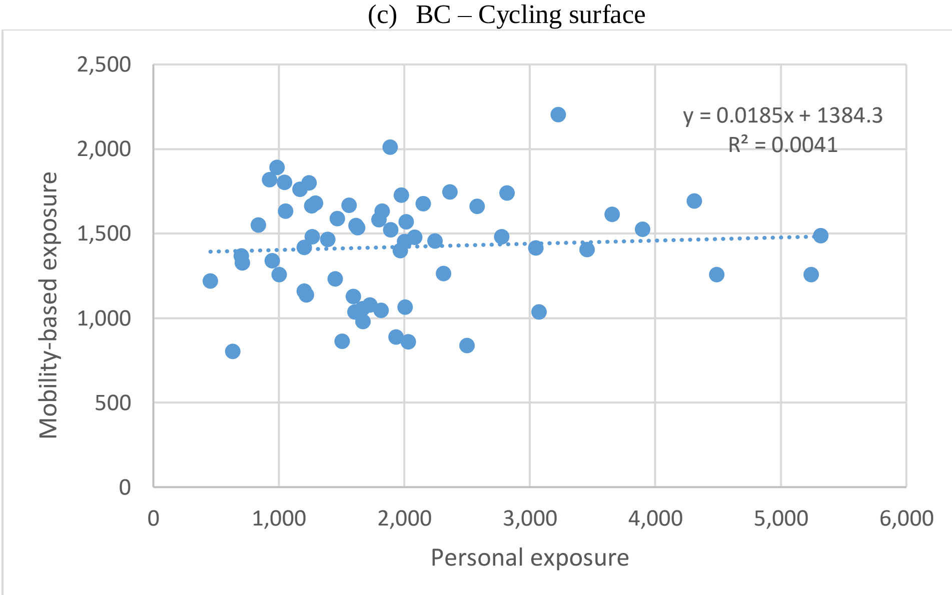
(a) UFP – Cycling surface



(b) UFP – Fixed points surface



(c) BC – Cycling surface



(d) BC – Fixed points surface

FIGURE S21 Scatterplots of the measured and predicted (mobility-based) exposures of the participants of the panel study.

5 DISCUSSION

TABLE S20 and TABLE S21 summarise the characteristics of various UFP and BC short-term monitoring campaigns and LUR models published in the recent literature.

TABLE S20 Summary of land-use regression (LUR) model protocols for Ultrafine Particles (UFP)

Study	**Location**	**Type of data collection**	**Number of segments/points sampled**	**Time spent per point/segment and/or number of visits**	**R^2 of the LUR model (* means adjusted R^2)**
Hankey and Marshall [2]	Minneapolis (U.S.A.)	Mobile (bike)	1,101 aggregation locations (spatial resolution: 100m, temporal resolution: 1s)	200 seconds (afternoon), and less than 100 seconds (morning) on average	0.50 (morning) and 0.48 (afternoon)
Sabaliauskas et al. [3]	Toronto (Canada)	Mobile (pedestrian)	112 road segments	5 to 10 minutes	0.72
Patton et al. [4]	Boston (U.S.A.)	Mobile (car)	Each one-second measurement was kept	1 second	0.23 to 0.42 (depending on neighbourhood considered)
Kerckhoffs et al. [5]	Amsterdam and Rotterdam (Netherlands)	Mobile (electric car)	2,964 road segments (745 visited twice)	18 seconds on average	0.13 (all segments) 0.18 (segments visited twice)
Farrell et al. [6]	Montreal (Canada)	Mobile (bike)	4,058 road segments	Between 1 and 52 visits	0.3812
Weichenthal et al. [7]	Montreal (Canada)	Mobile (bike in summer, cars in winter)	414 road segments	405 seconds on average (always more than 200)	0.62
Weichenthal et al. [8]	Toronto (Canada)	Mobile (car)	405 road segments	10 minutes on average (always more than 250 seconds)	0.67*
Rivera et al. [9]	Girona and close cities (Spain)	Fixed	644 fixed sites	15 minutes	0.36*
Saraswat et al. [10]	New Delhi (India)	Fixed	18 (morning) 37 (afternoon)	More than 1h	0.28 (morning) 0.23 (afternoon)
Ghassoun et al. [11]	Braunschweig (Germany)	Fixed	27 fixed points	45 minutes	0.74 (summer) 0.85 (winter)
Montagne et al. [12]	Amsterdam and Rotterdam (Netherlands)	Fixed	161 sites	90 minutes	0.37
Kerckhoffs et al. [5]	Amsterdam and Rotterdam (Netherlands)	Fixed	128 fixed sites	60 minutes	0.36
van Nunen et al. [1]>	Basel (Switzerland), Heraklion (Greece), Amsterdam, Maastricht, and Utrecht (“The Netherlands”), Norwich (United Kingdom), Sabadell (Spain), and Turin (Italy)	Fixed	160 in general, 240 sites for “The Netherlands”	90 minutes	0.28 to 0.48

TABLE S21 Summary of land-use regression (LUR) model protocols for Black Carbon (BC)

Study	**Location**	**Type of data collection**	**Number of segments/points sampled**	**Time spent per point/segment and/or number of visits**	**R^2 of the LUR model (* means adjusted R^2)**
Hankey and Marshall [2]	Minneapolis (U.S.A.)	Mobile (bike)	1,101 aggregation locations (spatial resolution: 100m, temporal resolution: 60s)	Less than 100 seconds on average	0.29 (morning) 0.43 (afternoon)
Larson et al. [1]>	Vancouver (Canada)	Mobile (gasoline car)	39 locations	5 to 13 minutes	0.68
Kerckhoffs et al. [5]	Amsterdam and Rotterdam (Netherlands)	Mobile (electric car)	2,336 segments (745 visited twice)	18 seconds on average	0.12 (all segments) 0.30 (segments visited twice)
Kerckhoffs et al. [5]	Amsterdam and Rotterdam (Netherlands)	Fixed	141 fixed sites	60 minutes	0.28
Montagne et al. [12]	Amsterdam and Rotterdam (Netherlands)	Fixed	161 sites	90 minutes	0.35
Saraswat et al. [10]	New Delhi (India)	Fixed	17 (morning) 25 (afternoon)	More than 1h	0.86 (morning) 0.69 (afternoon)

6 REFERENCES

(1) City of Toronto. Traffic Signal Vehicle and Pedestrian Volumes https://www1.toronto.ca/wps/portal/contentonly?vgnextoid=417aed3c99cc7310VgnVCM1000003dd60f89RCRD&vgnextchannel=1a66e03bb8d1e310VgnVCM10000071d60f89RCRD (accessed Dec 14, 2017).

(2) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194–9202.

(3) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a land-use regression model for ultrafine particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92.

(4) Patton, A. P.; Zamore, W.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Transferability and generalizability of regression models of ultrafine particles in urban neighborhoods in the boston area. *Environ. Sci. Technol.* **2015**, *49* (10), 6051–6060.

(5) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particle and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902.

(6) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near roadway air pollution across a spatially extensive road and cycling network.

Environ. Pollut. **2016**, *212*, 498–507.

(7) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A land use regression model for ambient ultrafine particles in Montreal, Canada: A comparison of linear regression and a machine learning approach. *Environ. Res.* **2016**, *146*, 65–72.

(8) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ. Pollut.* **2016**, *208*, 241–248.

(9) Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmos. Environ.* **2012**, *54*, 657–666.

(10) Saraswat, A.; Apte, J. S.; Kandlikar, M.; Brauer, M.; Henderson, S. B.; Marshall, J. D. Spatiotemporal Land Use Regression Models of Fine , Ultra fi ne , and Black Carbon Particulate Matter in New Delhi , India. *Environ. Sci. Technol.* **2013**, *47* (22), 12903–12911.

(11) Ghassoun, Y.; Ruths, M.; Löwner, M. O.; Weber, S. Intra-urban variation of ultrafine particles as evaluated by process related land use and pollutant driven regression modelling. *Sci. Total Environ.* **2015**, *536*, 150–160.

(12) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720.

(13) van Nunen, E.; Vermeulen, R.; Tsai, M.-Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; et al. Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ. Sci. Technol.* **2017**, *51* (6), 3336–3345.

(14) Larson, T.; Henderson, S. B.; Brauer, M. Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678.