On the spatial representativeness of NOX and PM10 monitoring-sites in Paris, France

Delphy Rodriguez, Myrto Valari, Sébastien Payan, Laurence Eymard

Atmospheric Environment: X 1 (2019) 100010

Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

On the spatial representativeness of NO_X and PM_{10} monitoring-sites in Paris, France

Delphy Rodriguez[a,*], Myrto Valari[b], Sébastien Payan[a], Laurence Eymard[a]

[a] LATMOS/IPSL, Sorbonne Université, UVSQ,CNRS, Paris, France
[b] LMD, Ecole Polytechnique, Palaiseau, France

HIGHLIGHTS

- A new method to quantify the spatial representativeness of urban monitoring-sites using high-resolution simulations.
- Measurements at urban background sites are representative of concentrations over the broad neighborhood around the monitor.
- Measurements at traffic sites are representative of concentrations over specific urban features: roads, pavements, etc.
- PM_{10} representativeness areas are larger than NO_X representativeness areas.

ARTICLE INFO

Keywords:
Urban air pollution
Surface monitoring networks
High resolution modeling
Spatial representativeness

ABSTRACT

Ambient pollutant concentrations in Paris, France are routinely measured by the local surface-monitor network of the AIRPARIF agency. Such networks, however dense, have a limited spatial representativeness around the monitoring-site and are not capable to represent the strong horizontal gradients of pollutant concentrations over urban areas. High resolution models simulate 3D pollutant concentration fields at a spatial resolution as fine as a few meters over the urban area by integrating the underlying emission sources and accounting for the effect of buildings on the dispersion patterns. These models, provide a good spatial variability over the urban area but suffer from uncertainties related to the emission inventories, meteorological fields and parametrizations of the physical and chemical processes.

In this paper, simulations conducted by ARIA Technologies with the Parallel Micro-Swift-Spray (PMSS) model (http://www.aria.fr/projets/aircity) are used to assess NO_X and PM_{10} representativeness areas around five urban background and five traffic-oriented monitoring-sites of the AIRPARIF network during ten days in March 2016. Commonly, the spatial representativeness of a monitor site is defined through homogeneity areas, namely the area around a monitoring-site where pollutant concentration is above 20% of the concentration at the location of the monitoring-site. Here, we propose a novel approach that uses similarity areas to define the spatial representation of monitor sites. Similarity areas integrate points that respect the additional condition to be highly correlated in time with the concentration at the monitor station. Thus, the criterion to select similarity areas is a combination of a high value of the correlation coefficient and a small value of the normalized root mean square error with regards to the concentration at the grid-cell corresponding to the location of the monitor. Criteria thresholds are determined through an iterative analysis and a representative area is defined through image processing that selects all the connected pixels that satisfy criteria thresholds and incorporate the grid-cell of the monitor.

Daily similarity areas estimated around each monitor are compared against homogeneity areas with regards to their shape, spatial extent, and urban specific characterization. Around urban background sites they are of the same order of magnitude, whereas around traffic sites similarity areas are generally larger than homogeneity areas. PM_{10} representativeness areas are found to be 2.2 times larger than the NO_X ones. Urban background areas are representative of the broad neighborhood around the monitoring-site, whereas traffic-oriented monitoring-sites are representative of specific urban features such as sections of roads and sidewalks along the road. Averaged over the 10 days of the study and across all monitoring-sites, representativeness areas for urban background monitoring-sites are about 8 times larger than traffic representativeness areas (0.6 km^2 vs. 0.07 km^2).

* Corresponding author.
E-mail address: delphy.rodriguez@latmos.ipsl.fr (D. Rodriguez).

https://doi.org/10.1016/j.aeaoa.2019.100010
Received 20 January 2019; Accepted 20 January 2019
Available online 23 January 2019

1. Introduction

The latest studies on the health impact of atmospheric pollution have pointed out that traffic-related atmospheric pollution is particularly dangerous for children asthma and cardiovascular diseases (APHEKHOM, 2012; Host, 2013). The health impacts due to exposure to particulate matter (PM), ozone (O_3) and nitrogen dioxide (NO_2) are presented in the technical report of the World Health Organization (WHO) « Review of Evidence on Health Aspects of Air pollution » (REVIHAAP Project, 2013). The International Agency for Research on Cancer (IARC), a specialized cancer agency of the WHO, has classified the outdoors atmospheric pollution as « carcinogenic to humans » (IARC gr.I, 2013). The annual financial and economic cost of the atmospheric pollution in France according to the Senate in July 2015 (Senate Report n°610, 2015) is 101.3 billion €. This amount is not only due to health impacts but also due to hidden costs such as the decrease of crop yield, deterioration of the buildings, etc.

This paper focuses on Nitrogen Oxides ($NO_X = NO_2 + NO$) and particulate matter with diameter lower than 10 μm (PM_{10}). NO_X are mostly emitted by fossil fuel combustion, industrial processes, and can be produced by chemical reactions involving NO and O_3. PM_{10} compounds are characterized of a large variability of chemical composition, solid or liquid state and size distribution. PM_{10} are directly emitted by urban heating, fossil fuel combustion, industrial and agriculture processes (CITEPA, French emissions inventory agency: https://www.citepa.org). Other, indirect sources of PM_{10} are chemical reactions, large-scale transport of dust and sea-salt and the resuspension of particles already deposited on the ground due to the wind blowing or traffic circulation along the streets. In this paper, we will focus on the local emission sources to provide guidelines for local policies.

The highly heterogeneous emissions from various sources (traffic, residential, industrial …) and the presence of buildings affect the dispersion of pollutants in the urban environment resulting to a high spatial and temporal variability of pollutant concentrations (Harrison, 2018). In Paris, France, ambient pollutant concentrations are routinely measured by the local network of the AIRPARIF agency (https://www.airparif.asso.fr). In the inner Paris area, this network has 15 and 8 permanent surface monitoring-sites, for NO_X and PM_{10} respectively. The spatial representativeness is limited around each monitoring-site and therefore, such networks are not able to capture the strong horizontal gradients of pollutant concentrations over the urban area (Vardoulakis et al., 2005; Hagenbjörk et al., 2017; Hagemann et al., 2014; Blanchard et al., 2014).

Pollutant concentration fields are simulated with regional-scale chemistry transport models such as CHIMERE (http://www.lmd.polytechnique.fr/chimere) under air-quality forecasting platforms (e.g. Prev'Air http:/www.prevair.org, or Esmeralda http://www.esmeralda-web.fr) or research projects. Such models integrate emission fluxes, the horizontal and vertical transport of pollutants, turbulent mixing, chemical reactions and deposition through complex numerical systems to calculate the three-dimensional atmospheric concentration fields. However, these models have insufficient horizontal resolution (maximum of 1 km × 1 km) to resolve the intra-urban variability of the pollutant concentration.

High-resolution models such as the Parallel Micro-Swift-Spray (PMSS) model integrated in local scale air-quality modeling platforms such ARIA City (http://www.aria.fr/aria_city.php) simulate pollutant concentration fields at a spatial resolution of a few meters over the urban area. They take into account the advection from various emission sources, integrate buildings as obstacles to the wind flow by generating vortexes downwind and upwind their location, calculate wet and dry deposition and in some cases deal with simple chemical reactions such as the Chapman cycle for ozone and NOx interactions. Given that the street model used in this study implements a simple chemical mechanism we chose to focus on the less reactive NO_X pollutant. However, over the highly urbanized area of the study and close to traffic sources, NO_X is a very good proxy of NO_2 since these pollutants are strongly correlated one with the other and with the traffic. Thus, chemistry is not expected to alter our results.

These models provide a good spatial variability within the urban environment but are still subjected to important biases due to the high uncertainties relating to the emission inventories, meteorological predictions and simplified chemistry.

Spatial representativeness assessment aims generally to improve the characterization of atmospheric pollution simulations by using surface measurements to test model performance, design and optimize monitoring-site deployment, better estimate population exposure to the atmospheric pollution, assess or report air quality, classify station, assimilate data for modeling.

In this paper, we focus on the first task, namely to define the spatial extent of the measurement around the monitoring-site. The European Commission stresses the need to assess the spatial representativeness around monitoring-sites without imposing a specific methodology (EC, 2011). An intercomparison exercise has been realized by FAIRMODE (Forum for Air Quality Modeling in Europe) and AQUILA (Air Quality Reference Laboratories) to harmonize a large set of 10 spatial representativeness approaches and provide a technical support to the implementation of Directive 2008/50/EC (EC, 2008) on ambient air quality and cleaner air for Europe. Study results obtained from the same common dataset reveal a strong variability between the different approaches (extent, shape, technical procedures). Most of these approaches use a common criterion to select representativeness areas: the pollutant concentration at each point of the representativeness area should not differ from the measurement at the monitoring-site location by more than a given threshold. Additionally, some approaches use also external parameters influencing air quality such emissions from different sources and/or climatic and topographic dispersion. Nguyen et al. (2012) calculate the means of Principal Component Analysis (PCA) to classify status station.

A brief overview of the most commonly used approaches found in the literature is presented here. These studies use a homogeneity criterion based on a certain threshold. Santiago et al. (2013) apply a methodology based on annual average simulated concentration maps obtained by means of weighted average of Computational Fluid Dynamics model for urban stations whereas Martin et al. (2014), use a Chemistry Transport Model (CHIMERE) for rural background stations. The applied threshold is ± 20% of the simulated concentration to the monitoring-site location. Vitali et al. (2016); Piersanti et al. (2015) used the Concentration Similarity Frequency (CSF) to assess representativeness areas from simulated concentrations. CSF approach is based on the definition given by Nappo et al. (1982): "a point measurement is representative of the average in a larger area (or volume) if the probability that the squared difference between point and area (volume) measurement is smaller than a certain threshold more than 90% of the time". Representativeness area contains a grid-cell if the difference between concentrations at the monitoring-site and the selected grid-cell does not exceed 20%, 90% of the time. Spangl et al. (2007) compared modeled annual averaged concentration to the concentration to the monitoring-site location with similarity thresholds depending on pollutant. Additional criteria emissions are applied.

Beauchamp et al., (2016), Bobbia et al. (2008), assesses spatial representativeness areas on annual averages by estimating spatial concentration variability and uncertainties from passive sample measurements. Then those concentrations are interpolated from modeling output data applying a kriging approach. The applied threshold to select spatial representativeness points is 30%.

Janssen et al. (2012) assess representativeness of stations by calculating the relation between annual averaged levels at the location of the station and a land use characteristics.

These approaches require long datasets, much longer than the 10 days available in this study. Also, such homogeneity criteria provide an average estimate, not taking into account the daily temporal variability

due to specific meteorological conditions or traffic emissions which may involve pollution episodes. Moreover they are generally focusing on a small number of monitoring-sites. Another limitation of these approaches is that they do not filter out the variability due to regional scale transport (especially high for PM_{10}) and the long time averaged. They, therefore, fail to highlight the impact of the urban features (emissions, urban landscape) on the characterization of the representativeness area.

In this study, daily representativeness areas are derived from PMSS simulations with at 3 m horizontal resolution. The study focuses on NO_X and PM_{10} around five traffic and five urban-background monitoring-sites of the AIRPARIF network in Paris for ten non-successive days in March 2016.

Modeled and measured data and methodological aspects are presented in Section 2. The results of the study are presented in Section 3. Section 4 consists of a discussion on the general conclusions and future perspectives of the study.

2. Data and methodology

2.1. Monitoring-sites

Five traffic-oriented and five urban background sites have been selected for the study, covering a diversity of urban landscapes in Paris. Traffic monitoring-sites are located on sidewalks along roads, major crossroads or the border of the Seine River. Urban background monitoring-sites are located on low emission areas such a parks, pedestrian squares, schoolyards and leisure parks. NO_X are measured at all monitoring-sites whereas PM_{10} are measured at four traffic stations and two urban background stations (Table 1 and Fig. 1). Hourly averaged NO_X and PM_{10} surface concentrations are routinely measured and available on the website of the AIRPARIF air-quality agency (http://www.airparif.fr).

Table 1
Characteristic features of the study-areas. Locations marked with an asterisk indicate the location of the monitoring-site.

Type	ID station	Characteristic features	Pollutants
Traffic-oriented stations	CELES	Seine banks*, the Seine River, bridge, side roads	NO_X
	HAUS	Bld Haussmann*, side roads	NO_X, PM_{10}
	OPERA	Crossroad*, road traffic	NO_X, PM_{10}
	BASCH		
	ELYS	Champs-Élysées Avenue*: large road bordered with trees, side roads	NO_X, PM_{10}
Urban background stations	PA15L	Stade* and leisure park, “Boluevard Peripherique”, traffic road	NO_X, PM_{10}
	PA12	Schoolyard*, railways, traffic road	NO_X
	PA13	Park*, traffic road, crossroad	NOX
	PA04C	Place*, park, the Seine River, traffic road	NO_X, PM_{10}
	PA07	Square close to the Eiffel Tower*, Champ-de-Mars Garden, Seine banks, the Seine River, Bridge	NO_X

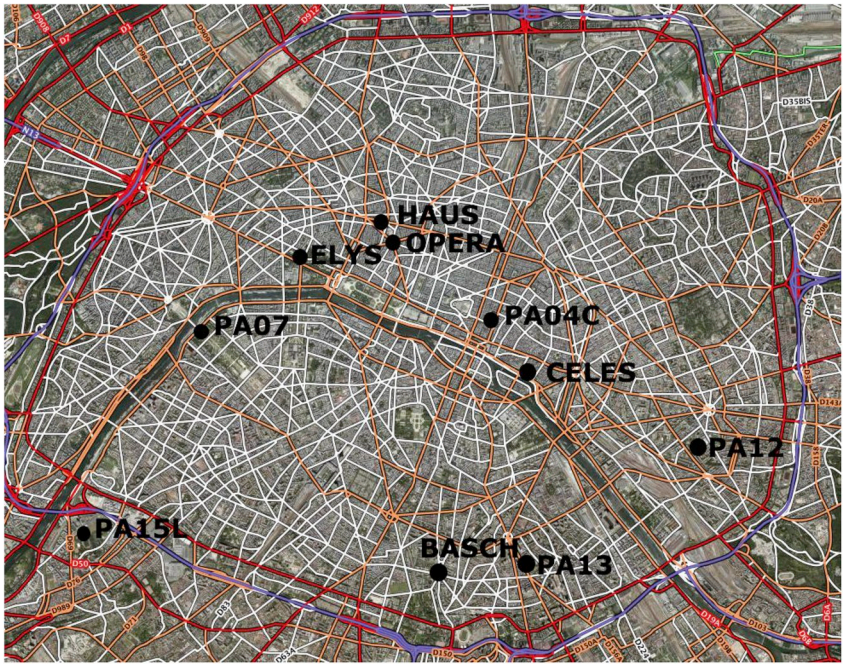

Fig. 1. Paris map with locations of studied Airparif stations, the Seine River and roads. The red thin line represents Paris city boundary “Boulevard Periphérique” and the white and orange line represents roads.

2.2. Model simulations

NOx and PM_{10} concentrations are simulated with the local-scale model PMSS, at a horizontal resolution of 3 m covering a $12 \times 10\ km^2$ domain over the city of Paris.

The model has thirty-five vertical layers from the ground up to an altitude of 800 m. The vertical resolution decreases with height with a surface layer of 2 m thick (Moussafir et al., 2015). Emissions fluxes are taken from the AIRPARIF inventory. Meteorological conditions are simulated with the MM5 weather mesoscale model. Terrain elevation and buildings' contours and height are taken from the BDTOPO database developed by the French National Geographic Institute. Atmospheric dispersion is simulated with the PSPRAY lagrangian particles model. Traffic emissions are calculated with the HEAVEN chain (Healthier Environment through the Abatement of Vehicle Emissions and Noise, http://www.airparif.asso.fr/etat-air/air-et-climat-emissions-heaven). HEAVEN chain calculates in near real-time traffic situation and allows obtaining pollutant emissions on main traffic road of Parisian area from a "bottom-up" approach. Hourly traffic emissions are calculated from the combination of traffic model (car flow, mean speed, cold engines percentage) and emission factors (car speed, road type, temperature, car type, etc.) applied by loop-based counting systems.

PMSS simulations are available for ten non-successive days in March 2016. Domain-averaged hourly variations of NO_X and PM_{10} simulated concentrations are shown in Fig. 2. A peak in NO_X concentration is observed on the 9th of March. A PM_{10} pollution episode is simulated from the 11th to the 15th of March 2016 reaching a peak of 60 $\mu g/m^3$ on the 12th of March. Another [PM_{10}] peak is observed on March 18.

Table 2 summarizes daily and domain-averaged NO_X and PM_{10} simulated concentrations, as well as mean daily surface pressure from the recent ECMWF model reanalyses ERA 5 (https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era5), wind direction and daily averaged velocity (superior to $3\ m\ s^{-1}$) simulated by the MM5 regional mesoscale weather forecasts model (http://www2.mmm.ucar.edu/mm5/). Wind patterns on these ten days change from West to North-East, accompanying the transition starting from a low-pressure system between 03-02 and 03-10 (994 hPa-1003 hPa) followed by a transition day 03-11-16 (1012 hPa), then the high-pressure situation took place on the 03-12-16 (1015 hPa). The beginning of a high-pressure situation is characterized by a strong atmospheric stability combined to a pollution episode due to agriculture activity from wide European area, as far as Eastern Europe, leading to an accumulation of PM_{10} within the boundary layer. A sharp rise in PM_{10} simulated concentration is observed on the 11th of March (more than twofold). These atmospheric conditions do not affect NO_X concentration, which is more driven by local traffic emissions. However, both PM_{10} and NO_X concentrations are sensitive to rainfall which occurs on the 2nd and 5th of March and results in pollutant deposition on the ground. During these days, NO_X concentration reaches its lowest levels.

NO_X concentration levels mostly depend on the density of the traffic network, with higher pollution levels during weekdays than during week-ends or holidays. Saturday the 5th of March, was a return-from-holiday and one would expect a peak in NO_X concentrations. However, this was attenuated by rainfall.

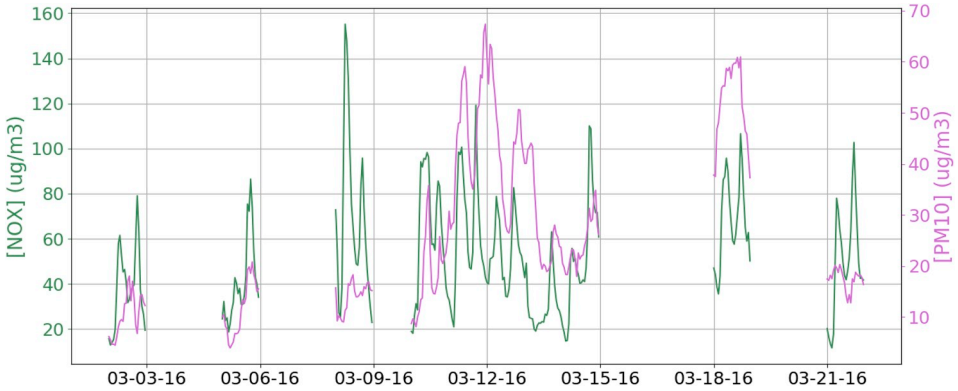

Fig. 2. Time-series of NO_X and PM_{10} simulated concentrations ($\mu g/m^3$) in Paris for 10 days in March 2016 (hourly data).

Table 2
The first column lists the days of March 2016, H designs Holiday. Rainy days are marked by an asterisk. In the second and the third columns are given the daily average simulated concentrations in Paris for NO_X and PM_{10}. The fourth column shows the surface pressure (hPa) from ERA reanalysis. The fifth column summarizes the mean wind direction and the range of daily-mean wind velocity modeled with the MM5 model.

March 2016	[NOX] (µg/m3)	[PM10] (µg/m3)	Surface Pressure (hPa)	Wind Direction [Daily average velocity (m.s-1)]
02 Wednesday* *H*	38.2	10.5	994	W [5.4–5.8]
05 Saturday* *H*	41.0	11.6	983	NNW [3.2–3.3]
08 Tuesday	68.8	14.0	1003	–
10 Thursday	62.2	**18.3**	1001	–
11 Friday	63.5	**46.2**	**1012**	–
12 Saturday	54.5	44.3	1015	–
13 Sunday	31.1	28.5	1016	NE [3.9–4]
14 Monday	52.1	24.2	1014	NE [4–4.3]
18 Friday	68.0	52.1	1010	–
21 Monday	50.4	17.4	1008	–

2.3. Model evaluation against local surface observations

PMSS simulations at the surface model layer are compared to measurements of the AIRPARIF network. A thorough work has been carried out to assign model grid-cells to each monitoring-site combining google street view and plans of the French National Geographic Institute. An error in the grid-cell selection would strongly affect the results at such high resolution (Duyzer et al., 2015). Two AIRPARIF stations have been omitted for the model evaluation process due to missing data in the emission inventory along street segments in the vicinity of these stations (BASCH and CELES).

PMSS simulations are in a good agreement with PM_{10} measurements as shows the squared correlation coefficient (R^2), equal to 0.83 and 0.92, and the root mean squared error (RMSE), equal to 9.57 $\mu g/m^3$ and 6.21 $\mu g/m^3$ for traffic and urban background sites respectively (Fig. 3). Model performance is not so satisfactory for NO_X measurements with a R^2 equal to 0.65 and 0.45 for traffic and urban background sites respectively. As shown in the scatter plots the model overestimates NO_X

maxima, especially in traffic sites, resulting in large overall error (RMSE equal to 81.57 µg/m^3). Moreover, PMSS simulations fail to reproduce some nocturnal peaks of the NO_X concentration. On the contrary, PMSS underestimates NO_X concentrations at urban background sites. More specifically, at urban sites, PMSS simulations fail to reproduce the morning peak of NO_X concentration corresponding to the rush hour (between 7AM and 8AM). The large difference in the bias between NO_x and PM_{10} is probably due to errors in the emissions inventory and to the simple chemical mechanism of the PMSS model, both having a larger impact on NO_x concentrations than on PM_{10}.

Although the model evaluation has pointed out some limitations of the PMSS simulations, overall the model performance is good for PM_{10}. NO_X simulations are not so satisfactory for peaks due to traffic emissions input. The reason is most probably the simple chemical mechanism of the PMSS model that fails to represent the concentration of the highly reactive pollutants such as NO. Note that the error for NO_2 is smaller than for NO for traffic stations (RMSE(NO_2) = 21.9 µg/m^3 vs RMSE(NO) = 72.3 µg/m^3) and similar for urban background stations.

Despite of these model limitations, we assume that the model captures the actual spatial variability of pollution. In the following, these simulations are used to assess the spatial representativeness of each AIRPARIF station during the 10 days of March 2016.

2.4. The spatial representativeness approach

The spatial representativeness area of a monitoring-site consists of all the points around the station where the concentration is linked to the concentration at the monitoring-site location, through a given criterion. Here we develop an alternative approach that focuses on the daily variability through the definition of "similarity" area. It aims at being a complementary method, with respect to those used to evaluate air quality directive limits.

This area integrates points for which the daily temporal variability is highly correlated to the temporal variability of the concentration at the reference point (monitor), and the daily bias with respect to the concentration at the monitor location, remains below a chosen threshold.

The model grid-cell corresponding to the location of the monitor station, called "pseudo-station" hereafter, is used as reference value, instead of the monitoring stations. Indeed, due to the discrepancies between observations and simulations, using actual measurement would lead to inconsistent results when studying the spatial representativeness of stations using simulations.

In two cases, (BASCH and CELES stations), the lack of emission sources along a street segment in the vicinity of the station led us to slightly shift the "pseudo-station". Those both virtual stations will be renamed "BASCH_v" and "CELES_v". CELES has been shifted by 600 m from the real monitoring-site position in the northwest direction, on the same road. BASCH has been moved by 60 m to be placed in another avenue (exit of the crossroad "Avenue Jean Moulin" instead of "Avenue du General Leclerc"). Those two transfers are made by respecting the same road and urban landscape conditions (subjective analysis). We indeed checked that there is no lack of emission sources along this new street segment in the vicinity of the virtual stations. Indeed, a further analysis with corrects simulations would be necessary for completing results around the actual station locations.

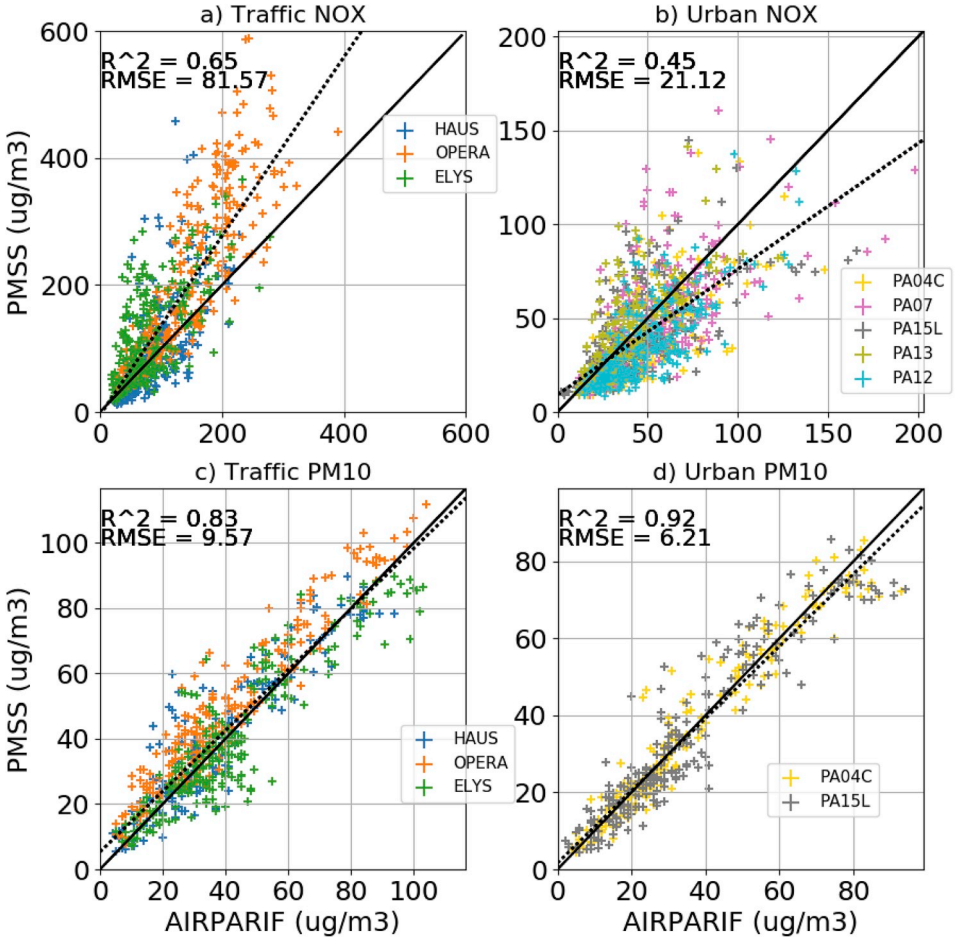

Fig. 3. Scatter plots of measured vs simulated hourly concentrations for NO_X (top) and PM_{10} (bottom) at traffic (left) and urban (right) monitoring sites. The black line represents y = x. The dotted line represents the linear regression between AIRPARIF measurements and PMSS simulations.

For evaluating the relevance of the similarity approach, with comparison to homogeneity optimal thresholds must be chosen, both for the correlation coefficient and the root mean square difference. We use the 20% threshold for homogeneity, to be consistent with previous studies. Criteria thresholds are applied over squared areas of $1.2 \times 1.2\,km^2$ centered around each pseudo-station. The size of these squares was chosen to be large enough to encompass both high (traffic network and dense residential areas) and low emission zones (parks or part of the Seine River).

Similarity areas are thus obtained by the combination of a (1) high correlation coefficient R (Eq (1)) and a (2) low normalized root mean square error NRMSE (Eq (2)).

$$R(ref, val_{(x,y)}) = \frac{\sum_{t=0}^{t=23}(ref_{(t)} - < ref_t >)(val_{(t,x,y)} - < val_{(t,x,y)} >)}{\sigma_{ref}\,\sigma_{val(x,y)}} \quad (1)$$

Where, $ref_{(t)}$ is the PM_{10} or NO_X hourly concentration at the pseudo-station, $val_{(t,x,y)}$ the PM_{10} or NO_X hourly concentration at each grid-cell, $<ref_{(t)}>$ the PM_{10} or NO_X daily concentration at the pseudo-station, $<val_{(t,x,y)}>$ the PM_{10} or NO_X daily concentration at each grid-cell, σ_{ref} the PM_{10} or NO_X daily standard deviation at the pseudo-station and $\sigma_{val_{(x,y)}}$ the PM_{10} or NO_X daily standard deviation at each grid-cell.

$$NRMSE(ref, val_{(x,y)}) = \frac{\sqrt{\frac{1}{n}\left(\sum_{t=0}^{t=23}\left(ref_{(t)} - val_{(t,x,y)}\right)^2\right)}}{<ref_{(t)}>} \quad (2)$$

where $ref_{(t)}$ is the PM_{10} or NO_X hourly concentration at the pseudo-station, $val_{(t,x,y)}$ the PM_{10} or NO_X hourly concentration at each grid-cell and $<ref_{(t)}>$ PM_{10} or NO_X the daily concentration at the pseudo-station.

Homogeneity areas are determined from daily-averaged simulated concentrations applying a threshold of 20% of the pseudo-measure, as recommended by the European Commission for uncertainty assessment in the Directive 2008/50/EC (EC, 2008) (from 15% to 25% depending on the measured pollutant) and used in literature (Vitali et al., 2016; Piersanti et al., 2015; Blanchard et al., 1999; Janssen et al., 2008). This criterion will be called *EC criterion* hereafter (Eq. (3)).

$$EC\ criterion = \frac{abs(< val_{(t,x,y)} > - < ref_{(t)} >)}{<ref_{(t)}>} * 100 \quad (3)$$

Where, $<ref_{(t)}>$ is the PM_{10} or NO_X mean daily concentration at the pseudo-station and $<val_{(t,x,y)}>$ the PM_{10} or NO_X mean daily concentration at each grid-cell.

PM_{10} concentrations are characterized by a large temporal variability due to dust episodes occurring on some of the days within the ten days period of the study (Fig. 1). To reduce this regional scale effect on the statistics and highlight local effects such as urban emission sources we normalized PM_{10} data on a daily basis (Eq. (4)).

$$[\boldsymbol{PM}10]_{(t,x,y)\boldsymbol{normalized}} = \frac{[\boldsymbol{PM}10]_{(t,x,y)} - [\boldsymbol{PM}10]_{\text{min area}}}{[\boldsymbol{PM}10]_{\text{max area}} - [\boldsymbol{PM}10]_{\text{min area}}} \quad (4)$$

Where, $[\boldsymbol{PM}10]_{(t,x,y)\boldsymbol{normalized}}$ is the PM_{10} hourly normalized concentration at each grid-cell, $[\boldsymbol{PM}10]_{(t,x,y)}$ is the PM_{10} hourly concentration simulated at each grid-cell, $[\boldsymbol{PM}10]_{\text{min area}}$ and $[\boldsymbol{PM}10]_{\text{max area}}$ are the 0.1st and 99.9th percentiles of the daily averaged PM_{10} concentrations over an area of $1.2 \times 1.2\,km^2$ around the station.

The previous criteria are applied on the simulated concentration fields around each traffic station. An algorithm selects the connected pixels that satisfy the aforementioned criteria and incorporate the « pseudo-station » pixel. Two pixels are considered "connected" when they are neighbors and they respect the threshold criteria. We thus focus on the representativeness area strongly connected to the station. In some cases with moderated traffic, both sidewalks are selected. However, traffic impact is not always symmetric over both sidewalks bordering the traffic road. It is difficult to fully assess this result, since it relies on PMSS simulations.

As an example, Fig. 4 shows the correlation coefficient (R) and NRMSE values estimated at each model grid-cell over an area of $1.2 \times 1.2\,km^2$ around the around Champs-Élysées Avenue monitoring-site (ELYS) on the 2nd of March 2016. The highest R values (purple color) are obtained in the vicinity of the station, along the avenue, and along the Seine bank. The minimal NRMSE values (from dark blue to green) are observed along a line inside the Champs-Elysées Avenue. The two criteria are thus complementary and their combination results in the definition of an area inside which concentration levels remain close to the concentration at the location of the station.

To choose thresholds for ρ and NRMSE, we consider two conditions: (i) that the size of the defined area lies within a reasonable range and (ii) that such an area is obtained most of the days of the study.

The European Air Quality Directive (EC/2008/50) specifies that «a sampling point must be sited in such a way that the air sampled is representative of air quality for a street segment no less than 100 m in length at traffic-oriented sites». Indeed, the selected area must be large enough to get beyond the pseudo-station. Here we fix the lower limit at $300\,m^2$ and the upper limit at $400 \times 10^3\,m^2$ to avoid incorporating the entire domain centered around the monitoring-site (squares of $1.2 \times 1.2\,km^2$). For example, on the 2nd of March 2016, around the monitoring-site located on the avenue des Champs-Elysées, the selected similarity area with R coefficient equal to 0.7 and NRMSE value of 0.45 is considered as a success with a selected area of $8.9 \times 10^3\,m^2$, superior to $300\,m^2$ and inferior to $400 \times 10^3\,m^2$ (Fig. 5).

Criteria thresholds to select similarity areas have been determined by an iterative analysis from all traffic-oriented monitoring-site areas over the 10-days with a step of 0.05 for both correlation and NRMSE values. This analysis is carried out on traffic-sites, because between road and sidewalk there is the highest degree of variability so we can check if the retained thresholds allow the distinction between the road and sidewalk and evaluate the impact on distance to the road. Then the same threshold values are applied on the urban-background sites.

For each threshold, a score is estimated by counting the number of days on which the size of the selected area ranges within the chosen boundaries. The maximum score for each site is 10. Finally, a total score

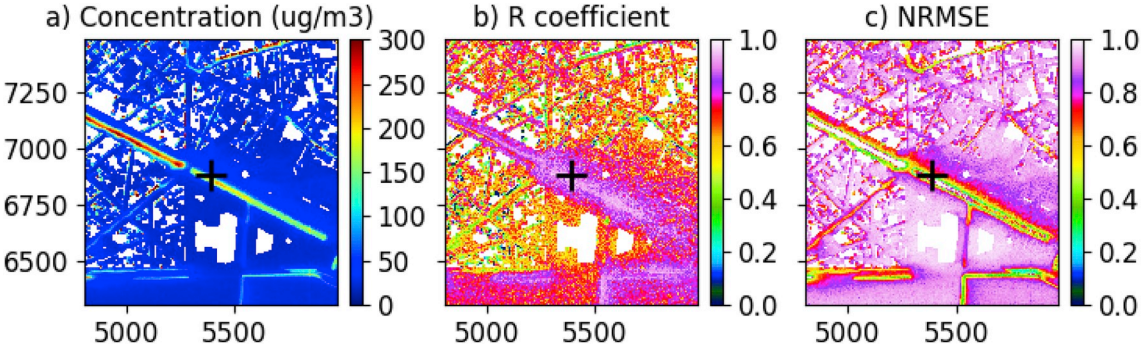

Fig. 4. Champs-Élysées Avenue area NO_X concentration (a), NO_X correlation coefficient (b), NO_X NRMSE on the 2nd of March 2016.

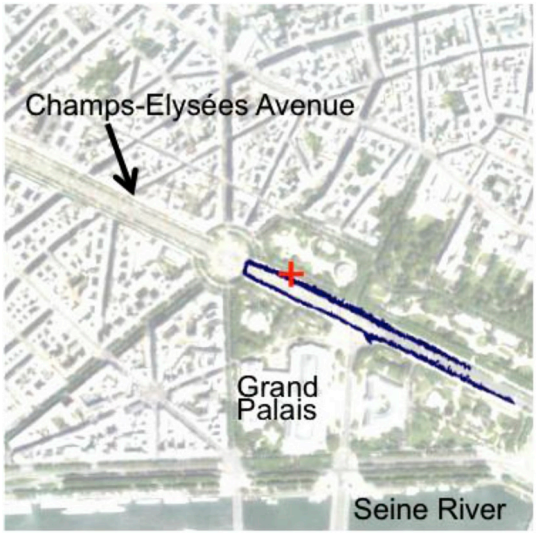

Fig. 5. NO_X similarity area around Champs-Élysées Avenue on the 2nd of March 2016. The red cross shows the pseudo-station location.

is calculated from all stations for the 10 days and reported as percentage of success.

The main difficulty is to find a combination of criteria thresholds that involves a good agreement for the maximum of days and the maximum of traffic-oriented areas.

In the case where two different combinations of criteria thresholds give the same score (percentage of success), the retained combination is the more restrictive (i.e. the higher *R* correlation coefficient and/or the lowest NRMSE value). Table 3 summarizes the results of this iterative analysis for the 5 traffic-oriented monitoring-sites and for the 10 days of the study.

For NO_X similarity area, with a constant correlation coefficient of 0.7 and a NRMSE varying from 0.4 to 0.5 with a 0.05 step, the total success rate is comprised between 90 and 100%. Combination of R coefficient equal to 0.7 and NRMSE value of 0.4 selects too restrictive areas 5 times: 1 time (1 day) for 2 monitoring-site areas: BASCH_v, ELYS, CELES_v and 2 days for HAUS monitoring-site area. With correlation coefficients of 0.75 and 0.8 and NRMSE equal to 0.5, success rate is 98%. The combination of R coefficient equal to 0.7 and NRMSE equal to 0.45 and 0.5 has a 100% success rate for the 10 days and 5 traffic-oriented areas. In that case, the most restrictive criterion threshold is retained, namely R = 0.7 and NRMSE = 0.45.

Success rates in the selection of similarity areas are lower for PM_{10} than for NO_X. This is probably due to the fact that part of the local variability is overshadowed by the homogeneous regional signal of PM_{10} concentrations. Thus traffic has a smaller contribution for PM_{10} than for NO_X resulting in similarity areas too large ($> 400 \times 10^3\ m^2$). For correlation coefficients of 0.7, 0.75 and 0.8 and NRMSE = 0.3, success rates are in the range of 76–78%. With a combination of a correlation coefficient of 0.75 and NRMSE values of 0.25, 0.3 and 0.35,

Table 3
Success rates in the selection of (a) similarity and (b) homogeneity areas for NO_X and PM_{10} of all traffic monitoring-sites areas on the 10 days of the study. The retained criteria thresholds for similarity areas are shown in bold. In case of failure, the number of days and the selected area (m^3) are mentioned in italics.

a)

Pollutant	Criteria thresholds		Success rates for the ten days Traffic oriented stations					Total (%)
	R	NRMSE	BASCH_v	CELES_v	ELYS	HAUS	OPERA	
NO_X	0.7	0.4	9 *1 = 241*	9 *1 = 198*	9 *1 = 106*	8 *2 < 30*	10	90
	0.7	**0.45**	10	10	10	10	10	**100**
	0.7	0.5	10	10	10	10	10	100
	0.75	0.5	10	10	9 *1 = 193*	10	10	98
	0.8	0.5	10	10	9 *1 = 121*	10	10	98
PM_{10}	0.7	0.3	9 *1 = 846x10³*	8 *1 = 148* *1 = 459x10³*	7 *3 > 833x10³*	6 *4 > 561x10 ³*	9 *1 = 161*	78
	0.75	**0.3**	9 *1 = 846x10³*	8 *1 = 148* *1 = 459x10³*	7 *3 > 826x10³*	6 *4 > 561x10³*	9 *1 = 161*	**78**
	0.8	0.3	9 *1 = 846x10³*	8 *1 = 148* *1 = 459x10³*	6 *3 > 833x10³* *1 = 133*	6 *4 > 561x10³*	9 *1 = 161*	76
	0.75	0.35	6 *4 > 876x10³*	8 *1 = 180* *1 = 459 x10³*	6 *4 > 833x10³*	4 *6 > 466x10³*	10	68
	0.75	0.25	10	7 *3 < 226*	8 *1 > 833x10³* *1 = 253*	4 *4 > 553x10³* *2 < 145*	7 *3 < 206*	72

b)

Pollutant	Success rates for the ten days Traffic oriented stations					Total (%)
	BASCH_v	CELES_v	ELYS	HAUS	OPERA	
NO_X	9 *1 = 97*	1 *9 < 96*	10	2 *8 < 30*	1 *9 < 152*	46
PM_{10}	10	10	8 *2 > 453x10³*	5 *5 > 506x10³*	7 *3 < 197*	80

total success rates vary between 68% and 78%. However, NRMSE = 0.3 is not restrictive enough, resulting in the selection of vast, non-realistic similarity areas whereas 0.25 is too restrictive (OPERA area). The final similarity criteria thresholds resulting from the iterative analysis for similarity for the 5 traffic-oriented areas are given in bold letters in Table 3a.

Success rates for the selection of the daily homogeneity representativeness area using the fixed 20% "EC criterion" are shown in Table 3b. Total success rate for homogeneity areas are generally lower than for similarity areas. Detection of NO_X homogeneity areas has a success rate of 46% (100% for similarity areas). PM_{10} success rates for homogeneity and similarity areas are of the same order of magnitude 80% and 78% respectively. NO_X homogeneity areas are particularly restrictive around Boulevard Haussmann monitoring-site (HAUS) and around the bank of the Seine River (CELES_v) due to the strong spatial variability of NO_X concentration. Respectively for HAUS and CELES_v, only 2 and 1 out of the 10 selected areas may be considered successful.

The retained thresholds for PM_{10} are more restrictive than for NO_X. This is due to the higher spatial variability of NO_X concentrations, which requires less restrictive thresholds in order to retain areas large enough to get beyond the pseudo-station.

3. Results

3.1. Quantitative results

3.1.1. NO_X representative areas

3.1.1.1. Traffic-oriented areas. NO_X homogeneity and similarity areas around traffic sites for the 10 days in March 2016 are summarized in Fig. 6. NO_X traffic-oriented monitoring-sites have a large variability between stations, criteria (homogeneity vs similarity) and across the ten days period. NO_X traffic-oriented representativeness areas are comprised between 15 m^2 (minimum HAUS homogeneity area) and 51.5 $\times 10^3$ m^2 (maximum ELYS homogeneity area). A large variability is observed between the five traffic-oriented areas. ELYS has the largest similarity area, whereas OPERA has the smallest one, averaged over the 10 days, ELYS area is five times larger than the OPERA area (16.6 $\times 10^3$ m^2 vs 1.7 $\times 10^3$ m^2). ELYS station is located on a sidewalk along the wide Champs-Elysées Avenue whereas OPERA station is located on a crossroad and integrates contributions from several streets. ELYS has the largest homogeneity area whereas CELES_v has the smallest one (18.3 $\times 10^3$ m^2 vs 74 m^2).

Similarity areas are larger than homogeneity areas for CELES_v (ratio similarity (S)/homogeneity (H) = 164), HAUS (ratio S/H = 18) and BASCH_v (ratio S/H = 6). Similarity and homogeneity areas are of the same order of magnitude for ELYS. Homogeneity area is negligible for OPERA with a representativeness area averaged across the ten days of 74 m^2.

In order to get a synthetized result, the "most realistic" representativeness area is taken as the intersection of representativeness areas obtained at least seven days out of ten (70%). Applying a more restrictive condition (e.g. 100% of the time) leads to very small areas, whereas using the mean representativeness area across the ten days leads to a non-realistic representativeness area, given the high variability of selected areas across the ten days. As shown in Table 4, CELES_v station has the smallest NO_X homogeneity area (18 m^2) whereas BASCH_v has the smallest similarity area (846 m^2). ELYS has the largest homogeneity area (2844 m^2) and CELES_v has the largest similarity area (2745 m^2). Similarity areas are larger than homogeneity areas with a ratio ranging between 3.6 (BASCH_v) and 152.5 (CELES_v), except the ELYS station, where both areas are of the same order of magnitude.

The most probable homogeneity area is in some cases not significant (18 m^2 and 27 m^2 for NO_X at CELES_v and HAUS monitoring-sites respectively). A possible explanation for this, is that the 20% threshold, stated for consistency with literature values, is not optimal for data issued from the PMSS simulation, contrary to the thresholds used for similarity areas.

3.1.1.2. Urban background areas. Fig. 7 summarizes the representativeness areas around NO_X urban background monitoring-sites for the 10 days. Across the study period and all stations, mean NO_X urban background areas are fifty-one times larger than NO_X traffic-oriented areas (411 $\times 10^3$ m^2 vs 8.2 $\times 10^3$ m^2). NO_X urban background stations representativeness areas range between 9.6 $\times 10^3$ m^2 (minimum PA12 homogeneity area) and 932.0 $\times 10^3$ m^2 (maximum PA12, similarity area) whereas traffic-oriented stations representativeness areas range between 15 m^2 and 51.5 $\times 10^3$ m^2.

In most cases, the representativeness areas of urban background stations cover the neighborhood of the station. Representativeness areas of monitoring-sites located on schoolyards (PA12 and PA07) are in some cases very small but they still cover part of the neighborhood with a minimum value of 9.6 $\times 10^3$ m^2 (minimum PA12 homogeneity area) and 19.9 $\times 10^3$ m^2 (minimum PA07 similarity area). In fact, those monitoring-sites are partly influenced by traffic emissions due to their proximity to the roadside.

Large differences between the representativeness areas of the NO_X urban background stations are found. A ratio of 2.1 is estimated between the monitoring-site located on a pedestrian square close to the Centre Pompidou (PA04C) and the PA12 site (513.0 $\times 10^3$ m^2 vs. 242.7 $\times 10^3$ m^2 averaged over ten days for homogeneity and similarity criteria).

By averaging across the ten days and all urban background areas, we find that NO_X similarity areas are 1.3 times higher than NO_X homogeneity areas (468.7 $\times 10^3$ m^2 vs 353.5 $\times 10^3$ m^2). Similarity area values span over a greater range across the ten days of the study than homogeneity areas, particularly for PA07 and PA12. Homogeneity area values at PA15L, a monitoring-site located within a stadium close to "Boulevard Periphérique" and another major road during the 10 days of the study have the smallest dispersion. This can be explained by the sharp spatial gradient of pollutant concentration at the proximity of

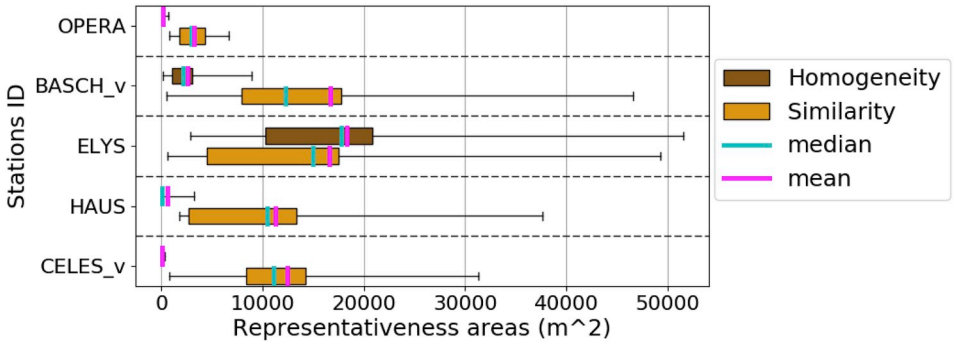

Fig. 6. Boxplots of NO_X homogeneity and similarity areas (m^2) around traffic sites across the 10 days of the study. Blue and pink solid lines inside the boxes represent the median and mean values of the distribution respectively, lower and upper bounds of the boxes represent the 25th and 75th percentiles, and low and high whiskers represent the minimum and maximum values.

Table 4
NO_X traffic-oriented representativeness areas (m^2) shared 70% of the time, average areas over the ten days of the study (m^2) and standard deviation (m^2).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m2)	Similarity (S) (m2)	Homogeneity: mean [standard deviation] (m2)	Similarity: mean [standard deviation] (m2)
OPERA	90	1125	165 [186]	3.2×10^3 [1.9×10^3]
BASCH_v	234	846	2.5×10^3 [2.4×10^3]	16.7×10^3 [13.6×10^3]
ELYS	2844	2529	18.3×10^3 [12.6×10^3]	16.6×10^3 [15.2×10^3]
HAUS	27	1728	615 [1189]	11.2×10^3 [10.3×10^3]
CELES_v	18	2745	74 [89]	12.4×10^3 [7.7×10^3]

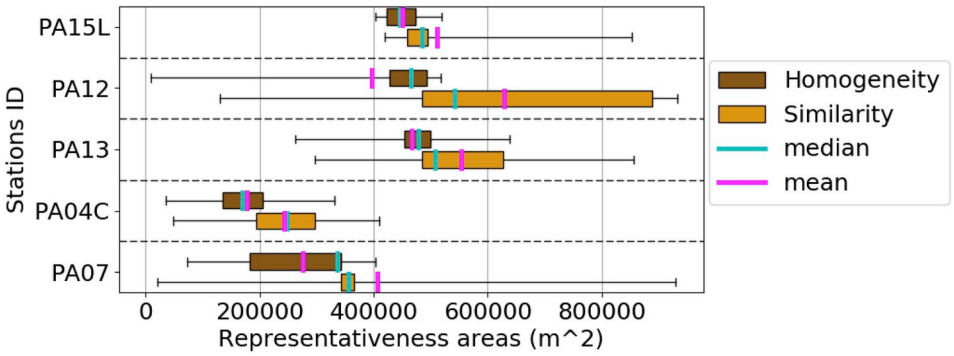

Fig. 7. Boxplots of NO_X homogeneity and similarity areas (m^2) around urban background monitoring-sites during the ten days in March 2016. Blue and pink solid lines inside the boxes represent the median and mean values of the distribution respectively, the lower and upper bounds of the boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 5
NO_X background representativeness areas (m^2) shared 70% of the time, average areas over the days (m^2) and standard deviation of the areas over the ten days (m^2).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m2)	Similarity (S) (m2)	Homogeneity: mean [standard deviation] (m2)	Similarity: mean [standard deviation] (m2)
PA15L	445.1×10^3	479.3×10^3	450.1×10^3 [34.7×10^3]	511.3×10^3 [117.5×10^3]
PA12	431.8×10^3	497.6×10^3	397.2×10^3 [161.7×10^3]	628.7×10^3 [252.9×10^3]
PA13	464.0×10^3	495.0×10^3	467.4×10^3 [95.3×10^3]	554.0×10^3 [154.9×10^3]
PA04C	121.7×10^3	193.8×10^3	176.8×10^3 [8.1×10^3]	242.7×10^3 [91.5×10^3]
PA07	186.0×10^3	350.7×10^3	275.9×10^3 [115.6×10^3]	406.9×10^3 [241.9×10^3]

these roads.

NO_X urban background representativeness areas obtained seven days out of ten (Table 5) are three hundred times wider than NO_X traffic representativeness areas ($1.2 \times 10^3\,m^2$ vs $366.5 \times 10^3\,m^2$). PA04C monitoring-site is the least representative ($121.7 \times 10^3\,m^2$ and $193.8 \times 10^3\,m^2$ for the largest homogeneity and similarity areas respectively). PA13 (monitoring-site located in the Choisy Park) and PA12 in the schoolyard have the larger homogeneity and similarity areas ($464.0 \times 10^3\,m^2$ and $497.6 \times 10^3\,m^2$ respectively). Homogeneity and similarity areas in most cases are of the same order of magnitude with the exception of PA07 station, where similarity area is two times larger than homogeneity area.

3.1.2. PM_{10} representativeness areas

3.1.2.1. Traffic-oriented areas. PM_{10} traffic-oriented areas are seventeen times larger than NO_X traffic-oriented areas. Fig. 8 summarizes PM_{10} representativeness areas of traffic sites for the 10 days of the study. In some cases, PM_{10} traffic-oriented areas cover a vast surface around the monitor (maximum of 846.2×10^3 m for BASCH_v similarity area). A strong variability is observed between the monitoring-sites areas, with the highest ratio of twenty-eight being found between the largest representative area around HAUS ($279 \times 10^3\,m^2$ averaged over ten days for both similarity and homogeneity) and the smallest one around OPERA ($9.9 \times 10^3\,m^2$). PM_{10} similarity areas are in general, larger than PM_{10} homogeneity areas with a ratio between 1.5 (ELYS) and 5 (BASCH_v) except for HAUS areas where they both are of the same order of magnitude with very vast areas. In most of the cases, representativeness areas have a high variability across the ten days (see boxplots width of BASCH_v similarity area, ELYS, HAUS and CELES_v).

By considering areas shared 70% of the time (Table 6), PM_{10} representativeness areas are nine times larger than NO_X representativeness area ($11.0 \times 10^3\,m^2$ vs $1.2 \times 10^3\,m^2$). A large variability is shown between stations with the OPERA site being the less representative monitor location ($171\,m^2$ and $1.2 \times 10^3\,m^2$ for homogeneity and similarity areas respectively). ELYS and BASCH_v have the largest homogeneity and similarity areas with $26.0 \times 10^3\,m^2$ and $38.4 \times 10^3\,m^2$ respectively. Similarity areas are larger than homogeneity areas for HAUS, OPERA and BASCH_v with a ratio of 4.8–13.6. Inversely, ELYS similarity areas are two times lower than the corresponding homogeneity areas. For CELES_v, similarity and homogeneity area are of the same order of magnitude.

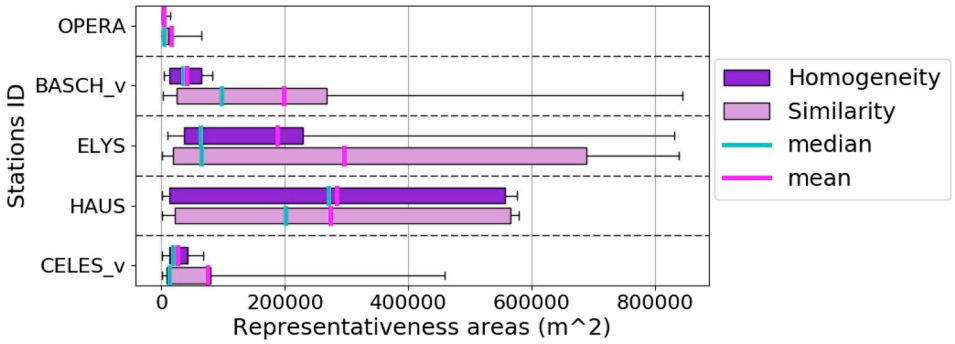

Fig. 8. Boxplots of PM_{10} homogeneity and similarity areas (m^2) around traffic station across the 10 days of the study. Blue pink solid lines inside the boxes represent the median and mean of the distributions respectively, lower and upper bounds of the boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 6
PM_{10} traffic-oriented representativeness areas (m^2) shared 70% of the time, average areas over the days (m^2) and standard deviation of the areas over the ten days (m^2).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m2)	Similarity (S) (m2)	Homogeneity: mean [standard deviation] (m2)	Similarity: mean [standard deviation] (m2)
OPERA	171	1.2×10^3	4.3×10^3 [4.4×10^3]	15.6×10^3 [23.6×10^3]
BASCH_v	2.8×10^3	38.4×10^3	40.5×10^3 [29.1×10^3]	197.7×10^3 [244.0×10^3]
ELYS	26.1×10^3	17.1×10^3	188.4×10^3 [252.4×10^3]	296.4×10^3 [360.0×10^3]
HAUS	2.6×10^3	12.3×10^3	283.8×10^3 [270.4×10^3]	274.4×10^3 [253.7×10^3]
CELES_v	5.2×10^3	4.6×10^3	25.7×10^3 [20.5×10^3]	74.4×10^3 [132.5×10^3]

3.1.2.2. Urban background areas. Fig. 9 summarizes PM_{10} representativeness areas around urban background monitoring-sites for the 10 days in March 2016. PM_{10} representativeness areas are generally larger than NO_X representativeness areas for urban background monitoring-site with a ratio of 1.9 across all days and urban background stations ($772 \times 10^3\,m^2$ vs $411 \times 10^3\,m^2$) and are comprised between $289 \times 10^3\,m^2$ and $988 \times 10^3\,m^2$.

Here again, we observe a large variability between representativeness areas of urban background monitoring-sites with a highest ratio of 1.8 between the most (PA12) and the less representative sites (PA04C). PM_{10} homogeneity and similarity areas are of the same order of magnitude with mean values across all days and urban background stations equal to $791.5 \times 10^3\,m^2$ and $753.2 \times 10^3\,m^2$ respectively. PM_{10} concentrations are more spatially homogeneous than NO_X concentrations that have a high spatial gradient. For PA15L and PA12, the 25th percentile values of similarity areas are higher than the ones for homogeneity areas but the minimum and maximum values are very close.

By considering areas shared seven days out of ten (Table 7), PM_{10} urban background representativeness areas are two times larger than NO_X urban background representativeness areas ($799.4 \times 10^3\,m^2$ vs $366.5 \times 10^3\,m^2$). PM_{10} urban representativeness areas are also seventy-five times larger than PM_{10} traffic representativeness areas ($799.4 \times 10^3\,m^2$ vs $10.6 \times 10^3\,m^2$). PA04C has the lowest representativeness area ($410.4 \times 10^3\,m^2$ and $428.2 \times 10^3\,m^2$ for homogeneity and similarity areas respectively). PA12 has the highest homogeneity ($931.7 \times 10^3\,m^2$) and similarity area ($955.9 \times 10^3\,m^2$).

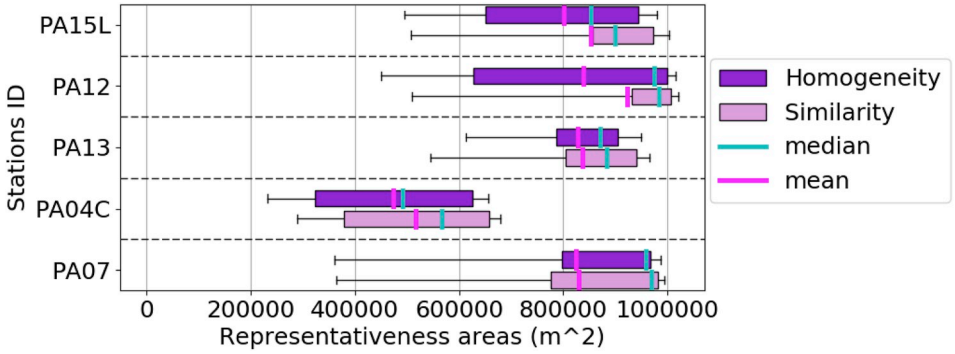

Fig. 9. Boxplots of PM_{10} homogeneity and similarity areas (m^2) on 10 days in March 2016 around urban background monitoring-sites. Blue and pink solid lines inside the boxes represent the median and mean of the distributions respectively, lower and upper bounds of boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 7
PM_{10} background representativeness areas (m^2) shared 70% of the time, average areas over the days (m^2) and standard deviation of the areas over the ten days (m^2).

ID STATIONS	Areas shared seven days on ten		Statistics over ten days	
	Homogeneity (H) (m2)	Similarity (S) (m2)	Homogeneity: mean [standard deviation] (m2)	Similarity: mean [standard deviation] (m2)
PA15L	833.0×10^3	881.9×10^3	801.7×10^3 [177.1×10^3]	852.4×10^3 [167.7×10^3]
PA12	931.7×10^3	955.9×10^3	838.3×10^3 [226.1×10^3]	922.8×10^3 [147.3×10^3]
PA13	849.4×10^3	855.7×10^3	828.0×10^3 [108.4×10^3]	836.5×10^3 [136.5×10^3]
PA04C	410.4×10^3	428.2×10^3	473.8×10^3 [156.9×10^3]	516.3×10^3 [146.6×10^3]
PA07	920.2×10^3	928.1×10^3	824.4×10^3 [238.1×10^3]	829×10^3 [242.8×10^3]

3.2. Characterization of urban areas

A large variability among stations is observed for the most probable traffic-oriented similarity areas (i.e. mean areas and areas shared seven days on ten), ranging from 18 m^2 (NO_X homogeneity area at CELES_v station) to 38.4×10^3 m^2 (PM_{10} similarity area at BASCH_v station), see Tables 4 and 6. This large variability may be related to the monitoring-site location and its vicinity (see Table 1). Two traffic-oriented monitoring-sites are located on a crossroad (OPERA is in the middle of the crossroad and BASCH_v is at the exit of the crossroad). Three traffic-oriented monitoring-sites are located on a sidewalk along roadside (HAUS, ELYS and CELES_v).

In this section, we will investigate the possible impact of specific urban features around the station on the differences in similarity areas. The specific features characterizing the representativeness areas selected around traffic-oriented monitoring-sites are listed in Table 8. Their only common feature is the sidewalk in the immediate proximity to the road.

Table 8
Characteristic features selected around traffic-oriented monitoring-sites.

Sites	Characteristic features
BASCH_v – crossroad exit	Sidewalk and part of the road leading to the crossroad
OPERA – crossroad	Small area close to the monitoring-site
CELES_v – border of the Seine River	Sidewalk and sometimes Seine river and river bank
ELYS – Champs-Elysées Avenue	Sidewalk on one or both sides of the road
HAUS – Boulevard Haussmann	Sidewalk or all elements in the area (PM_{10})

3.2.1. Traffic-oriented monitoring-sites located on crossroad (BASCH_v and OPERA)

As shown in section 3.1.1 NO_X representativeness areas are smaller than the PM_{10} representativeness areas for both monitoring-sites (0.6×10^3 m^2 vs 10.6×10^3 m^2 averaged across criteria and stations). NO_X and PM_{10} similarity areas at the OPERA monitoring-site are of the same order of magnitude (1125 m^2 and 1179 m^2, see Tables 4 and 6) and both include a roadside segment of 140 m along the large crossroad. However, as shown in Fig. 10 the diurnal variability of the selected PM_{10} similarity area is much higher than for NO_X (see also the standard deviations of similarity areas in Tables 4 and 6).

Two times out of ten, PM_{10} similarity area around the OPERA monitoring-site includes main traffic-roads and secondary axes over a vast area around the station (Fig. 10d).

The most probable NO_X similarity area around the BASCH_v station is smaller than the OPERA one. It includes a small portion of the sidewalk and the street along a distance of about 100 m. The PM_{10} similarity area around BASCH_v is larger than the one around the OPERA monitor and also, larger than the NO_X area around BASCH_v. It includes a part of the street side and the sidewalk where the monitoring-site is located (avenue Jean Moulin) over a length of 700 m but also two other large avenues, one leading to the crossroad, and another one further down Jean Moulin avenue on the south-east direction. Thus, traffic sources from different avenues more or less distant to the monitor site may contribute to the pollutant concentration measured at the BASCH_v monitoring-site. On the contrary, OPERA is isolated in the middle of the crossroad and does not integrate pollution sources from the nearby roadside except for a small road segment. These differences in the topography of the two stations explain the large discrepancy in PM_{10} representativeness areas between the two sites (most probable PM_{10} similarity area 38.4×10^3 m^2 (BASCH_v) vs 1179 m^2 (OPERA)).

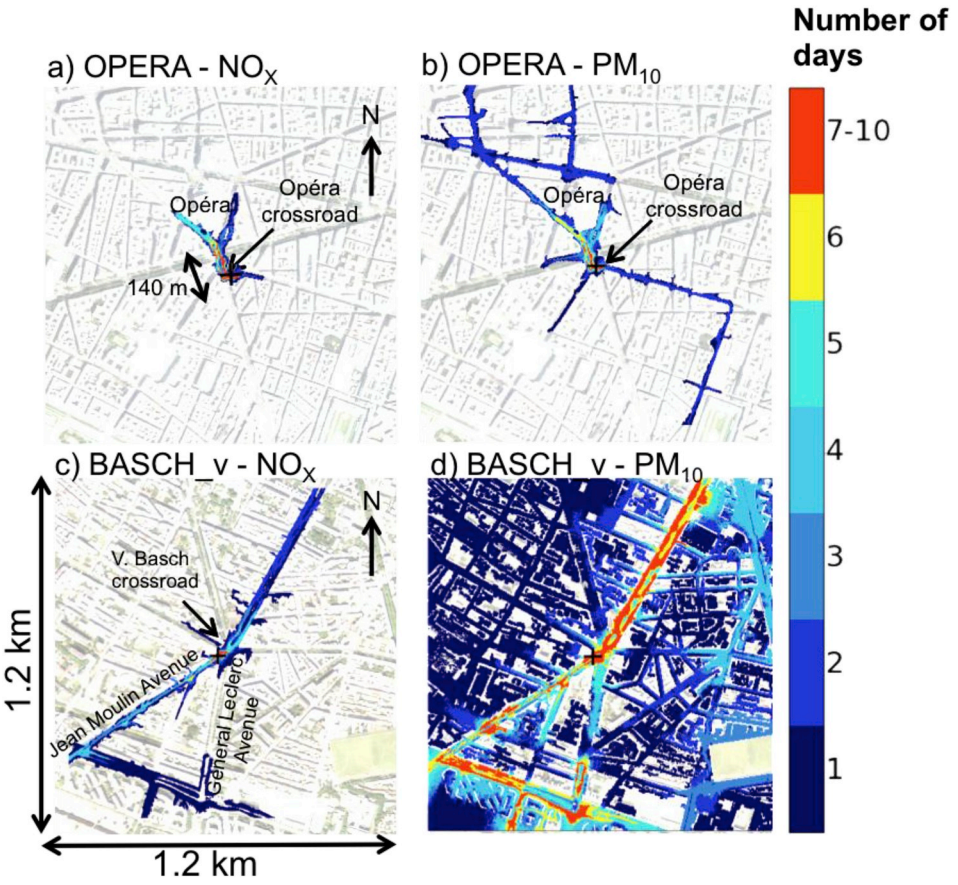

Fig. 10. NO_X (left) and PM_{10} (right) similarity areas cumulated over the 10 days of the study for OPERA monitoring-site (top) and BASCH_v monitoring-site (bottom). The black cross indicates the location of the pseudo-station. The color bar indicates the number of days for which a street portion is selected. The most probable area (selected at least 7 days) is shown in red color in each case. Black arrow indicates the north wind direction.

3.2.2. Traffic-oriented monitoring-sites located along a roadside (ELYS, HAUS and CELES_v)

ELYS monitor station is located on the sidewalk of a wide urban boulevard. NO_X similarity area includes the sidewalk of the station along a distance of 500 m, while the much larger PM_{10} area (2.5×10^3 m^2 vs 17.1×10^3 m^2 for NO_X and PM_{10} respectively) includes both sidewalks of the road along a distance of 700 m as well as a 400 m segment of a sidewalk from the station in northwest direction (Fig. 11). PM_{10} similarity area at the ELYS monitor has a stronger variability during the ten days of the study than NO_X similarity area with a standard deviation over ten days of 360×10^3 m^2 (PM_{10}) and 15.2×10^3 m^2 (NO_X). Secondary arteries perpendicular to the primary axe are included in the selected area 10% of the time for NO_X and 50% of the time for PM_{10}. Four days out of ten, all the streets of the domain shown in Fig. 11b as well as part of the Seine River are included in the PM_{10} similarity area.

The most probable PM_{10} similarity area around Boulevard Haussmann monitoring-site (HAUS) is larger than the NO_X one (12.3×10^3 m^2 and 1.7×10^3 m^2, respectively). NO_X similarity area includes the sidewalk of the station along a distance of 200 m from the station and a small street perpendicular to the Haussmann boulevard, in the northwest direction. PM_{10} similarity area includes a portion of sidewalk over 400 m and secondary axes (Fig. 11). The PM_{10} similarity area has a higher variability for these ten days than the NO_X area (see standard deviations on Tables 4 and 6). 20–30% of the time, NO_X similarity areas include small segments of secondary axes, whereas 40 and 50% of the time, PM_{10} similarity areas include all streets segments except from the busy roadside near the monitoring-site.

The extents of the most probable NO_X and PM_{10} similarity areas around the Seine River Bank monitoring site (CELES_v) are close to each other (2.7×10^3 m^2 vs 4.6×10^3 m^2 for NO_X and PM_{10} respectively). They both include a segment of the sidewalk and roadside south-east from the monitoring-site along a distance of 640 m. Similarity areas extend along the bank of the Seine river four days out of ten for PM_{10} whereas for NO_X selected similarity is not as widely spread. One day out of ten, PM_{10} similarity area selects secondary axes and crosses the bridge. Note that the Southeast part of the road is not selected, due to the lack of emission data in the simulation.

The most probable NO_X similarity areas are of the same order of magnitude for the three monitoring sites (around 2×10^3 m^2). However, the CELES_v PM_{10} area is smaller than the two other (4.6×10^3 m^2 (CELES_v) vs. 17.1×10^3 m^2(ELYS) and 12.2×10^3 m^2 (HAUS)). This difference may be attributed to the road configuration. ELYS is located at the edge on the 20 m wide sidewalk of a 30 m wide avenue, whereas HAUS is located at the edge on the 9 m wide sidewalk of a 18 m wide road. The CELES_v monitor has the particularity to be located on the sidewalk (6 m wide) of a road (12 m wide) from which it is separated with a 3 m wide bicycle path.

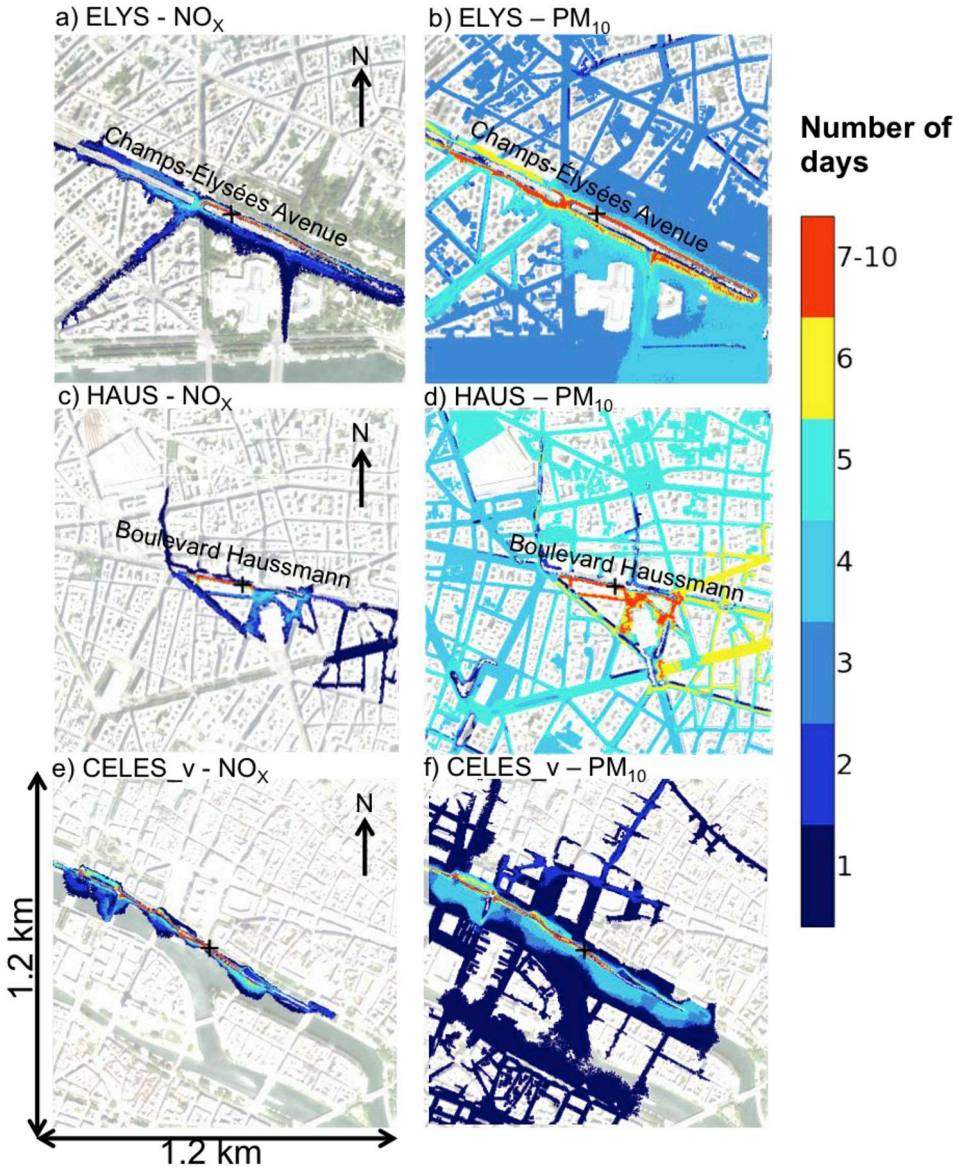

Fig. 11. Same as Fig. 10, but for three monitor sites, ELYS (top), HAUS (middle) and CELES_v (bottom).

4. Discussions and conclusions

A new methodology has been developed to assess daily NO_X and PM_{10} representativeness areas around traffic and urban-background monitoring-sites of the air-quality network of AIRPARIF in the city of Paris, France. It is based on a high-resolution model simulation (PMSS model) during ten days in March 2016, including a typical PM_{10} winter pollution episode due to high-pressure conditions. The PMSS modeling system provided pollutant concentration fields at a 3 m horizontal resolution with in a 2 m thick surface layer. Modeled concentrations were compared to measurements at 10 monitoring-sites, showing that simulations are in good agreement with the observations for PM_{10} and stay realistic for NO_X even if model has some difficulty to reproduce NO_X concentration peaks. To perform this study, we assume that the model properly reproduces the pollution spatial variability in Paris.

A representativeness area should be seen as an area where pollution variations are very close to those at the monitoring-site. Two kinds of such “resemblance” were investigated in this study: homogeneity and similarity. Homogeneity areas integrate points where daily concentrations do not differ more than 20% of the daily averaged value at the monitoring-site location. Similarity areas incorporate points satisfying the additional condition to be strongly correlated with concentrations at the location of the monitoring-site. We first determined the optimal thresholds on correlation and normalized root mean square, considering that the obtained area surface must be confined within the [300 and 400 $\times$ 10^3 m^2] range, for a maximum number of days and a maximum number of sites. A sensitivity analysis was performed to choose the lowest NRMSE and the highest value of correlation coefficient specific for NO_X and PM_{10} traffic-oriented sites.

This new approach has been found to be robust despite some gaps in the traffic emissions and the small dataset (only ten days). The same criteria thresholds were applied for the different monitoring-sites of the AIRPARIF network leading to realistic representativeness areas for both urban background and traffic sites. The similarity criteria thresholds applied in the present study are most probably dependent of the study area and period as well as of the dispersion model. However the proposed methodology is generic and could be easily repeated under a different study setup.

Homogeneity and similarity areas are most of the time of the same order of magnitude for both NO_X and PM_{10} urban background monitoring-site. The difference between these two representativeness areas is more pronounced around traffic monitor-sites and in most cases, the

most probable similarity area is larger than the most probable homogeneity area. The definition of similarity areas is composed of two components, a threshold on the correlation coefficient, to ensure that the time series inside the similarity area have correlated variations, and a threshold on the NRMSE, to ensure that the time series variations have similar values. As a consequence, the daily average does not strongly differ as well. The overall criterion to select similarity areas integrates the condition of homogeneity with an additional constraint on the correlation coefficient. To check this hypothesis, we compared representativeness areas selected on the basis of a condition only on the NRMSE against homogeneity areas. For PM_{10} of urban background stations and for NO_X of 3/5 urban background stations (PA15L, PA12, PA13), NRMSE and homogeneity criteria select an area (km^2) similar with respectively an averaged ratio between the most probable NRMSE area and the most probable homogeneity area to 1 and 1.1. NRMSE most probable NO_X areas are wider than most probable NO_X homogeneity ones for 2/5 urban stations with a ratio to 1.6 (PA04C) and 1.9 (PA07), and much larger for traffic stations with an averaged ratio to 77.2 (range ratio between 1.3 (ELYS) and 213 (CELES_v)).

Most probable PM_{10} NRMSE areas are larger than most probable PM_{10} homogeneity area for OPERA, BASCH_v and HAUS with an averaged ratio between the most probable NRMSE area and the most probable homogeneity area of 8.6. For CELES_v they have the same value. For ELYS monitoring-site, most probable PM_{10} homogeneity areas are slightly larger than most probable PM_{10} NRMSE areas with a ratio of 0.7.

NRMSE thresholds were determined using a sensitivity analysis which integrates the daily variability and the diversity of urban elements landscape (by defining areas limits $> 300\,m^2$ and $< 400 \times 10^3\,m^2$). Given the retained thresholds, this criteria is less stringent than homogeneity criteria. However, homogeneity criteria do not integrate any area specificities or daily peak.

In the FAIRMODE and AQUILA inter-comparison studies, representativeness areas of one traffic and two urban background monitoring-sites were estimated from annual year data (in Belgium). In this framework, results obtained with eleven methodologies have been compared, and show an important variability (FAIRMODE report, Kracht et al. 2017): PM_{10} traffic-oriented representativeness areas are estimated between 0.001 km^2 ($10^3\,m^2$) and 160 km^2 whereas here, most probable PM_{10} traffic-oriented similarity area range between 0.001 km^2 ($1.2 \times 10^3\,m^2$) and 0.003 km^2 ($38.4 \times 10^3\,m^2$). In the FAIRMODE report, PM_{10} urban background representativeness areas are estimated between 1 km^2 ($10^6\,m^2$) and 700 km^2. In our study, most probable PM_{10} similarity and homogeneity areas of urban background monitoring-sites range between 0.4 km^2 and 0.9 km^2 due to the limitation of the study domain fixed 1.2 km × 1.2 km around each station.

The representativeness areas estimated in this study show a large daily variability. Traffic variations between working days, week-ends and national holidays as well as meteorological conditions (mainly the wind) are probably the two principal causes of this variability for both PM_{10} and NO_X. Unfortunately, the small dataset did not allow us to further investigate the daily traffic variation impact on representativeness areas.

Among the causes of this day-to-day variability, the wind can be suspected to play a role. The influence of the wind speed and wind direction has been investigated. Wind direction was not found to have a significant effect based on this small sample (only ten days and four windy days). The influence of the wind speed on the representativeness areas has been investigated by comparing similarity areas on windy days against non-windy days, we used 3 m/s as threshold to distinguish between the two. Preliminary results suggest that similarity areas are smaller in case of windy days and that the wind impact is stronger for traffic-oriented sites. The ratio between the difference of no windy area and windy area on the mean area for traffic stations is 1.7 and 1.8 respectively for NO_X and PM_{10} and for urban background stations is 1.1 and 0.6 for NO_X and PM_{10}. However, such an analysis should be carried out over longer simulation periods to provide statistically significant results.

The study showed that in general, urban background areas are representative of the neighborhood in the vicinity of the monitoring-site whereas traffic-oriented monitoring-sites are representative of specific urban features such as sections of roads and sidewalks along the road. Averaged over all monitoring-sites, most probable similarity areas for urban background monitoring-sites are about seventy times larger than traffic representativeness areas. Averaged over all stations, PM_{10} most probable similarity areas are 2 times larger than NO_X areas around urban monitors and 8 times around traffic sites. Even if PM_{10} concentrations were normalized to remove part of the regional signal (see Sect. 2.4) the remaining trace of large-scale effects seem to overshadow local effects such as the sharp concentration gradients over the road network.

As a next step, AIRPARIF monitoring-site measurements will be assimilated to the model simulations in order to correct some model artifacts and provide more reliable concentrations as input to the computation of the similarity areas. The present study should be expanded over a much greater number of simulated days, to provide more general conclusions and confirm our assumptions on the influence of parameters such as wind speed and traffic patterns. Then, measurements at monitor sites could be spread across similarity areas depending on the wind vector and provide an information particularly relevant for human exposure estimation.

Acknowledgements

Acknowledgments to ARIA Technologies for giving those simulations for ten days.

References

APHEKOM, 2012. Improving Knowledge and Communication for Decision Making on Air Pollution and Health in Europe. http://aphekom.org.

Beauchamp, M., Malherbe, L., 2016. Annexe technique au rapport intitulé estimation de l'exposition des populations aux dépassements de seuils réglementaires. In: Interpolation des sorties de modèles urbains par krigeage avec dérive polynomiale. note lcsqa, . http://www.lcsqa.org.

Blanchard, C.L., Tanenbaum, S., Hidy, G.M., 2014. Spatial and temporal variability of air pollution in Birmingham, Alabama. Atmos. Environ. 89, 382–391. https://doi.org/10.1016/j.atmosenv.2014.01.006.

Bobbia, M., Cori, A., de Fouquet, C., 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. Pollut. Atmos. https://doi.org/10.4267/pollution-atmospherique.

CITEPA. Center Interprofessionnel Technique d'Etudes de la Pollution Atmospherique https://www.citepa.org/en/air-and-climate/pollutants-and-ghg.

Duyzer, J., van den Hout, D., Zandveld, P., van Ratingen, S., 2015. Representativeness of air quality monitoring networks. Atmos. Environ. 104, 88–101. https://doi.org/10.1016/j.atmosenv.2014.12.067.

European Commission (EC), 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. The framework directive. Off. J. Eur. Union En. Ser. L152/51.

European Commission (EC), 2011/850/EU. Commission Implementing Decision of 12, December 2011. Laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality. Off. J. Eur. Union En. Ser. L 335/86.

Hagemann, R., Corsmeier, U., Kottmeier, C., Rinke, R., Wieser, A., Vogel, B., 2014. Spatial variability of particle number concentrations and NOx in the Karlsruhe (Germany) area obtained with the mobile laboratory "AERO-TRAM". Atmos. Environ. 94, 341–352. https://doi.org/10.1016/j.atmosenv.2014.05.051.

Hagenbjörk, A., Malmqvist, E., Mattisson, K., Sommar, N.J., Modig, L., 2017. The spatial variation of O3, NO, NO2 and NO x and the relation between them in two Swedish cities. Environ. Monit. Assess. 189. https://doi.org/10.1007/s10661-017-5872-z.

Harrison, R.M., 2018. Urban atmospheric chemistry: a very special case for study. Clim. Atmos. Sci. 1, 5. https://doi.org/10.1038/s41612-017-0010-8.

Host, S., 2013. « Exposition à la pollution atmosphérique liée au trafic routier et risques sanitaires », *VertigO - la revue électronique en sciences de l'environnement*. http://journals.openedition.org/vertigo/12816.

IARC report, 2013. https://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf.

Janssen, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D., Trimpeneers, E., Mensink, C., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. Atmos. Environ. 59, 492–500. https://doi.org/10.1016/j.atmosenv.2012.05.028.

Kracht, O., Santiago, J.-L., Martin, F., Piersanti, A., Cremona, G., Righini, G., Vitali, L., Delaney, K., Basu, B., Ghosh, B., Spangl, W., Brendle, C., Latikka, J., Kousa, A., Pärjälä, E., Meretoja, M., Malherbe, L., Letinois, L., Beauchamp, M., Lenartz, F., Hutsemekers, V., Nguyen, L., Hoogerbrugge, R., Eneroth, K., Silvergren, S., Hooyberghs, H., Viaene, P., Maiheu, B., Janssen, S., Roet, D., Gerboles, M., 2017. Spatial Representativeness of Air Quality Monitoring Sites - Outcomes of the FAIRMODE/AQUILA Intercomparison Exercise. EUR 28987 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/60611, JRC108791. ISBN 978-92-79-77218-4.

Martin, F., Fileni, L., Palomino, I., Vivanco, M.G., Garrido, J.L., 2014. Analysis of the spatial representativeness of rural background monitoring stations in Spain. Atmos. Pollut. Res. 5, 779–788. https://doi.org/10.5094/APR.2014.087.

Moussafir, J., Olry, C., Nibart, M., Albergel, A., Armand, P., Duchenne, C., Mahé, F., Thobois, L., Loaëc, S., Oldrini, O., 2014. AIRCITY: A Very High Resolution Atmospheric Dispersion Modeling System for Paris, vol. 1 American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FEDSMhttps://doi.org/10.1115/FEDSM2014-21820.

Nappo, C.J., Caneill, J.Y., Furman, R.W., Gifford, F.A., Kaimal, J.C., Kramer, M.L., Lockhart, T.J., Pendergast, M.M., Pielke, R.A., Randerson, D., Shreffler, J.H., Wyngaard, J.C., 1982. The workshop on the representativeness of meteorological-observations, June 1981. Boulder, Colorado. Bull. Am. Meteorol. Soc. 63, 761e764.

Nguyen, P.L., Stefess, G., de Jonge, D., Snijder, A., Hermans, P.M.J.A., van Loon, P., Hoogerbrugge, R., 2012. Evaluation of the Representativeness of the Dutch Air Quality Monitoring Stations. The National, Amsterdam, Noord-Holland, Rijnmond-Area, Limburg and Noord-Brabant Networks. RIVM Report 680704021/2012.

Piersanti, A., Vitali, L., Righini, G., Cremona, G., Ciancarella, L., 2015. Spatial representativeness of air quality monitoring stations: a grid model based approach. Atmos. Pollut. Res. 6, 953–960. https://doi.org/10.1016/j.apr.2015.04.005.

REVIHAAP Project, 2013. Review of evidence on health aspects of air pollution. http://www.euro.who.int/_data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1.

Santiago, J.L., Martín, F., Martilli, A., 2013. A computational fluid dynamic modelling approach to assess the representativeness of urban monitoring stations. Sci. Total Environ. 454–455, 61–72. Senat Report n°610, 2015. https://doi.org/10.1016/j.scitotenv.2013.02.068. http://www.senat.fr/fileadmin/Fichiers/Images/commission/enquete/pollution_air/Synthese_CE_Pollution_de_l_air.pdf.

Spangl, W., Schneider, J., Moosmann, L., Nagl, C., 2007. Representativeness and Classification of Air Quality Monitoring Stations – Final Report. Service Contract to the European Commission - DG Environment Contract No. 07.0402/2005/419392/MAR/C1. Umweltbundesamt, Wien. Reports, Bd. REP-0121.

Vardoulakis, S., Gonzalez-Flesca, N., Fisher, B.E.A., Pericleous, K., 2005. Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. Atmos. Environ. 2725–2736. Fourth International Conference on Urban Air Quality: Measurement, Modelling and Management, 25–28 March 2003 39. https://doi.org/10.1016/j.atmosenv.2004.05.067.

Vitali, L., Morabito, A., Adani, M., Assennato, G., Ciancarella, L., Cremona, G., Giua, R., Pastore, T., Piersanti, A., Righini, G., Russo, F., Spagnolo, S., Tanzarella, A., Tinarelli, G., Zanini, G., 2016. A Lagrangian modelling approach to assess the representativeness area of an industrial air quality monitoring station. Atmos. Pollut. Res. 7, 990–1003. https://doi.org/10.1016/j.apr.2016.06.002.