



HAL
open science

NOX and PM10 Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France

Delphy Rodriguez, Éric Parent, Laurence Eymard, Myrto Valari, Sébastien Payan

► **To cite this version:**

Delphy Rodriguez, Éric Parent, Laurence Eymard, Myrto Valari, Sébastien Payan. NOX and PM10 Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France. Atmospheric environment: X, 2019, 3, pp.100038. 10.1016/j.aeaoa.2019.100038 . insu-02158416

HAL Id: insu-02158416

<https://insu.hal.science/insu-02158416>

Submitted on 19 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

NO_x and PM₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France

Delphy Rodriguez, Eric Parent, Laurence Eymard, Myrto Valari, Sébastien Payan



PII: S2590-1621(19)30041-3

DOI: <https://doi.org/10.1016/j.aeaoa.2019.100038>

Article Number: 100038

Reference: AEAOA 100038

To appear in: *Atmospheric Environment: X*

Received Date: 24 January 2019

Revised Date: 27 May 2019

Accepted Date: 29 May 2019

Please cite this article as: Rodriguez, D., Parent, E., Eymard, L., Valari, M., Payan, Sé., NO_x and PM₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France, *Atmospheric Environment: X* (2019), doi: <https://doi.org/10.1016/j.aeaoa.2019.100038>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NO_x and PM₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France

Delphy Rodriguez¹ (1), Eric Parent (2), Laurence Eymard (1), Myrto Valari (3), Sébastien Payan (1)

(1) LATMOS/IPSL, Sorbonne Université, UVSQ, CNRS, Paris, France

(2) AgroParisTech, 16 rue Claude Bernard, Paris, France

(3) LMD, Ecole Polytechnique Palaiseau, France

HIGHLIGHTS

- A Bayesian scheme combining simulated pollutant concentrations with measured ones.
- Bias correction for high-resolution air quality modeling.
- Enhanced concentration estimates of air quality monitoring network.
- Spatially resolved concentration estimates around urban monitoring sites.

Abstract

Air quality over cities is mainly monitored by in-situ surface measurements. However, these stations are too sparse to properly capture the inhomogeneity of pollutant concentrations over urban areas. The need for high-resolution concentration estimate has grown in recent years, together with the awareness of the harmful effects of air pollution. In this study, we develop a Bayesian scheme that combines the high-resolution (3x3 m²) Particulate Micro SWIFT SPRAY numerical air quality simulations (PMSS) with operational surface measurements. The goal is to improve NO_x and PM₁₀ PMSS concentrations estimates over monitoring stations and within their vicinity. For this purpose, we simulate pollutant concentrations over the city of Paris for ten days over the period of March 2016. The Bayesian model provides an enhanced estimate of pollutant concentration in space and time. At the monitoring stations location, these estimates are characterized by lower temporal dispersion compared to the simulated data. Within the vicinity of the monitor stations, enhanced concentration estimates are closer to observations. For NO_x, the improvement is stronger and occurs in a larger area for urban background stations than for traffic stations. Overall, NO_x improvement is higher than PM₁₀ improvement. The initial PMSS model prediction is more biased for NO_x than for PM₁₀ due to large uncertainties in NO_x emissions over the traffic network.

¹ Corresponding author.

Email address: delphy.rodriguez@latmos.ipsl.fr and delphy.rodriguez@gmail.com (D. Rodriguez)

40 **Keywords: Urban air pollution; Surface monitoring networks; High-resolution**
41 **modeling; Spatial representativeness; Bayesian modeling**

42
43

44 1 Introduction

45

46 In large cities, air quality is often monitored by ground station networks measuring
47 pollutant concentrations. Several types of areas are sampled: traffic, residential, etc.
48 The spatial representativeness of these measurements is limited to a small area
49 around the station, especially in dense urban areas: emissions from various sources,
50 as well as air flow channeling in street canyons, result in sharp horizontal gradients
51 of pollutant concentrations (Harrison, 2018). The need for high-resolution gridded
52 estimates of pollutant concentrations over big cities has grown in recent years,
53 together with the increasing awareness of the harmful health effects of atmospheric
54 pollution.

55 Various approaches have been developed to address the spatial distribution of
56 pollutant concentration around monitoring sites. These approaches use real-time
57 measurements from local networks and expand them in space and/or time based on
58 additional relevant information such as land-use, weather conditions, and
59 topography. They may be divided mainly into first-principle geophysical modeling
60 and statistical ones. Atmospheric pollutant dispersion models describe the physical
61 and chemical processes that govern the emission, transport, mixing, chemical
62 reaction and deposition of pollutants. High-resolution models, such as the Particulate
63 Micro Swift Spray (PMSS) air quality model integrated into the ARIA City modeling
64 platform (http://www.aria.fr/aria_city.php), provide realistic spatial distributions of
65 pollutant concentrations. However, they suffer from large biases due to uncertainties
66 on the emissions and on the parameterizations of the complex atmospheric
67 processes governing urban air quality. Statistical approaches include land-use
68 regression (Janssen, *et al.* 2012 and Son, *et al.* 2018), kriging interpolation schemes
69 (Beauchamp *et al.* 2017 and Wu *et al.* 2018, Martin *et al.* 2012), methods for
70 assimilation of measurements in models such as Kalman filtering (Hanea *et al.* 2004)
71 or 4D variational assimilation algorithms (Elbern *et al.* 2001) and Bayesian models
72 (Beloconi *et al.* 2018).

73 The scope of this study is thus to improve NO_x and PM₁₀ pollutant concentrations
74 simulated with the PMSS high-resolution air quality model over and in the vicinity of
75 the monitoring stations of the AIRPARIF local air quality agency in Paris, France
76 (<https://www.airparif.asso.fr>). To spatially expand the measurement around
77 monitoring stations, we use the concept of “representativeness areas” defined in our
78 previous work (Rodriguez *et al.* 2019). These areas consist of points where
79 concentration levels are similar and time variations correlated to those at the
80 monitoring site’s location. In practice, they are defined by applying criteria thresholds
81 on the Normalized Root Mean Square Error (NRMSE) and on the correlation
82 coefficient between simulated concentrations over all the pixels around the
83 monitoring sites and simulated concentration at the monitoring site’s location.

84

85 Frequentist or Bayesian statistical approaches may be used. In both approaches, the
86 purpose is to determine as accurately as possible an unknown parameter θ . Here,
87 the unknown parameter θ is the concentration in the so-called “representativeness

88 area” of an AIRPARIF monitoring site, i.e. at the vicinity of a station. In the
89 frequentist setting, the unknown parameter θ is pre-assigned to a fixed single value.
90 A statistical procedure is chosen to establish a confidence interval from any set of
91 data collected in the same conditions, under the same model. It ensures the
92 probability of this random interval to cover the fixed value θ , i.e. the reliability of the
93 chosen statistical procedure. Then, the confidence interval is calculated from the
94 available data as a random draw from this estimation procedure. All the information
95 is contained in data. The frequentist procedure assumes a reproducible experience
96 and a large data set. Frequentists claim that this is an *objective* way of reasoning.
97 Within the Bayesian paradigm, the unknown parameter θ is considered uncertain.
98 What is known about θ is represented by a random variable conditioned upon the
99 available information. The major issue is to quantify the uncertainty from all available
100 data (Boreux *et al.*, 2010). One of the most significant differences between
101 frequentist and Bayesian approaches is the choice of a “prior” (O’Hagan, 2008). Prior
102 characterizes our expert knowledge by giving suggestions on an unknown
103 parameter θ (as a distribution of θ values) before observing data and is *subjective*
104 since it reflects a personal probabilistic judgment.

105
106 In recent years, Bayesian statistics have become a standard methodology in
107 environmental science, it has been used to improve spatial predictions of pollutant
108 concentration by Pirani *et al.*, 2018, Amin *et al.*, 2015, Millan *et al.*, 2009, Sahu *et al.*,
109 2005, among others, and deposition (Cowles *et al.*, 2003). In this study, we propose
110 a Bayesian model that updates the distribution of current and past ground station
111 measurements with the high-resolution output of the PMSS model ($3 \times 3 \text{m}^2$) within the
112 city of Paris for ten days over the period of March 2016.

113
114 In Section 2, the AIRPARIF air quality network and PMSS model simulations are
115 presented. Then, representativeness areas, as obtained in Rodriguez *et al.* (2019),
116 are summarized, and the Bayesian framework is described. Section 3 gathers
117 results: first, the pollutant concentrations at the monitoring site locations (Sect. 3.1)
118 and then, concentrations within representativeness areas (Sect. 3.2). Conclusions
119 are given in section 4.

120

121 **2 Data and methods**

122

123 As in Rodriguez *et al.* (2019), a simulation set of ten non-consecutive days in March
124 2016 are used. During this period (March 2 – 21), a high-pressure episode occurred,
125 characterized by strong atmospheric stability and the accumulation of PM_{10} within
126 the boundary layer. It resulted in a sharp rise in PM_{10} concentration on March 11.
127 The effect of this synoptic-scale episode is less pronounced for NO_x than for PM_{10} .
128 Indeed, in highly urbanized areas, the spatial and temporal variability of pollutant
129 concentration is more driven by traffic patterns for NO_x than for PM_{10} . We note that
130 the contribution of the traffic sector to the total emission concentrations is two times
131 higher for NO_x than for PM_{10} in the Paris area (56% vs. 28% AIRPARIF, 2012). In
132 the close proximity to roads, the NO_x emissions decrease much higher for NO_x
133 (Pasquier *et al.*, 2017). Data from eight local monitoring stations from the AIRPARIF
134 network have been collected for the same period.

135

136
137
138

139 2.1 Surface measurements

140

141 The location of AIRPARIF monitoring sites have been chosen to sample various
142 urban landscapes in Paris. They are classified as “traffic” or “urban background”
143 depending on their proximity (or not) to the street network. A thorough description of
144 the AIRPARIF network can be found on the agency’s website
145 (<https://www.airparif.asso.fr/en/stations/index>). Data from three traffic-oriented and
146 five urban background monitoring stations are used in this study as listed in **Table1**.
147 Traffic monitoring stations are located on sidewalks along roads and on major
148 crossroads, whereas urban background monitors are located on low emission areas
149 such as parks, pedestrian squares, schoolyards, and leisure parks. All traffic
150 monitors measure both NO_x and PM₁₀, whereas PM₁₀ is measured only at two out of
151 five urban background sites.

152 These measurements are considered as reference, or “ground truth” of pollutant
153 concentration, assuming that the measurement error is negligible.

154
155

| Type | ID station | Characteristic features | Pollutants |
|---------------------------|------------|---|------------------------------------|
| Traffic-oriented stations | HAUS | Bld Haussmann*, side roads | NO _x , PM ₁₀ |
| | OPERA | Crossroad*, road traffic | NO _x , PM ₁₀ |
| | ELYS | Champs-Élysées Avenue*: large road bordered with trees, side roads | NO _x , PM ₁₀ |
| Urban background stations | PA15L | Stadium* and leisure park, “Boulevard Périphérique”, traffic road | NO _x , PM ₁₀ |
| | PA12 | Schoolyard*, railways, traffic road | NO _x |
| | PA13 | Park*, traffic road, crossroad | NO _x |
| | 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 |

156 **Table1: Characteristic features of the study areas. Locations marked with an**
157 **asterisk indicate the location of the monitoring site.**

158

2.2 Numerical air quality simulations

NO_x and PM₁₀ concentrations were simulated using the high-resolution model PMSS, in a 12×10 km² grid covering the city of Paris. The model set-up is the same as in Rodriguez *et al.* (2019).

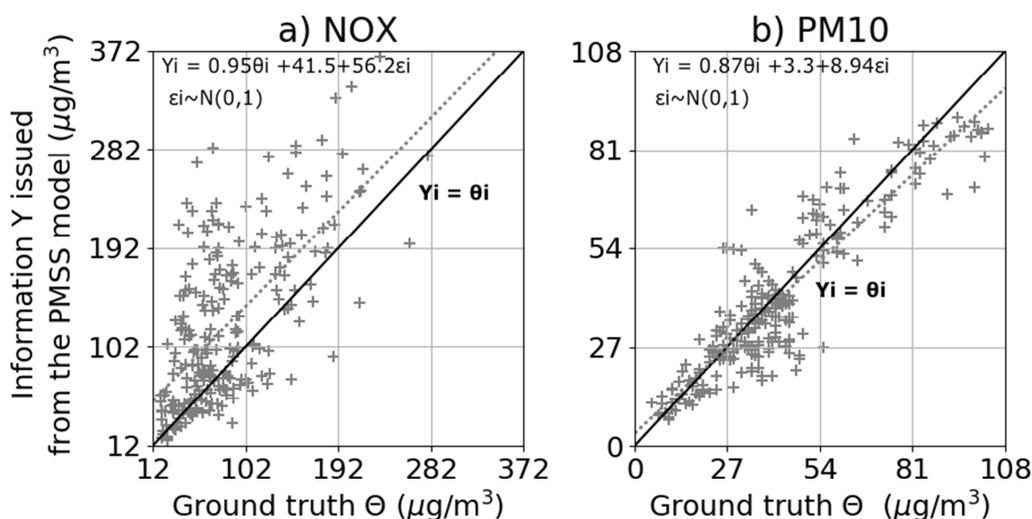
The model has thirty-five vertical layers from the ground up to an altitude of 800 m. The thickness of the surface layer is two meters and increases with the altitude (Moussafir *et al.*, 2015). The model simulates the turbulent flow around the buildings and pollutant dispersion through a three-meters horizontal resolution grid. Emission fluxes from urban sources are taken from the AIRPARIF inventory. Boundary conditions are taken from the CHIMERE chemistry-transport model (Mailler *et al.*, 2017). Meteorological conditions are simulated with the MM5 weather mesoscale model (<http://www2.mmm.ucar.edu/mm5/>). Building contours and heights are also accounted for as obstacles for the air flow. They 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 calculates in near real-time the traffic circulation and allows obtaining pollutant emissions on main traffic roads of the Paris area based on a “bottom-up” approach. Hourly traffic emissions are derived 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.

At each monitor, simulated concentrations are compared to measurements for both pollutants. PMSS simulations are in a very good agreement with PM₁₀ measurements, both for traffic and urban background sites, with a correlation coefficient (R^2) equal to 0.83 and 0.92 respectively, and a root mean squared error (RMSE) equal to 9.57 µg/m³ and 6.21 µg/m³ respectively (Rodriguez *et al.*, 2019). Model performance is not as good for NO_x with an R^2 value of 0.65 and 0.45, and RMSE of 81.57 µg/m³ and 21.12 µg/m³ for traffic and urban background sites, respectively.

Fig.1 shows scatter plots of concentrations at the ELYS monitoring station. The PMSS model fails to represent NO_x concentration maxima. As discussed in Rodriguez *et al.*, (2019), this may be due to the simple chemical mechanism of the PMSS model that fails to represent the concentration of the highly reactive pollutants such as NO and/or to uncertainties in the NO_x emission derived by the HEAVEN traffic emission modeling chain. Note that the error for NO₂ is smaller than for NO for traffic stations (RMSE(NO₂)=21.9 µg/m³ vs RMSE(NO)=72.3 µg/m³). A similar result is obtained for urban background stations.

The model is not always able to represent the high concentration levels during pollution episodes at the studied monitoring sites in Paris due to omissions in emission sources, errors in the meteorological predictions, or inaccuracy of the parameterization. However, it still provides realistic dispersion patterns of pollutant according to other studies during specific measurements campaigns in wind tunnel and road field experiment (Trini Castelli *et al.*, 2018) and for academic project (Hanna *et al.*, 2011). To perform this study, we assume that the model captures the

209 actual spatial variability of pollutant in Paris (Rodriguez *et al.*, 2019).
 210
 211



212
 213 **Fig.1: Scatter plots of measured vs simulated hourly concentrations for NO_x**
 214 **(a) and PM₁₀ (b) of Champs-Élysées avenue monitoring station during the ten**
 215 **days of March 2016. The black line represents $Y=\theta$. The grey dotted line**
 216 **represents the linear regression between AIRPARIF measurements and PMSS**
 217 **simulations.**

218

219 2.3 Spatial representativeness areas

220

221 PMSS simulations were used in Rodriguez *et al.* (2019) to determine
 222 representativeness areas around the AIRPARIF monitoring sites. The daily spatial
 223 representativeness area around a monitoring site was defined as the ensemble of
 224 adjacent pixels where the pollutant concentration is (1) strongly correlated with the
 225 concentration at the pixel of the station; (2) has a low statistical error (Normalized
 226 Root Mean Squared Error) with respect to the concentration at the pixel of the
 227 monitoring station. Correlation coefficient and NRMSE thresholds selection are
 228 based on an analysis over the 10-day with a step of 0.05 for both criteria. For each
 229 threshold, a score is estimated by counting the number of days on which the size of
 230 the selected area including the monitoring site is comprised between 300 m² and
 231 400×10³m².

232 The maximum score for each site is 10. Finally, a total score is calculated from all
 233 stations for the 10 days and reported as the percentage of success. In the case
 234 where two different combinations of criteria thresholds give the same score
 235 (percentage of success), the retained combination is the more restrictive (i.e. the
 236 higher ρ correlation coefficient and/or the lowest NRMSE value).

237 **Table2** shows the applied criteria thresholds to select representativeness areas.

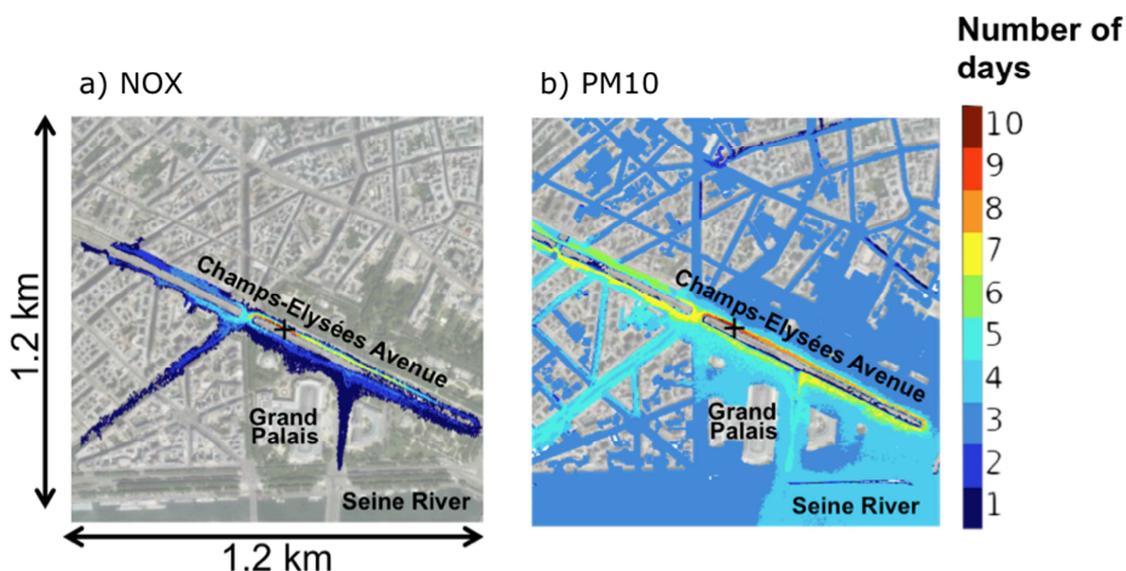
238

| Pollutants | ρ | NRMSE |
|------------------|--------|-------|
| NO _x | 0.7 | 0.45 |
| PM ₁₀ | 0.75 | 0.3 |

239 **Table2: Thresholds of representativeness criteria: correlation coefficient (ρ)**
 240 **and Normalized Root Mean Square Error (NRMSE) for NO_x and PM_{10}**

241 The retained thresholds for PM_{10} are more restrictive than for NO_x . This is due to the
 242 higher spatial variability of NO_x concentrations, which requires less restrictive
 243 thresholds in order to retain areas large enough to get beyond the pseudo-station.
 244

245 This method is applied independently each day of the simulation to select the
 246 representative pixels. A representativeness probability is assigned to each pixel,
 247 depending on the frequency (i.e. number of days out of ten) at which each pixel is
 248 selected. An example of representativeness areas around the monitoring site of
 249 “Avenue des Champs-Élysées” is given in Fig.2. This monitoring station is located on
 250 the sidewalk of a wide urban boulevard. NO_x most representative area includes the
 251 sidewalk where the instrument is located along a distance of 500 m (yellow to red
 252 color, indicating that this area is selected at least 7 days), whereas the dark blue
 253 area shows street portions which have been selected only one day. The much larger
 254 PM_{10} area among the most probable representative ones (yellow to red, 7 days at
 255 least) includes both sidewalks along a distance of 700 m (boulevard portion including
 256 the station) as well as a 400-meters segment of sidewalk in northwestern direction
 257 from the station.



258 **Fig.2: NO_x (a) and PM_{10} (b) representativeness areas cumulated over the 10**
 259 **days of the study for ELYS monitoring site. The black cross indicates the**
 260 **location of the monitoring site.**
 261

262
 263
 264

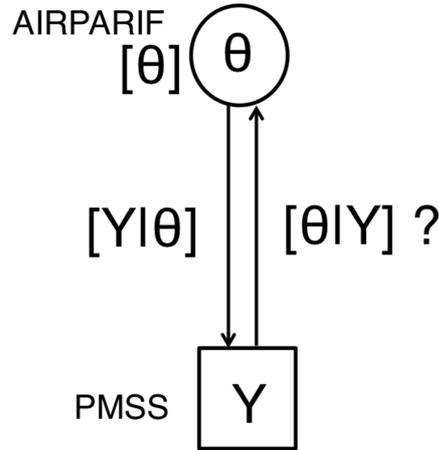
265 **2.4 The Bayesian framework**

266 **2.4.1 Model formulation**

267
 268
 269

The goal of this study is to improve the PMSS pollutant simulations at, and around,
 each AIRPARIF station by using past and current station measurements. The

270 scheme of the Bayesian model applied at the monitoring station location is
 271 represented with the Directed Acyclic Graph (DAG) of **Fig.3**.



272 **Fig.3: Directed Acyclic Graph (DAG) describing our scientific question with $[\theta]$:**
 273 **prior distribution, $[Y|\theta]$: likelihood, $[\theta|Y]$: posterior distribution.**
 274

275
 276 Past AIRPARIF measurements represent the information on which the prior expert
 277 knowledge of the unknown parameter θ (i.e. pollutant concentration at a given point
 278 in space and time) is based (before getting the predicted data from the PMSS
 279 simulated concentration).
 280

281 After integrating data information from the PMSS simulation, the prior information
 282 (past measurements from the station) is updated through the likelihood function to
 283 obtain the posterior knowledge. The likelihood function is the probability distribution
 284 of the data (PMSS simulations) conditional on the parameter θ (AIRPARIF
 285 measurements to the monitoring station), commonly used in other statistical
 286 approaches. Simple linear regression is used as a probabilistic model (PMSS
 287 simulated concentrations vs AIRPARIF measurements, **Eq.1**).
 288

$$289 \quad Y_i = a\theta_i + b + \sigma \epsilon_i \quad (\text{Eq.1})$$

290 with error term $\epsilon_i \sim N(0,1)$
 291

292 where Y_i is the initial PMSS simulations ($\mu\text{g}/\text{m}^3$), θ_i is the AIRPARIF monitoring
 293 station measurements ($\mu\text{g}/\text{m}^3$) during the ten days in March 2016 and a , b and σ
 294 respectively the slope, intercept, and the standard deviation of the residuals of the
 295 linear regression. Parameter values (i.e. a high slope, a low intercept, and standard
 296 deviation) for which the likelihood is high, are those that have a high probability of
 297 producing the observed data.
 298

299 Bayes'rule (**Eq.2**) updates the prior knowledge by learning from data (Y:PMSS
 300 simulated concentration) and provides a posterior distribution from which the *a*
 301 *posteriori* most probable concentration is estimated, associated with its uncertainty
 302 (the posterior standard deviation) (O'Hagan, 2008).
 303

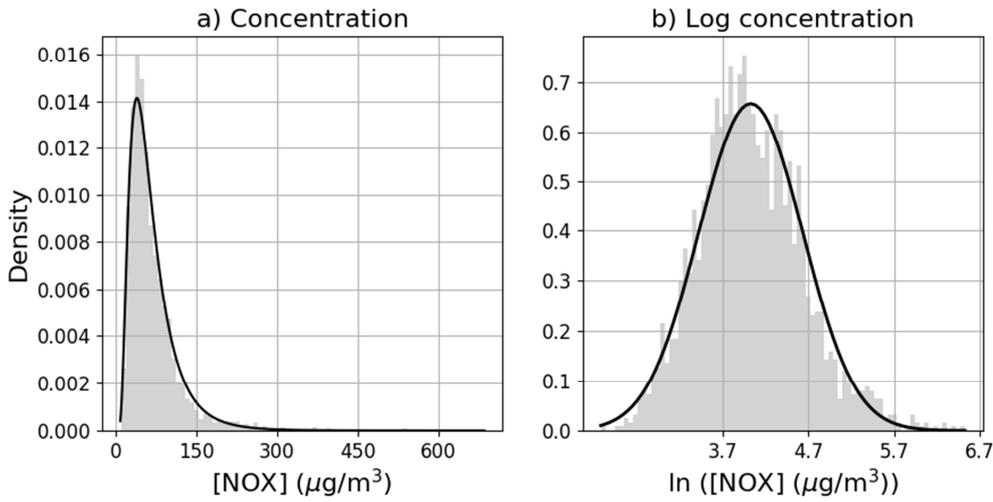
304 All three, prior, likelihood and posterior are distributions describing the probability
 305 that an event occurs. According to the Bayes'rule, the posterior distribution $[\theta|Y]$ is

306 the product of the prior distribution $[\theta]$ and the likelihood function $[Y|\theta]$, except for a
 307 constant scaling factor (Eq.2).

$$308 \quad [\theta|Y] = \frac{[Y|\theta][\theta]}{[Y]} \quad (\text{Eq.2})$$

309 Where $[\theta|Y]$ is the posterior distribution, $[Y|\theta]$ is the likelihood, $[\theta]$ prior
 310 distribution, $[Y]$ a constant of normalization that we can omit to formulate Bayes'rule
 311 by saying that the posterior is proportional to the prior times the likelihood.

312
 313 The distribution of NO_x and PM_{10} concentrations are close to log-normal (Fig.4a).
 314 Neperian logarithm concentrations (Fig.4b) are used instead of concentrations in
 315 order to use normal distributions, which are convenient to get an explicit posterior
 316 distribution, following the *a posteriori* formula of the conjugate normal model
 317 (O'Hagan, 2008) (Eq.3).



318
 319 **Fig.4: NO_x prior mean concentration of Champs-Élysées avenue monitoring**
 320 **site over ten years (2006 to 2015) at 1 AM during Monday and Friday: (a)**
 321 **concentration and (b) Neperian logarithm concentration.**

322
 323 The posterior distribution is a normal distribution with a mean ($m_{y-\log}$) and a
 324 standard deviation ($s_{y-\log}$) obtained from the Bayesian updating formulae Eq.3.
 325 Model precision is calculated as the inverse of the variance of the *posterior*
 326 distribution ($\frac{1}{s_{y-\log}^2}$). It reflects the degree of uncertainty related to the obtained mean
 327 of the estimated parameter.

328

$$\frac{m_{y-\log}}{s_{y-\log}^2} = \left(\frac{m_{o-\log}}{s_{o-\log}^2} + \frac{(Y - b) a^2}{a \sigma^2} \right)$$

329

(Eq.3)

$$\frac{1}{s_{y-\log}^2} = \frac{1}{s_{o-\log}^2} + \frac{a^2}{\sigma^2}$$

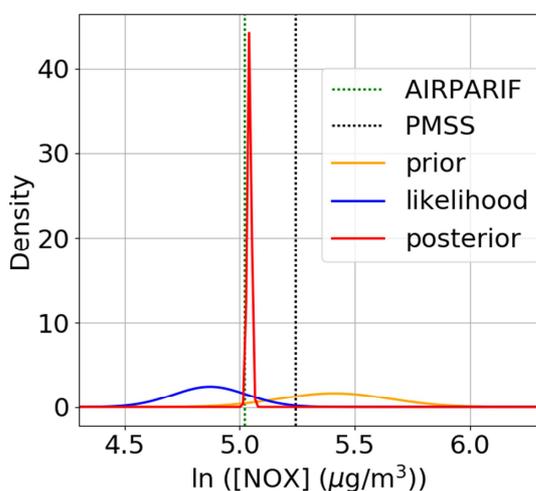
330

331 Where m_{o-log} and s_{o-log}^2 are respectively the mean and variance of the prior log
 332 concentration at the monitoring site's location, Y is the Neperian logarithm of the
 333 initial PMSS simulated concentration at a given pixel, a , b and σ^2 are respectively
 334 the slope, the intercept, and the variance of the residuals of the linear regression
 335 model used as likelihood function (see Section 2.4.2).
 336

337 At the end of the processing, the mean and standard deviation of the posterior
 338 distribution of NO_x and PM_{10} concentration are obtained by applying a log-normal
 339 transformation to the model output.
 340

341
 342 **Fig.5** shows an example of the application of Bayes'rule for NO_x log-concentration
 343 with concentration expressed in $\mu\text{g}/\text{m}^3$ at the OPERA monitoring station at 3 PM on
 344 March 2th. It is shown that the initial belief (prior) lies within a value ranging between
 345 4.8 and 6 (Neperian logarithm concentration) with a maximum weight of 5.4. The
 346 likelihood values range between 4.4 and 5.3 with a maximal weight of 4.9. The
 347 posterior distribution is the combination between prior and likelihood. It is a peaky
 348 distribution centered around 5, which appears as a compromise between likelihood
 349 and prior information. In this case, data (likelihood) are more informative than the
 350 prior because the variance of the prior distribution is high. The posterior mean
 351 properly fits the AIRPARIF measurements whereas the PMSS mean initial simulated
 352 concentration was biased, at 5.2.

353
 354
 355



356
 357 **Fig.5: Example of Bayesian update of the NO_x log-concentration at the Opera**
 358 **crossroad monitoring station at 3 PM on March 2, 2016.**

359
 360
 361

2.4.2 Finding the optimal probabilistic model at the monitor location

A major challenge when using Bayesian statistics is to properly choose the prior distribution $[\theta]$ and the likelihood function $[Y|\theta]$.

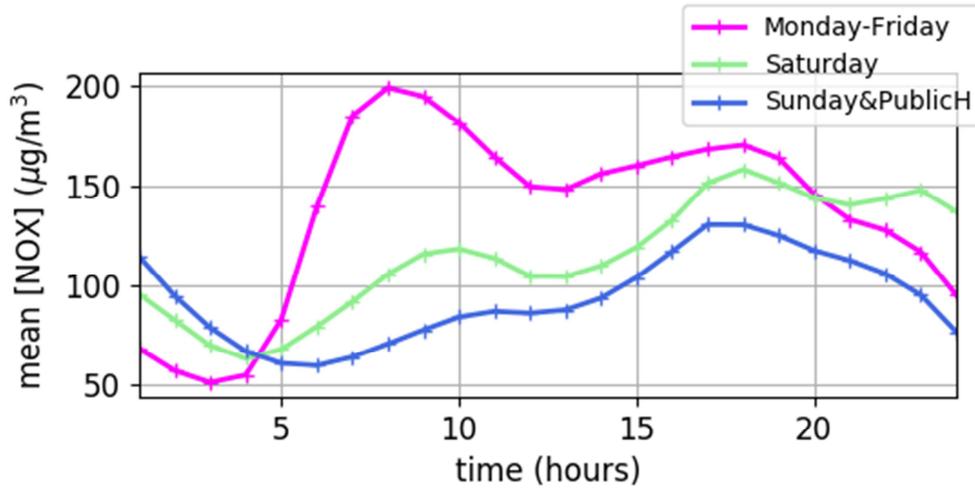
Prior distribution $[\theta]$

The prior is a distribution of possible values of the unknown parameter θ . It describes the expert knowledge before the observation. In the present case, to build a prior, we must answer the question: which data classification is the best source of information, to predict hourly NO_x and PM_{10} concentration during the ten days of March 2016 at any monitoring station location?

NO_x and PM_{10} concentrations in the center of Paris are strongly correlated to traffic emissions. They also depend on meteorological conditions responsible for their transport and mixing. Consequently, two priors are tested, both based on data collected at the AIRPARIF station in past years: (1) the first one based on traffic variations, by classifying concentration data along three classes: working days, Saturdays, Sundays and public holidays; (2) the second one based on weather conditions, by distinguishing low and high-pressure days, as well as fast-flow situations. Because of the high spatial and temporal variability within the studied domain, different priors for each station and time of the day must be considered.

AIRPARIF monitoring stations have been operating starting from different years: 2006 for PA13, PA12, PA07 and ELYS stations, 2011 for OPERA and PA04C, 2010 for HAUS and 2014 for PA15L.

Fig.6 shows daily mean profiles of NO_x concentration based on ten years of data (2006-2015) at the ELYS monitoring station. The diurnal cycle of mean concentrations is marked with a morning and afternoon peak corresponding to rush hours during working days. During Saturday, Sunday and public holidays, the night peak is more marked than the morning peak which is also shifted later on. The prior with weather classification (not shown) gives similar results for PM_{10} , but does not work properly for NO_x . The prior classification which best predicts NO_x and PM_{10} hourly concentrations for the ten days in March 2016 is, therefore, a classification per day-of-week.



397
398 **Fig.6: Daily profiles of mean concentration for NO_x prior at the Champs-**
399 **Élysées Avenue monitoring site. The magenta line represents working days**
400 **data (Monday to Friday), the green line represents Saturday data and blue line**
401 **Sunday and Public holidays data.**

402
403 **Likelihood function [Y|θ]**

404 The second challenge is the choice of the appropriate likelihood function. This is the
405 probability distribution of the data (PMSS simulation) conditional to the unknown
406 parameter θ (AIRPARIF measurements from the monitoring station), commonly used
407 in most statistical approaches. Through this function, the knowledge about the
408 pollutant concentrations (AIRPARIF measurements from the monitoring station
409 during the studied period) is updated, based on new information (PMSS simulation).

410
411 A simple linear regression of PMSS simulated concentrations vs. AIRPARIF
412 measurements is used here as a probabilistic model. A strong constraint is imposed
413 by the small size of our dataset (hourly concentrations during ten days) to define the
414 appropriate classification method. We couple simulated and observed concentration
415 data over time-slots so that at least fifteen points are used for the linear regression
416 (Eq.1). For each station, different time frames are tested to group the hourly data.
417 Time windows of 2, 3, 4, 6, 8 and 12 consecutive hours, and whole day are tested as
418 shown in Table 3.

419
420 A sensitivity analysis is performed for each type of likelihood and both proposed
421 priors (traffic and weather conditions). NO_x and PM₁₀ PMSS simulated concentration
422 at the monitoring station's location are compared with AIRPARIF monitoring station
423 measurements before and after the Bayesian updating. To evaluate the best
424 configuration, we calculate a percentage of Root Mean Square Error (RMSE)
425 difference in each case (Eq.4). For the most relevant choice, to complete this first
426 evaluation approach, we also calculate a percentage of standard deviation
427 difference.

428
429
$$\% \text{ RMSE difference} = \frac{(RMSE_{after \text{ bayesian updating}} - RMSE_{initial})}{RMSE_{initial}} * 100 \text{ (Eq.4)}$$

430 where $RMSE(ref, val) = \sqrt{\frac{1}{n} (\sum_{t=0}^{t=240} (ref_{(t)} - val_{(t)})^2)}$ is the root mean square error,
 431 $ref_{(t)}$ is the PM₁₀ or NO_x hourly concentration at the AIRPARIF monitoring site.
 432 For $RMSE_{initial}$, $val_{(t)}$ is the PM₁₀ or NO_x hourly PMSS simulated concentration at
 433 the monitoring station location for the ten studied days. For
 434 $RMSE_{after\ bayesian\ update}$, $val_{(t)}$ is the PM₁₀ or NO_x hourly posterior mean
 435 concentration resulting from the Bayesian updating.

436
 437 **Table 3** summarizes results of the sensitivity analysis, associating the previously
 438 chosen prior and the time frames tested to construct the likelihood function. The best
 439 data classification corresponds to the largest decrease of the %RMSE (the highest
 440 negative value), meaning that the Bayesian updating provides less biased
 441 concentration estimates compared to the initial PMSS simulation. Note that we only
 442 focus on the evaluation of the most probable concentration (posterior distribution
 443 mean). The overall best model performance is obtained when gathering the hourly
 444 data in 2-hours time frames. Data from all ten days corresponding to each 2-hours
 445 time frame are gathered to obtain twenty points for calculating the linear regression.

446
 447
 448
 449
 450
 451
 452

a) NO_x

| STATION | | Initial RMSE [Initial standard deviation (%)] | RMSE difference (%) [Standard deviation difference (%)] | | | | | | |
|-------------------------|-------|--|--|-----------------------|------|------|------|------|------|
| TYPE | ID | | Day | 2h | 3h | 4h | 6h | 8h | 12h |
| Traffic | HAUS | 60.0 [42%] | -49% | -45% [-31%] | -43% | -43% | -43% | -41% | -40% |
| | OPERA | 108.6 [50%] | -68% | -72% [-67%] | -71% | -70% | -69% | -69% | -67% |
| | ELYS | 67.8 [75%] | -27% | -29% [-21%] | -27% | -27% | -27% | -25% | -22% |
| Urban backg round | PA12 | 20.1 [24%] | -19% | -31% [-35%] | -27% | -27% | -26% | -21% | -20% |
| | PA13 | 16.6 [33%] | -13% | -26% [-22%] | -21% | -22% | -18% | -15% | -14% |
| | PA04C | 20.8 [35%] | -22% | -30% [-31%] | -25% | -25% | -28% | -21% | -21% |
| | PA07 | 25.5 [36%] | -27% | -17% [-12%] | -14% | -15% | -11% | -13% | -12% |
| | PA15L | 23.3 [45%] | -17% | -26% [-24%] | -22% | -25% | -12% | -19% | -13% |

453

b) PM₁₀

| STATION | | Initial RMSE [Initial standard deviation (%)] | RMSE difference (%) [Standard deviation difference (%)] | | | | | | |
|---------|----|--|--|----|----|----|----|----|-----|
| TYPE | ID | | Day | 2h | 3h | 4h | 6h | 8h | 12h |

| | | (%) | | | | | | | |
|-------------------------|-------|---------------|------|----------------|------|------|------|------|------|
| Traffic | HAUS* | 8.6 [23%] | 12% | -12% [-13%] | -14% | -12% | -7% | -10% | -5% |
| | OPERA | 10.5 [30%] | -30% | -38% [-49%] | -36% | -36% | -37% | -34% | -32% |
| | ELYS | 9.6 [19%] | 12% | -10% [-8%] | -8% | -8% | -5% | -7% | -4% |
| Urban backg round | PA04C | 5.1 [13%] | 4% | -1% [-8%] | 3% | 2% | 5% | 3% | 5% |
| | PA15L | 7.1 [21%] | 19% | -8% [-19%] | -3% | -4% | -4% | 1% | 3% |

454 **Table3: % RMSE and standard deviation (in brackets) difference between**
455 **simulated and observed data before and after the Bayesian updating for**
456 **several time frames of classification of the hourly data (per day, by 2,3,4,6,8**
457 **and 12-hours period for the ten days). [HAUS* - PM₁₀ results are averaged on**
458 **8/10 days, due to the lack of AIRPARIF measurements for two days].**

459
460
461
462

463 2.4.3 Expanding the measured value within the representative area

464
465 We applied the previous Bayesian model, for each monitoring station and each hour
466 of the day, to each pixel of the representativeness areas (see Section 2.3), by
467 accounting for the probability of each pixel to belong to this area.

468
469 The Bayesian model update is performed on every pixel within the
470 representativeness area, using the initial PMSS simulated concentration at the given
471 pixel and with the selected prior and likelihood function at the station (Section 2.4.2).
472 The fraction of the number of days, for which the pixel is selected to belong to the
473 representativeness areas on the total number of days of the study, is used as
474 weighting coefficient (Eq. 5). Thus, for pixels within a representativeness area
475 selected ten days out of ten, we apply a weight equal to one. In this case, the
476 concentration at the pixel is the posterior mean concentration at the given pixel (i.e.
477 concentration after the Bayesian updating). For pixels selected only one day out of
478 ten the final concentration will be calculated as the sum of 1/10th of the posterior
479 mean concentration at the pixel and 9/10th of the initial PMSS simulated
480 concentration at the pixel.

481

$$482 C_{final} = a * C_{after\ bayesian\ updating} + (1 - a) * C_{initial} \text{ (Eq5)}$$

483

484 where C_{final} is the new concentration estimate at the pixel, a the weighting
485 coefficient, $C_{after\ bayesian\ updating}$ the enhanced concentration estimate at the pixel
486 and $C_{initial}$ the PMSS simulated concentration at the pixel.

487

488 We note here that the weighting coefficient used to expand the measured value in
489 the representativeness areas depends on representativeness criteria statistics (daily
490 correlation coefficient and NRMSE) calculated from the ten-day period of the study.

491 Given the shortness of the time period and the random nature of meteorology, we
 492 expect to underestimate and/or misplace the daily variability of the concentration
 493 field. However, by taking the most probable representativeness area
 494 (representativeness area shared seven days on ten) instead of daily
 495 representativeness area, the uncertainty on daily representative area and thresholds
 496 effects are minimized.

497

498 The methodology should be applied on longer study periods covering a larger variety
 499 of atmospheric conditions, to reduce uncertainties. Representativeness areas could
 500 then be defined based on dispersion pattern regimes. Alternative approaches include
 501 geostatistical kriging such as in Beauchamp *et al.* (2018c).

502

503 3 Results

504

505 3.1 Pollutant concentration estimates at the monitoring site location

506

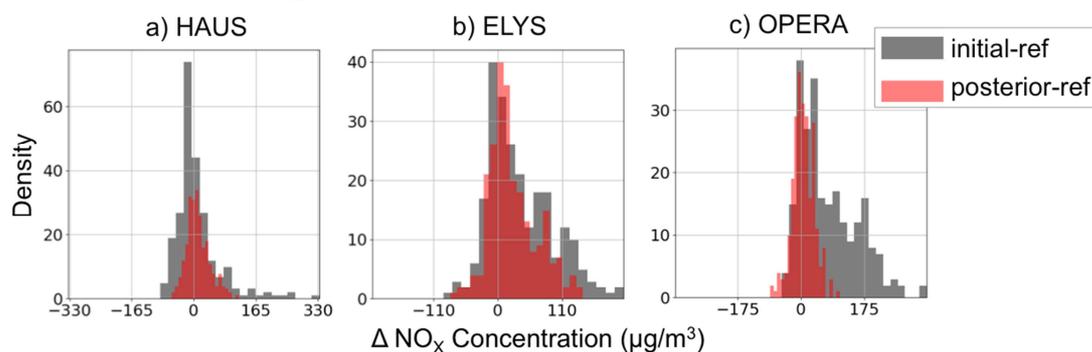
507 As shown in **Table 3**, concentration estimates by Bayesian updating present in
 508 almost all cases a lower RMSE than the initial PMSS simulated hourly
 509 concentrations. For both NO_x and PM_{10} pollutants, this improvement is higher for
 510 traffic stations (up to 72% decrease in RMSE at the OPERA monitor for NO_x) than
 511 for urban stations (31% at PA12 for NO_x). As shown in the histograms of **Fig.7**, the
 512 distribution of the PMSS simulation bias with respect to the monitoring station is not
 513 normal (grey histograms). This suggests the presence of systematic errors in the
 514 PMSS simulation. Especially for traffic stations, and for both pollutants, we find
 515 extreme values, suggesting high model overestimation of the measured value. After
 516 the Bayesian updating, the bias distribution is closer to a normal distribution with
 517 errors centered around zero. This shows that the applied methodology is an efficient
 518 bias-correction method for the PMSS model.

519

520

521

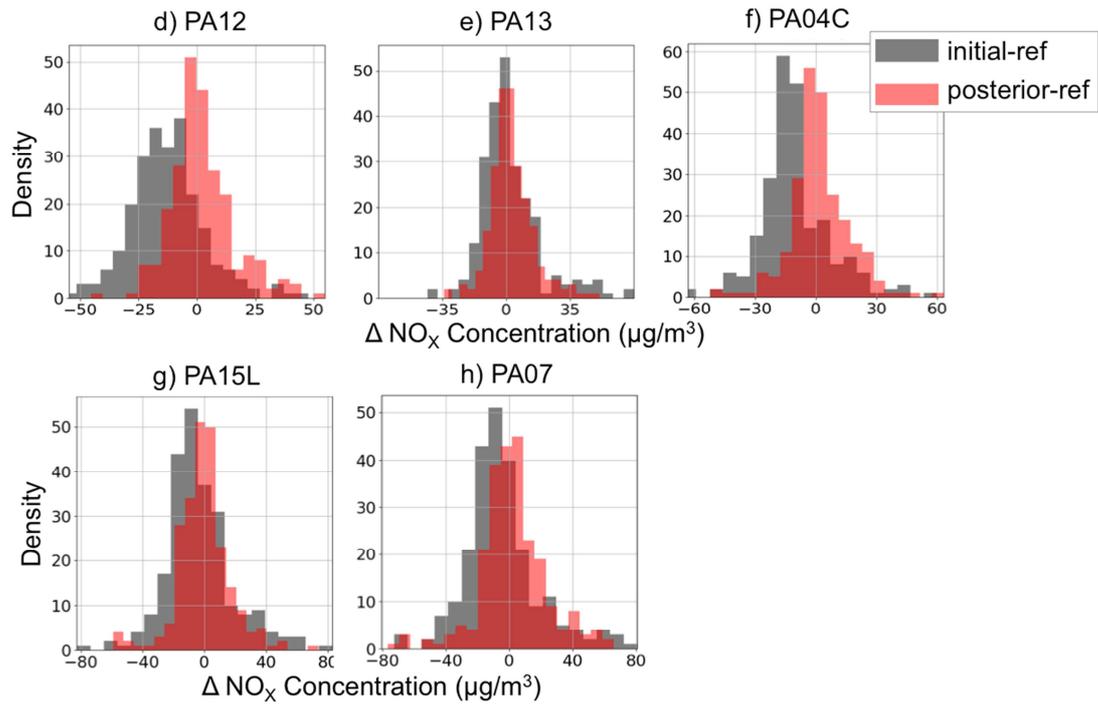
521 NO_x traffic stations



522

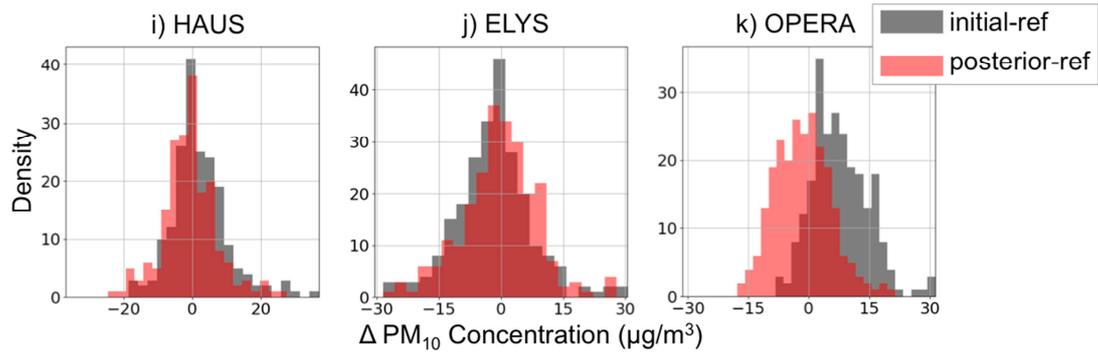
523

523 NO_x urban background stations



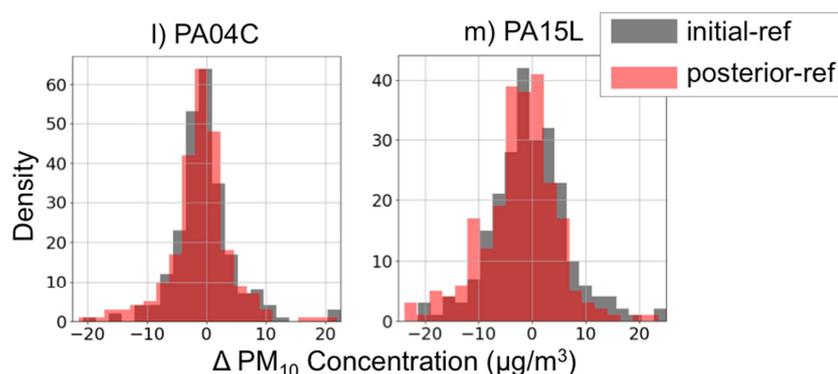
524
525
526

PM₁₀ traffic stations



527
528
529
530
531
532
533

PM₁₀ urban background stations



534
535 **Fig.7: Distribution of the bias of initial PMSS simulated concentrations (grey)**
536 **and of the posterior mean concentration (red), with respect to the monitoring**
537 **station measurements.**

538

539 **3.2 Pollutant concentration estimates in the representativeness areas**

540

541 The same Bayesian model (same prior and likelihood) is applied to the pixels of the
542 representativeness areas in the vicinity of the monitoring station to update NO_x and
543 PM_{10} PMSS simulated concentrations. To account for the probability of each pixel to
544 belong in the representativeness area, the weighting function is applied to the
545 posterior distributions to provide the final concentration estimates (see Section
546 2.4.3).

547

548 Maps of the differences between the updated concentration estimates and the initial
549 PMSS simulation within the representativeness areas are shown in **Fig.8** for NO_x
550 and PM_{10} pollutants respectively. For NO_x , around traffic stations, a small number of
551 pixels are modified, due to the small size of the representativeness area.
552 Changes can be seen in larger areas around the urban stations. The change in PM_{10}
553 concentrations is smaller than in NO_x concentrations with a maximal correction of
554 about 15% for PM_{10} vs. 50% for NO_x . As mentioned in Section 2.2, this is most
555 probably due to the better PMSS model performance for PM_{10} than for NO_x .
556 Moreover, for NO_x , around traffic stations, we only find small changes in the initial
557 PMSS concentration (by 20-30% in average) whereas around urban background
558 stations, up to 50% changes are observed on large areas for PA12 and PA04C. For
559 PM_{10} , the highest modification is obtained for PA15L with 15% change and around
560 the three traffic stations within a small-sized area surrounding the station.

561

562

563

564

565

566

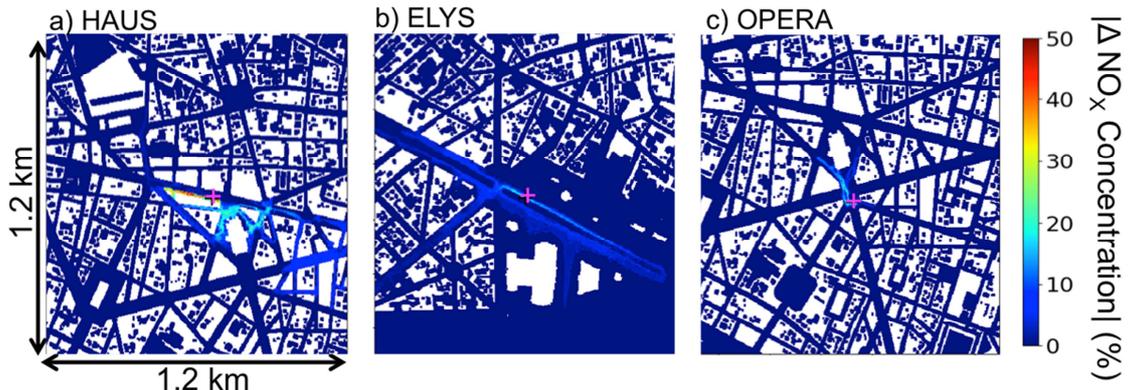
567

568

569

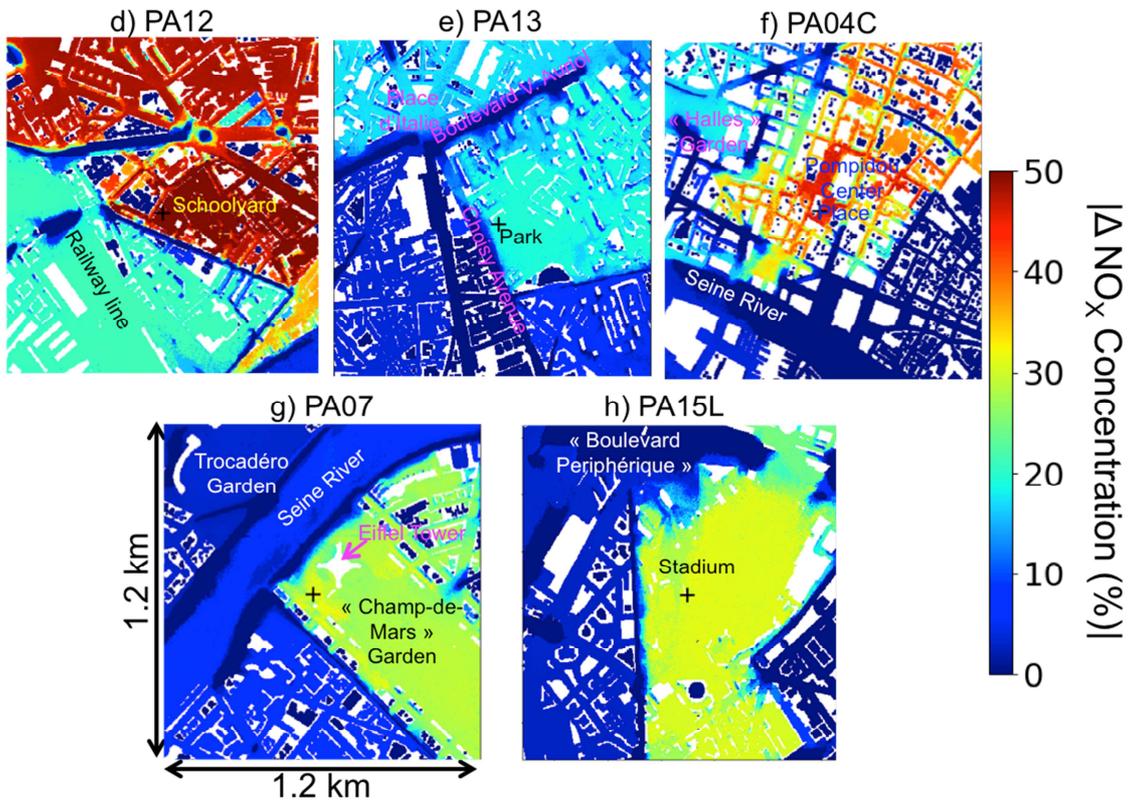
570 **NO_x traffic stations**

571



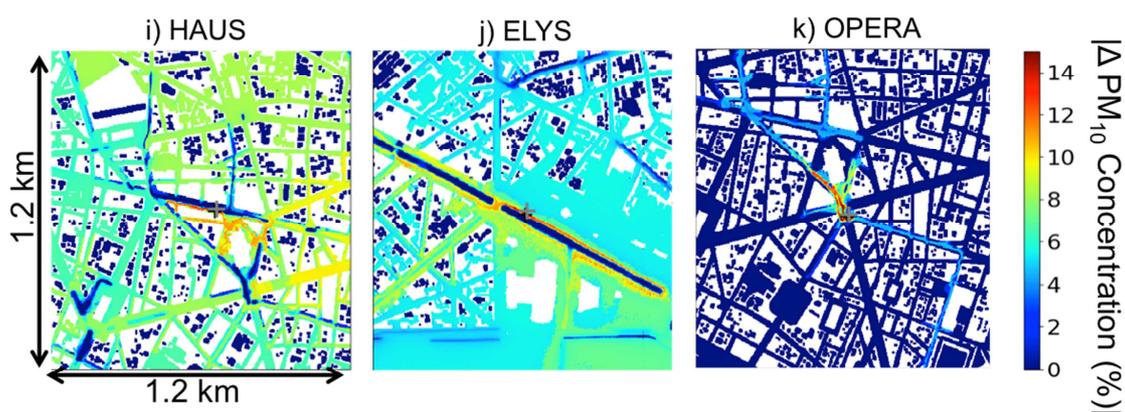
572
573
574
575

NO_x urban background stations

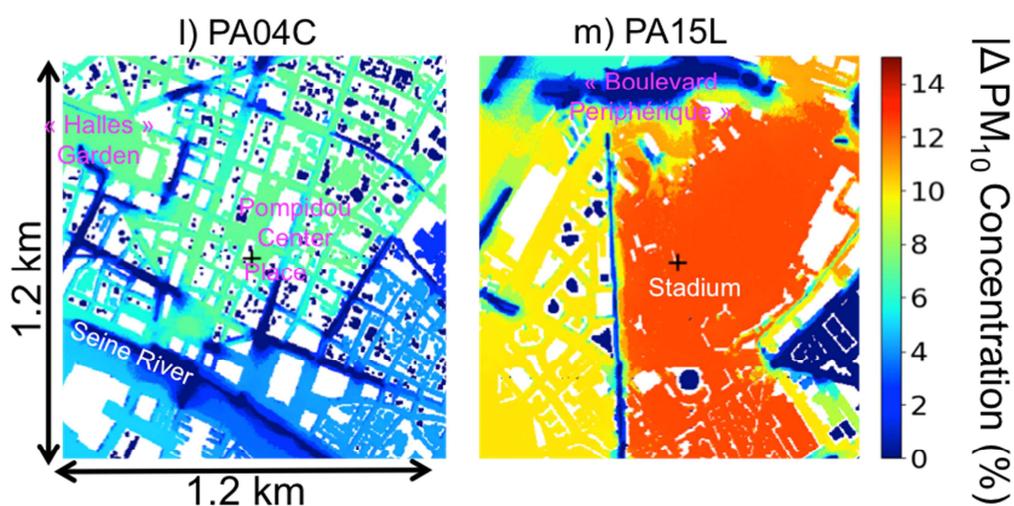


576
577
578
579
580
581
582
583
584
585
586

PM_{10} traffic stations



587
588 **PM₁₀ urban background stations**
589



590
591 **Fig.8: Percentage difference in PM₁₀ concentrations between the updated**
592 **estimate and the initial PMSS simulation averaged over the ten days of the**
593 **study. Crosses indicate the location of the monitoring sites.**

594
595 **Table4** shows results obtained within the most probable representativeness area
596 around each monitoring station, i.e. the area selected at least seven days out of ten
597 (see Rodriguez *et al.*, 2019). The results are given, as in Section 2.4.2, in terms of
598 percentage decrease of the RMSE and standard deviation values between the
599 updated concentration estimate and the initial PMSS simulated concentration with
600 respect to the AIRPARIF measurement (**Eq.4**). RMSE and standard deviation shown
601 in **Table4** are spatially averaged across the pixels of the most probable
602 representativeness areas.

603
604 NO_x concentrations after Bayesian updating are improved in all cases, with a
605 maximal correction for the OPERA station (-46% for RMSE difference and - 39% for
606 the standard deviation difference).

607 PM₁₀ concentrations are also closer to the measured values after Bayesian updating
 608 within the most probable representativeness area by considering RMSE and
 609 standard deviation (**Table4.b**).

610

611

a) NO_x

| TYPE STATION | ID STATION | Initial RMSE [Initial standard deviation (%)] | RMSE difference (%) [Standard deviation difference (%)] |
|------------------|------------|--|--|
| Traffic | HAUS | 72.0 [48%] | -37% [-33%] |
| | OPERA | 105.5 [47%] | -46% [-39%] |
| | ELYS | 64.0 [71%] | -20% [-14%] |
| Urban background | PA12 | 20.4 [34%] | -21% [-25%] |
| | PA13 | 18.5 [36%] | -21% [-19%] |
| | PA04C | 25.3 [37%] | -17% [-22%] |
| | PA07 | 26.6 [36%] | -19% [-22%] |
| | PA15L | 24.6 [48%] | -22% [-19%] |

612

b) PM₁₀

| TYPE STATION | ID STATION | Initial RMSE [Initial standard deviation (%)] | RMSE difference (%) [Standard deviation difference (%)] |
|------------------|------------|--|--|
| Traffic | HAUS | 25.3 [52%] | -8% [-12%] |
| | OPERA | 10.0 [26%] | -18% [-24%] |
| | ELYS | 10.5 [21%] | -12% [-13%] |
| Urban background | PA04C | 5.5 [13%] | -5% [-8%] |
| | PA15L | 7.2 [21%] | -10% [-18%] |

613

614

615

616

Table4: Average RMSE, standard deviation and % RMSE, and standard deviation differences between simulated and observed NO_x (a), and PM₁₀ (b) concentrations, before and after Bayesian updating, spatially averaged within the most probable representativeness area, for the ten days of the study.

617

618

619

620

4 Conclusion

621 In this study, our main goals are (1) to improve NO_x and PM_{10} simulated
622 concentrations over AIRPARIF monitoring stations and (2) to extend this enhanced
623 concentration estimate within representative areas at the vicinity of the station.

624 We show that the bias in the ten-day PMSS simulation with respect to the AIRPARIF
625 measurements does not follow a normal distribution. Uncertainties due to the
626 emission inventory, meteorological conditions and model parameterizations lead to
627 systematic errors. Bayesian statistics is especially appropriate to handle model
628 uncertainties and provide bias correction in such cases.

629 The proposed Bayesian model combines PMSS model simulations with current and
630 past surface pollutant concentration measurements of the AIRPARIF stations (3
631 traffic-stations, 2 and 5 urban background stations respectively for PM_{10} and NO_x).
632 Combination of these two sources of information results in PMSS model error
633 reduction at the station location and provides a spatially resolved concentration
634 estimate in the vicinity of the monitoring site.

635
636 The most probable NO_x and PM_{10} concentrations at the monitoring station location
637 are given by the Bayesian posterior distribution. A sensitivity analysis is performed to
638 find the optimal probabilistic model at each station, and determine the parameters of
639 the prior and likelihood distributions. Past hourly AIRPARIF pollutant concentration
640 measurements are used to establish the prior distribution for each monitoring station.
641 Two different data classifications, based either on the intensity of traffic circulation or
642 meteorological conditions, are tested. The classification per weekday, accounting for
643 traffic circulation appears more appropriate. The hourly data were grouped in several
644 time frames for the linear regression in order to test different likelihood functions. The
645 two-hour period likelihood is found to give the best results by reducing RMSE and
646 standard deviation between simulated concentrations and AIRPARIF measurements.
647 For example, at the OPERA crossroad site, a -72% difference in the RMSE is
648 obtained for NO_x concentrations.

649
650 We spatially extend the updated concentration estimates within the
651 representativeness areas of each monitoring site, by applying a weighting function
652 that considers the probability of each pixel to belong to the area. We propose to
653 estimate the updated concentration at each pixel, by taking a weighted average
654 between the posterior mean concentration and the initial PMSS simulation at the
655 given pixel. The fraction of the number of days for which the pixel is selected to
656 belong to the representativeness areas on the total number of days of the study is
657 used as the weighting coefficient.

658 Results show that final NO_x and PM_{10} concentration estimates within the most
659 probable representative area (pixels selected seven days among ten) are closer to
660 AIRPARIF measurements than the initial PMSS simulation with a reduced error
661 (decrease of % RMSE and of % standard deviation).

662 Bias correction is larger for NO_x concentrations than for PM_{10} because the initial
663 PMSS model error is smaller for PM_{10} . Modifications are observed over larger areas
664 around the urban stations than around traffic ones due to the size of the
665 representativeness area.

666 The Bayesian model developed in this study is an innovative and low computational
667 cost method to spatially extend pollutant concentration measurements in the vicinity
668 of the station. By providing low-bias high-resolution pollutant concentration estimates
669 over urban areas, the method could contribute to a better assessment of human

670 exposure to atmospheric pollution. This method should be further validated, by
671 performing local measurements inside representativeness areas.

672

673 Finally, longer PMSS simulations would increase the available dataset for the linear
674 regression, leading to more robust likelihood functions. In particular, the impact of
675 the specific air pollution episode with a sharp PM₁₀ increase in the whole region
676 would be attenuated in favor of more general statistics reflecting the baseline
677 conditions.

678

679

680 **Acknowledgments:**

681 Acknowledgments to ARIA Technologies for providing the PMSS simulations for ten
682 days.

683 **References / bibliography**

684

685 Amin, N.A.M., Adam, M.B., Aris, A.Z., 2015. Bayesian Extreme for Modeling High
686 PM10 Concentration in Johor. *Procedia Environmental Sciences, Environmental*
687 *Forensics* 2015 30, 309–314. <https://doi.org/10.1016/j.proenv.2015.10.055>

688

689 Beauchamp, M., de Fouquet, C., Malherbe, L., 2017. Dealing with non-stationarity
690 through explanatory variables in kriging-based air quality maps. *Spatial Statistics* 22,
691 18–46. <https://doi.org/10.1016/j.spasta.2017.08.003>

692

693 Beauchamp, M., Malherbe, L., de Fouquet, C., and Létinois, L, 2018. A necessary
694 distinction between spatial representativeness of an air quality monitoring station and
695 the delimitation of exceedance areas. *Environmental Monitoring and Assessment*,
696 190(7):441, Jun 2018c. ISSN 1573-2959. [https://doi.org/10.1007/s10661-018-6788-](https://doi.org/10.1007/s10661-018-6788-y)
697 *y*.

698

699 Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., Vounatsou, P., 2018.
700 Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in
701 Europe using high-resolution satellite-derived products. *Environment International*
702 121, 57–70. <https://doi.org/10.1016/j.envint.2018.08.041>

703

704 Boreux, J.J., Parent, E., Bernier, J., 2010. *Pratique du calcul bayésien, Statistique et*
705 *probabilités appliquées*. Springer-Verlag.

706

707 Cowles, M.K., Zimmerman, D.L., 2003. A Bayesian space-time analysis of acid
708 deposition data combined from two monitoring networks. *Journal of Geophysical*
709 *Research: Atmospheres* 108. <https://doi.org/10.1029/2003JD004001>

710

711 Elbern, H., and Schmidt, H., 2001. Ozone episode analysis by four-dimensional
712 variational chemistry data assimilation. *J. Geophys. Res.*, 106, No. D4, 3569-3590.
713 <https://doi.org/10.1029/2000JD900448>

714

715 Fuentes, M., Raftery, A.E., 2005. Model Evaluation and Spatial Interpolation by
716 Bayesian Combination of Observations with Outputs from Numerical Models.
717 *Biometrics* 61, 36–45. <https://doi.org/10.1111/j.0006-341X.2005.030821.x>

- 718
719 Hanea, R.G., Velders G.J. and Heemink, A., 2004. Data assimilation of ground-level
720 ozone in Europe with a Kalman filter and chemistry transport model J. Geophys. Res
721 109, D10302. doi:10.1029/2003JD004283
722
- 723 Hanna, S., White, J., Trolier, J., Vernot, R., Brown, M., Gowardhan, A., Kaplan, H.,
724 Alexander, Y., Moussafir, J., Wang, Y., Williamson, C., Hannan, J., Hendrick, E.,
725 2011. Comparisons of JU2003 observations with four diagnostic urban wind flow and
726 Lagrangian particle dispersion models. Atmospheric Environment 45, 4073–4081.
727 <https://doi.org/10.1016/j.atmosenv.2011.03.058>
728
- 729 Harrison, R.M., 2018. Urban atmospheric chemistry: a very special case for study.
730 npj Climate and Atmospheric Science 1, 5. [https://doi.org/10.1038/s41612-017-0010-](https://doi.org/10.1038/s41612-017-0010-8)
731 [8](https://doi.org/10.1038/s41612-017-0010-8)
732
- 733 Janssen, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D.,
734 Trimpeneers, E., Mensink, C., 2012. Land use to characterize spatial
735 representativeness of air quality monitoring stations and its relevance for model
736 validation. Atmospheric Environment 59, 492–500.
737 <https://doi.org/10.1016/j.atmosenv.2012.05.028>
738
- 739 Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G.,
740 Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L.,
741 Markakis, K., Meleux, F., 2017. CHIMERE-2017: From urban to hemispheric
742 chemistry-transport modeling. Geoscientific Model Development. 10. 2397-2423.
743 10.5194/gmd-10-2397-2017.
744
- 745 Martín, F., Palomino, I. and Vivanco M.G. ,2012. Combination of measured and
746 modelling data in air quality assessment in Spain. Int. J. Environment and Pollution,
747 Vol. 49, Nos. 1/2, 36-44. <https://doi.org/10.1504/IJEP.2012.049773>
748
- 749 McMillan, N.J., Holland, D.M., Morara, M., Feng, J., 2010. Combining numerical
750 model output and particulate data using Bayesian space–time modeling.
751 Environmetrics 21, 48–65. <https://doi.org/10.1002/env.984>
752
- 753 Moussafir, J., Olry, C., Nibart, M. Albergel, A., Armand, P., Duchenne, C., Mahé,
754 F.,Thobois, L., Loaëc, S., Oldrini, O., 2014. AIRCITY: A Very High Resolution
755 Atmospheric Dispersion Modeling System for Paris. American Society of Mechanical
756 Engineers, Fluids Engineering Division (Publication) FEDSM. 1.
757 10.1115/FEDSM2014-21820.
758
- 759 O'Hagan, A., 2008. The Bayesian approach to statistics. In Rudas, T. *Handbook of*
760 *probability: Theory and applications* (pp. 85-100). Thousand Oaks, CA: SAGE
761 Publications, Inc. doi: 10.4135/9781452226620
762
- 763 Pasquier, A.,and Ré, M., 2017. Considering criteria related to spatial variabilities for
764 the assessment of air pollution from traffic. Transportation Research Procedia, World
765 Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016 25,
766 3354–3369. <https://doi.org/10.1016/j.trpro.2017.05.210>
767

- 768 Pirani, M., Gulliver, J., Fuller, G.W., Blangiardo, M., 2014. Bayesian spatiotemporal
769 modelling for the assessment of short-term exposure to particle pollution in urban
770 areas. *J Expo Sci Environ Epidemiol* 24, 319–327.
771 <https://doi.org/10.1038/jes.2013.85>
772
- 773 Rodriguez, D., Valari, M., Payan, S., Eymard, L., 2019. On the spatial
774 representativeness of NO_x and PM₁₀ monitoring-sites in Paris, France. *Atmospheric*
775 *Environment: X*. <https://doi.org/10.1016/j.aeaoa.2019.100010>
776
- 777 Sahu, Sujit K., and Kanti V. Mardia. A Bayesian Kriged Kalman Model for Short-
778 Term Forecasting of Air Pollution Levels. *Journal of the Royal Statistical Society.*
779 *Series C (Applied Statistics)* 54, no. 1, 2005: 223-44.
780 <http://www.jstor.org/stable/3592609>
781
- 782 Son, Y., Osornio-Vargas, Á.R., O'Neill, M.S., Hystad, P., Texcalac-Sangrador, J.L.,
783 Ohman-Strickland, P., Meng, Q., Schwander, S., 2018. Land use regression models
784 to assess air pollution exposure in Mexico City using finer spatial and temporal input
785 parameters. *Science of The Total Environment* 639, 40–48.
786 <https://doi.org/10.1016/j.scitotenv.2018.05.144>
787
- 788 Trini Castelli, S., Armand, P., Tinarelli, G., Duchenne, C., Nibart, M., 2018. Validation
789 of a Lagrangian particle dispersion model with wind tunnel and field experiments in
790 urban environment. *Atmospheric Environment* 193, 273–289.
791 <https://doi.org/10.1016/j.atmosenv.2018.08.045>
792
- 793 Wu, C.-D., Zeng, Y.-T., Lung, S.-C.C., 2018. A hybrid kriging/land-use regression
794 model to assess PM_{2.5} spatial-temporal variability. *Science of The Total*
795 *Environment* 645, 1456–1464. <https://doi.org/10.1016/j.scitotenv.2018.07.073>