HAL open science

# Enhancing air quality forecasts by geomatic downscaling: an application to daily PM10 concentrations in France

Daniel Joly, Daniel Gilbert, Maria Diaz-De-Quijano, Mohamed Hilal, Mathieu Joly, Nadine Bernard

## To cite this version:

Daniel Joly, Daniel Gilbert, Maria Diaz-De-Quijano, Mohamed Hilal, Mathieu Joly, et al.. Enhancing air quality forecasts by geomatic downscaling: an application to daily PM10 concentrations in France. Theoretical and Applied Climatology, 2021, 143 (1), pp.327-339 (2021). 10.1007/s00704-020-03418-7 . hal-02973090

**HAL Id: hal-02973090**

**https://hal.inrae.fr/hal-02973090v1**

Submitted on 28 Feb 2023

**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.

Manuscript Click here to access/download;Manuscript;PM10-FRANCE_2019-07-19.docx

Click here to view linked References



# Enhancing air quality forecasts by geomatic downscaling: an application to daily $PM_{10}$ concentrations in France

Daniel Joly[1*], Daniel Gilbert[2], Maria Diaz-de-Quijano[1], Mohamed Hilal[3], Mathieu Joly[4]
Nadine Bernard[2,1]

1 Laboratoire ThéMA, UMR 6049 CNRS, Université Bourgogne Franche-Comté, Besançon, France

2 Laboratoire Chrono-Environnement, UMR 6249 CNRS, Université Bourgogne Franche-Comté, Besançon, France

3 Laboratoire CESAER, UMR 1041 INRA, AgroSup Dijon, Université Bourgogne Franche-Comté, Dijon, France

4 Centre National de Recherches Météorologiques, UMR3589, CNRS–Météo-France, Toulouse, France

* Corresponding author: Daniel Joly. e-mail: daniel.joly@univ-fcomte.fr

**Abstract**

Public health institutions need high-resolution next-day forecasts so they can order appropriate measures when there is a risk of air pollution exceeding regulatory thresholds. The MOCAGE model, the chemistry transport model developed by Météo-France, forecasts hourly surface $PM_{10}$ concentrations at a resolution of 0.1° throughout France (7.6 km). To obtain more efficient forecasts, a downscaling method is applied using topographic data (250 m resolution) and inventory data (2.2 km). All these disparate inputs are spatially standardized in a geographical information system to construct continuous daily fields at 250 m resolution. This method is suitable for large territories with widely varying environments (mountains, lowlands, coastlines, urban areas, etc.) and areas with a low density of monitoring stations. The parameters used to improve MOCAGE forecasts are derived from "global" and "local" regressions describing the links between the daily $PM_{10}$ concentration averages collected at 325 monitoring stations and seven explanatory variables (three topographic and four emission-inventory variables). One of the main results shows that the topographic and emission variables respectively explain 6% and 13% of $PM_{10}$ variance in France. Analysis by local regression accounts for 74% of the spatial variation of $PM_{10}$ concentration while the global regression accounts for 49%. The results show above all that, if the authorities responsible for human health protection had used the downscaling method instead of MOCAGE raw forecasts in 2016, they would have informed or alerted ten times as many people about the information and recommendation threshold (50 $\mu g.m^{-3}$) and alert threshold (80 $\mu g.m^{-3}$) being exceeded.

**Keywords**: particulate matter; land use regression methods; spatialization; public health.

## 1 Introduction

Due to their harmful effects on health (Bentayeb et al. 2015; Riviere et al. 2019; Meier-Girard et al. 2019), it is especially important to monitor concentrations of air pollutants and to enhance air quality forecasts. Air pollutant effects on human health vary with the duration and level of exposure (Barba-Vasseur et al. 2017; Mariet et al. 2018). Among the major air pollutants, particulate matter and ozone have the largest impact on human health at present (Pascal et al. 2013). Exposure to particulate matter (PM) with an aerodynamic diameter of 10 µm or less ($PM_{10}$) has notably been linked to respiratory and

cardiovascular diseases and related deaths (Analitis et al. 2006; Valavanidis et al. 2008; Ayres-Sampaio et al. 2014; Fischer et al. 2015; Dehghan et al. 2018; Riant et al. 2018). In this context, particle measurements and monitoring levels have featured prominently in European Legislation Directives that impose a duty on the authorities to inform the population when a certain threshold of mean annual or daily concentrations of pollutants is reached, especially in urban areas (Marco and Bo 2013; Kuklinska et al. 2015). The European and French air pollution networks result from European legislation (Directive 2008/50/EC on ambient air quality and cleaner air for Europe; French decree n°2010-1250 concerning air quality) making monitoring mandatory in densely populated areas. Moreover, Member States of the European Union are encouraged to predict air pollution events using chemistry transport models (CTMs) in order to warn the population, with a particular focus on at-risk populations. CTMs are numerical tools that simulate pollutant transport and dispersion, providing estimated concentrations for a particular area (El-Harbawi 2013) and time period (Zhang et al. 2017). Several air quality models such as HYSPLIT (Makra et al. 2013; Waked et al. 2018), Polyphemus/Polair3D (Lecoeur and Seigneur 2013), CHIMERE (Monteiro et al. 2007; Menut et al. 2013; Mailler et al. 2017; Potier et al. 2019), MOCAGE (Sič et al. 2015; Guth et al. 2016), CAMx (Nopmongcol et al. 2012; Milford et al. 2013), MSC-W (Simpson et al. 2012), and CMAQ (Appel et al. 2012) have been used to simulate $PM_{10}$ concentrations across France and Europe. These CTMs with coarse spatial resolutions (> 4 km) are typically used to analyze population exposure to pollutants in epidemiological studies, health risk assessments, and environmental damage studies (Liu et al. 2015). However, recent improvements have generated air quality models (for example HERMESv2.0) with spatial resolutions of up to 1–2 km (Baldasano et al. 2014; Guevara et al. 2013). These output results are still perfectible both in terms of resolution and quality of the estimation, for example for mountain areas, where the complexity of relief largely influences local air pollution (Gryning and Millán 2012) by causing local temperature inversions (Chemel et al. 2016; Largeron and Staquet 2016; Paci et al. 2016), or for high-density urban areas.

The satellite based data used to spatialize PM provide greater spatial-temporal coverage compared to sparse ground monitoring stations. Indeed, the PM estimations are mostly at $1 \times 1$ km spatial resolution and can thus meet the demand expressed by the health services for $PM_{10}$ estimates of good quality at high spatial resolution (Stafoggia et al. 2017; Shtein et al. 2018). However, the problem with satellite data is that they do not provide ground pollution data when the cloud cover is compact. Under these circumstances, the use of geomatic techniques applied to low-resolution CTM model outputs is another solution for downscaling CTM that has proven its effectiveness (Denby et al. 2011; Theobald et al. 2016). They are based on regressions between numerical models of $PM_{10}$ concentrations and meteorological predictors such as temperature, wind speed, boundary layer height, etc. (Stadlober et al. 2008), whose resolution is finer than that of the numerical models. The land use regression (LUR) models use air pollution measurements and land use predictor variables to estimate pollutant concentrations at unsampled locations at different temporal (day, month, year) resolutions (Hoek et al. 2008; Ryan 2007). LUR models are usually used to estimate pollutant variations at the within-city scale or at local scale (Hoek et al. 2011; Eeftens et al. 2012; de Hoogh et al. 2013; de Hoogh et al. 2016; Shahraiyni and Sodoudi 2016; Miri et al. 2019) or for isolated and sparsely populated areas (Diaz-de-Quijano et al. 2014). However, few studies have modeled broad geographic areas (European sub-domains) at high resolutions (< 0.5 km) (Vienneau et al. 2010; 2013).

The main objective of this study is to enhance $PM_{10}$ forecasts at country scale using LUR methods in order to obtain the best possible statistical modeling of the spatial variation of $PM_{10}$ at regional scale over large areas characterized by regional and local spatial heterogeneity. Another aim is to improve the information given to populations about air quality in accordance with recommended thresholds and to better predict risks for the population. Here, a LUR geomatic method based on a set of explanatory

variables has been used to estimate spatial variations of $PM_{10}$ concentrations through regressions. The findings are then applied to downscale MOCAGE, a chemistry-transport model (CTM) that forecasts mean daily $PM_{10}$ concentrations. The case study is applied to the whole of continental France at regional scale with a 250 m spatial resolution.

## 2 Material and methods

### 2.1 Data input

#### 2.1.1 Monitored $PM_{10}$ concentration

In France, the PM10 observation network measurements are managed by the air quality monitoring association network. The PM10 data used consist of daily averages calculated from data monitored at 325 stations in 2016 (366 days), using hourly averages. When there are fewer than 20 hourly measurements for a day for a given station, data for that day are considered to be missing; otherwise the daily average is calculated from the available hourly values. The stations sample urban (50%), peri-urban (10%), and rural (8%) areas; 10% and 22% are industrial and traffic stations respectively. All monitored data have been validated by the AASQA (Associations Agréées pour la Surveillance de la Qualité de l'Air - Authorized Associations for Air Quality Monitoring) and the LCSQA (Laboratoire Central de Surveillance de la Qualité de l'Air - Central Air Quality Monitoring Laboratory). Validated data are then collected by the European Environment Agency (EEA) as part of its Air Quality e-Reporting. PM10 monitoring stations are unevenly distributed (Fig. 1). Clusters appear around the main cities (Paris, Lyon, Marseille, Lille, Toulouse), while rural and mountain areas (Alps, Pyrenees, Massif Central, Jura, etc.) are not very densely sampled. For this reason, only a comprehensive study of France as a whole will be carried out.

**Figure 1**. Location of the 325 $PM_{10}$ monitoring stations in continental France

#### 2.1.2 Simulated $PM_{10}$ concentration

MOCAGE (*MOdèle de Chimie Atmosphérique à Grande Echelle* - Large Scale Atmospheric Chemistry Model) is the chemistry transport model developed by Météo-France. The model has been used for a wide range of scientific studies on tropospheric and stratospheric chemistry at various spatial and temporal scales; it simulates a lot of chemical species and, among them, $PM_{10}$ concentration. It has been applied for example for studying the impact of climate on chemistry (Teyssèdre et al. 2007; Lacressonnière et al., 2012; Lamarque et al. 2013) or tropospheric-stratospheric exchanges using data assimilation (El Amraoui et al. 2010; Barré et al. 2013). Météo-France also contributes to the Copernicus Atmosphere Monitoring Service (http://www.regional.atmosphere.copernicus .eu/). MOCAGE is one of the seven models providing input for the regional overall forecasting system for Europe (Marécal et al. 2015).

The configuration of MOCAGE used in the present study has been run daily since 2005 for air quality operational 4-day forecasts in the framework of the PREV'AIR consortium (Honoré et al. 2008; Rouïl et al. 2009, http://www2.prevair.org/). The main characteristics of the configuration are the following. Three nested domains are used, with decreasing resolutions (globe at 1°, Europe at 0.5°, and France at 0.1°). For the global and the European domain, the hourly input meteorological forcing fields are Météo-France ARPEGE forecasts. For France, and for the first day of forecast, MOCAGE uses the operational outputs of Météo-France's non-hydrostatic AROME model. Emissions are prescribed at the surface, based on state-of-the-art inventories (MACCity over the globe and TNO-MACC-III over Europe), or dynamically depending on the local meteorology (e.g., for sea salts and desert dusts). MOCAGE

simulates the concentration of gases (Josse et al. 2004; Dufour et al. 2005), primary aerosols (Martet et al. 2009; Sič et al. 2015), and secondary inorganic aerosols (Guth et al. 2016).

MOCAGE outputs are expressed in the WGS 84 geodetic system at a resolution of 0.1° in X and Y, with an average distance of 7.6 km from one point to another. The domain covered by MOCAGE extends over 10.9° in Y and 14.9° in X (West = 9.95°; East = -4.95°; North = 51.95°; South = 41.05°). Preliminary filtering has extracted the points belonging to continental France from the databases. Data refer to the 366 days of monitored data.

### 2.1.3 Explanatory variables

**$PM_{10}$ emissions register.** The inventory of $PM_{10}$ emissions, at the resolution of 0.025°, concerns the whole area covered by the monitored $PM_{10}$ concentration and MOCAGE. Data derived from data collected by the National Spatial Inventory (INS, http://emissions-air.developpement-durable.gouv.fr/) were made available to the CNRM by INERIS (*Institut national de l'Environnement et des Risques* - National Institute for Environment and Risk Management) as part of the PREV'AIR consortium. They consist of 101 696 points whose values are daily averages expressed in µg.km² that concern the 10 sectors of SNAP (Selected Nomenclature for Air Pollution) emissions (Sup. material 1). Because some sources of pollution in France (i) are related to similar activities and (ii) some of them are too scarce, some SNAPs have been grouped together to form (i) the "industry + waste treatment" type (SNAPs 1+3+4+5+6+9) and (ii) the "road" type (SNAPs 7+8). SNAPs 2 (non industrial combustion plant) and 10 (agriculture) are left unchanged (Sup. material 2). Inventory points are also projected in RGF-L93. The value of each point resulting from the inventory (2.2 km resolution) is transferred to the 81 pixels of the resulting 9 x 9 pixels window (250 m resolution). The values are then smoothed by reference to a moving average (radius of 8 pixels).

**Topographic variables.** The 250 m resolution Digital Elevation Model (DEM) provided by the French National Geographic Institute (IGN) is used to construct a large number of topographic variables (Joly et al. 2012). Preliminary Pearson correlation processing was done to select the variables that most frequently explain the spatial distribution of $PM_{10}$ concentrations at the $p<0.05$ threshold. There are three variables: elevation [elev], magnitude of the positive landforms (hump amplitude [hump]), and valley depth [valley]. The hump value corresponds to the deviation in elevation between ground elevation and the virtual surface connecting all the thalwegs. Conversely, the valley value corresponds to the deviation in elevation between the ground and the virtual surface connecting all the ridges (Joly et al. 2012). These variables are intended to describe the capacity of topographic settings to impede or facilitate particle dispersion.

## 2.2 Methods

### 2.2.1 Regressions

The available information and the processing applied to obtain data for the regressions and downscaling are presented in Figure 2. The information provided in the form of dots is transformed into continuous spatial fields by kriging or geomatic processing. Values corresponding to monitored $PM_{10}$ concentration (1.1.1 in Figure 2), simulated $PM_{10}$ concentration (1.1.2 in Figure 2), $PM_{10}$ emissions register and topographical variables (1.1.3 in Figure 2) for each of the 325 pixels where monitoring stations are located are extracted from the raster layers made up of 86 366 691 pixels. The result is three tables, all comprising 325 rows (number of monitoring stations) and either 366 days (in the case of monitored and simulated $PM_{10}$ concentration) or seven columns (explanatory variables, four for the different types of $PM_{10}$ emissions and three for the topographical variables).

To control the degree of collinearity between variables, the variance inflation factor (VIF, an index quantifying the level of collinearity of a variable when it is added to the others in a multiple regression) is calculated. A maximum VIF value of 5 (Hair et al. 2006) has been recommended. Collinearities between the seven explanatory variables are very low (<1.5), except for elevation and hump, which have VIF values of 4.1 and 3.3. These values are high, but below 5 and so remain acceptable.

**Figure 2**. Available data and their transformation for further statistical processing

Based standard ordinary least squares (OLS) regression models are then used to establish the statistical links between the explanatory variables and the monitored $PM_{10}$ concentration (Arslan and Akyürek 2018). A simple linear regression is performed for each of the 366 days between the 325 $PM_{10}$ concentrations (response variables) and each of the seven explanatory variables. Only significant variables ($p < 0.05$) are selected and included in a multiple regression (final daily model). The "global" regressions referred to above provide poor results when applied to climatologically heterogeneous areas such as France. The solution chosen to overcome this difficulty is to carry out "local" regressions (Joly et al. 2011; Diaz-de-Quijano 2016). This method is based primarily on the recognition of the *n* nearest monitored stations to each pixel according to a neighborhood criterion. With n low (e.g. n = 20), the catchment area of the stations being limited, the stations are climatologically coherent. On the other hand, the statistics are not very accurate (Joly et al., 2011). With high n (e.g. n = 100), the statistics are strengthened but the catchment area of the stations becomes very large and the probability of integrating different climatic zones increases. This poses a problem for the consistency of the models especially in rural areas where the density is 1 station per 1700 km². In this study, parameter *n* was set to 40, to provide the best empirical trade-off between statistical and geographical characteristics. Pixels characterized by the same corpus of *n* stations are grouped together in the same polygon. France is thus segmented into a total of 13,434 polygons.

Regressions provide the parameters that describe these relationships. There are two types of parameters: the intercept and the slope of each explanatory variable included in the regression. The $PM_{10}$ concentration at any point in the study area can be estimated from these two coefficients.

$$PM10_{ij} = a_1X_1 + a_2X_2 + \cdots + a_nX_n + \mathrm{b} \quad (1)$$

Where $PM10_{ij}$ denotes the monitored $PM_{10}$ concentration at point $_{ij}$,

$X_1, X_2, X_n$ = variables included in the multiple regression,

$a_1, a_2, \ldots, a_n$ = slope of variables $X_1, X_2, \ldots, X_n$,

b = intercept.

Then, the analysis by multiple regression and the application of coefficients to the pixels in order to estimate $PM_{10}$ concentrations is performed once for the global regression and as many times as there are polygons for the local regressions. The parameters obtained by multiple regression are used to compute the estimated $PM_{10}$ concentrations and their corresponding residuals are calculated by cross validation (Plutowski et al. 1994; Stone 1974). Two goodness-of-fit measures (RMSE and $R^2$ calculated using the Bravais-Pearson R) are computed to assess the performance of each model.

Upon completion of the regression analysis, 325 residuals are obtained for each of the 366 days. Joly et al. (2013) have shown that the application of an autoregressive process can reduce these residues. In time series such as daily $PM_{10}$ concentrations, Neal et al. (2014) have demonstrated that the value recorded on any day at each station is dependent on the values recorded the day before. Physical

mechanisms such as inertia account for this property of bodies that tend to stabilize or to repeat the same patterns over the course of time. This general principle holds for the climatic characteristics of a location which are, in part, dependent upon those of the previous day. The autoregressive process aims to estimate the residuals of the day d (response variables) by those of day d-1 (explanatory variable).

#### 2.2.2 Downscaling

The objective of downscaling is to improve the resolution and quality of the $PM_{10}$ fields simulated by MOCAGE. The downscaling of $PM_{10}$ concentration is based, for each of the days analyzed, on the resolution of equation (1) whose regression parameters are adopted as such, while the value of the intercept returned by the regression is replaced by the value of $PM_{10}$ from MOCAGE:

$$PM10_{ij} = a_1X_1 + a_2X_2 + \cdots + a_nX_n + PM10_{MOCAGE} \quad (2)$$

in which the variables X and a are identical to those of the regression (1) while $PM10_{MOCAGE}$ replaces b.

The $R^2$ and RMSE (Root Mean Square Error calculated by cross validation) are used to measure the quality of fit.

## 3 Results

### 3.1 Monitored $PM_{10}$ concentration overview

The national-averaged annual mean **concentration** measured in France in 2016 is 18.2 µg.m$^{-3}$ and the three quartiles (Q1 to Q3) are 13.1, 16.9, and 22.1 µg.m$^{-3}$.

Among the 366 daily averages, the lowest value was 8.0 µg.m$^{-3}$, on January 31 and the highest was 47.1 µg.m$^{-3}$ (December 1). High daily values (median > 40 µg.m$^{-3}$) occurred three times, especially in December. On the other hand, values are low from April to November with a quartile 3 rarely exceeding 30 µg.m$^{-3}$. The average daily amplitude (calculated from data from the 325 monitoring stations) is 46.6 µg.m$^{-3}$. The lowest amplitude (22.1 µg.m$^{-3}$ on May 1) reflects high spatial homogeneity of $PM_{10}$ concentrations throughout the country; however, the maximum amplitude (144.8 µg.m$^{-3}$ on December 1) reflects very high spatial heterogeneity.

The minimum and maximum of the 325*366 daily monitored values are 0.4 and 194.2 µg.m$^{-3}$ found in a rural station and in the Paris region, respectively. In 2016, 1.85% and 0.11% of the monitored data exceeded the information (50 µg.m$^{-3}$) and alert (80 µg.m$^{-3}$) thresholds for human health protection, respectively.

### 3.2 $PM_{10}$ estimation by global regression

The three most frequently selected variables in the 366 global daily multiple regressions are road-traffic emissions (94%), plant combustion emissions (59%), and elevation (57%). Then come valley depth, agricultural emissions, and magnitude of positive landforms with a frequency close to 30%. Industrial emissions are rarely significant (9%).

Estimating $PM_{10}$ concentrations by global regression provides poor results. The topographic variables and the emission variables provide low explanatory power: the $R^2$ is on average 0.17 and 0.13 respectively (Tab. 1). The multiple regressions grouping the seven variables (3 topographic + 4 emission inventory) slightly improve the scores which remain modest ($R^2$ = 0.27). Note that this value is not the sum of the previous two because of the collinearities between variables. The situation changes after the

autoregressive process: the annual average of $R^2$ is 0.49. The RMSE of the global regressions is between 8.3 and 7.7 µg.m$^{-3}$ depending on the model (tab. 1).

| | Topography (T) | Emissions (E) | T + E | T+E + autoregressive term |
|---|---|---|---|---|
| $R^2$ average - GLOB | 0.06 | 0.13 | 0.17 | 0.49 |
| $R^2$ average -LOC | 0.46 | 0.53 | 0.57 | 0.74 |

**Table 1**. Mean of the $R^2$ depending on the four local regressions steps

### 3.3 $PM_{10}$ estimation by local regressions

The hierarchy that has been described for global regressions is roughly respected with the exception of industrial emissions, which are much more frequently significant (from 9% to 24%). Conversely, topographic variables are significant in only 13% of cases. The local regressions produce excellent results: $R^2$ values reach 0.74 as an annual average when the regressions relate to the seven explanatory variables and when the autoregressive process is used. The residues are low on average (RMSE = 4.7 µg.$^{m-3}$), but, on certain days and at certain stations, they can be high: the minimum is -32.3 µg.m$^{-3}$ while the maximum reaches 67.6 µg.m$^{-3}$. The highest errors (<-10 and >10 µg.m$^{-3}$) have respective frequencies of 0.9 and 1.6%. However, 87% of the errors are between -4.9 and +4.9 µg.m$^{-3}$.

### 3.4 MOCAGE outputs and downscaling

In 2016, the correlation coefficient ($R^2$) between the monitored PM10 concentration and the PM10 concentrations simulated by MOCAGE is 0.15. In 81% of the cases, daily $R^2$ values are less than 0.2. They are higher than 0.5 in 1.4% of cases. Downscaling using the topographic and emission variables associated with the autoregressive process improves the estimate since the $R^2$ annual mean is 0.23. If 5% of the daily $R^2$ are less than 0.2, 38% are now greater than 0.5, the maximum being 0.81.

Residuals (1.6 µg.m$^{-3}$ as an annual average) indicate that the downscaling process does not produce statistical bias. They can be occasionally high: the greatest errors (< -10 µg.m$^{-3}$ and +10 µg.m$^{-3}$) have a frequency of 9.2 and 16% respectively. However, 60% of errors range from -4.9 to + 4.9 µg.m$^{-3}$. The RMSE of the downscaling estimates is 9.2 µg.m$^{-3}$.

### 3.5 Mapping of $PM_{10}$ concentration in the most polluted situation

The case study selected for $PM_{10}$ mapping is that of December 1, the most polluted day of the year 2016: the minimum, Q1, median, Q3, and maximum $PM_{10}$ concentrations observed among the 325 stations are 3.2, 31.0, 40.4, 55.0, and 148.0 µg.m$^{-3}$ respectively. The $R^2$ between the 325 monitored $PM_{10}$ concentration and the values simulated by MOCAGE is 0.46 (Fig. 3); but MOCAGE, which does not currently represent all aerosol species, returns much lower values than the monitored ones: the average is only 24.3 µg.m$^{-3}$ (as against 47.1) and the maximum is just over 56.0 µg.m$^{-3}$ (as against 148.0).

**Figure 3**. Scatter plot between the 325 monitored PM10 versus the values simulated by Mocage and the values predicted by downscaling; case of December 1, 2016

The southeast quarter of France is used to illustrate the results of the models. This regional zoom shows the value of the method in providing regional estimates of $PM_{10}$ based on national information. Indeed, health agencies have a regional scope of intervention. The $PM_{10}$ simulated by MOCAGE (Fig. 4A) shows that the highest values (≥ 30 µg.m$^{-3}$) are located around Lyon with a clear extension to the south,

along the Rhône valley and eastward to Marseille. Areas with lower pollution levels are located in the mountainous and rural areas.

After downscaling, the average downscaled values are 56.3 µg.m$^{-3}$ and the $R^2$ is 0.80 (Fig. 3). The spatialization of $PM_{10}$ concentration remains virtually unchanged in its main features (Fig. 4B). In detail, it should be noted that pollution in the main cities (Lyon, Marseille, Nice, etc.) and the Arve valley, one of the most polluted valleys in France, is prominent. The map shows more spotted areas, reflecting the contrast in emissions. The Rhône-Saône river valley (north and south of Lyon) stands out because of the road network which is taken into account in the multiple regression.

**Figure 4**: Restitution of $PM_{10}$ simulated by MOCAGE for December 1, 2016 by (A) kriging and (B) downscaling.

### 3.6 Population at risk of exceeding the 50 and 80 µg.m$^{-3}$ threshold in France

| | Mocage kriged | Mocage downscaled | LUR geomatic method |
|---|---|---|---|
| **pixels > 50 µg.m$^{-3}$ (%)** | 0.5 | 8.3 | 13.7 |
| **pixels > 80 µg.m$^{-3}$ (%)** | 0.07 | 0.7 | 3.0 |
| **Population exposed > 50 µg.m$^{-3}$** | 5 880 300 | 10 434 500 | 13 901 000 |
| **Population exposed > 80 µg.m$^{-3}$** | 2 484 400 | 4 467 200 | 7 833 400 |

**Table 2**. Percentage of PM10 exceeding the information and recommendation (50 µg.m-3) and alert (80 µg.m-3) thresholds on December 1, 2016 and number of inhabitants affected by these exceedances.

The proportion of pixels with a value above the information and alert threshold (50 and 80 µg.m-3, respectively) on December 2, 2016 is 0.5% and 0.07% with the MOCAGE raw forecast (tab. 2). It rises to 8.3% and 0.7% after downscaling. Using French 200 m gridded population data (https://www.insee.fr/fr/statistiques/2520034), the number of people to be alerted after exceeding the information and recommendation threshold of 50 µg.m-3 also depends on the model used: fewer than 5.9 million people with the MOCAGE raw forecast versus more than 10 million with the downscaling forecast. Alerts would be triggered for fewer than 2.5 million people with MOCAGE, but for almost 4.5 million with the downscaling technique. These estimates at d+1 allow decision-makers to take appropriate measures to prevent adverse health effects from exposure to PM10 (Beloconi et al. 2018).

## 4 Discussion

The LUR geomatic method used, and based on a set of explanatory variables, provides the best estimate spatial variations of $PM_{10}$ concentrations at country scale with a 250 m of resolution by using the so-called local regression with autoregression. The combination of methods (global and local regressions, autoregressions, alone or associated) and explanatory variables ($PM_{10}$ emissions register, topographic variables) makes it possible to forecast $PM_{10}$ concentrations (explained variables) with different qualities of estimates.

### 4.1 Quality of regression estimates

The global regression estimates provide low results: topography ($R^2$ = 0.17) and emission sources ($R^2$ = 0.13) account for little of the spatial variation in $PM_{10}$ concentrations across France. The global approach cannot describe phenomena that depend on local conditions (Stadlober 2008; Czarnecka and Nidzgorska-Lencewicz 2017). After autocorrelation, the estimates result in an $R^2$ and RMSE of 0.49 and 7.7 µg.m$^{-3}$. These two indicators are similar to those derived from statistical adjustments intended to correct bias in the output of numerical models. But the estimates of the latter are based on annual or seasonal averages with resolutions of several kilometers while ours are daily with a resolution of 250 m. As an example, let us mention the estimate of $PM_{10}$ concentrations in Great Britain and the Netherlands which provide $R^2$ values between 0.3 and 0.4 (Vienneau et al. 2010). In a similar experiment concerning variation in $PM_{10}$ concentrations throughout the EU in 2001, the model explains 46% of the variation in $PM_{10}$ with RMSE =5.19 µg.m$^{-3}$ (Beelen et al. 2009). At a European scale and with a resolution of 20 km, the mean daily RMSE for April and June is around 13 and 7 µg.m$^{-3}$ with values as low as 8 and as high as 20 µg.m$^{-3}$ (Hamm et al. 2015). In similar work, Denby et al. (2005) provide an optimal estimate of the $PM_{10}$ spatial distribution over Europe at 35x25 km resolution using a statistical interpolation method; the total RMSE of the mean daily concentrations of $PM_{10}$ at the validation stations is 9.2 µg.m$^{-3}$.

The situation is much better with local regressions which take into account the topographic and $PM_{10}$ emission factors at a fine scale. The latter are not stationary but vary from one place to another across France. Thus, the influence of topography is more pronounced in mountainous areas (Largeron and Staquet 2016) and that of emissions in areas where dynamic or thermal turbulence is lowest. There is thus a spatial differentiation that is well described by local regressions, but poorly captured by global regressions. There is a second problem with global regressions. Because the explanatory variables fail to describe the spatial variation of $PM_{10}$ concentrations, the (single) intercept calculated by regression is close to the mean. However, depending on atmospheric conditions, there are significant variations in particulate matter from one point to another. Another merit of the local regression is that it suitably fits the intercept to the local $PM_{10}$ concentration.

The downscaling carried out concerns the MOCAGE raw forecasts at day+1 and is part of a time scale analogous to high-resolution urban studies. It is also similar to the adaptations of the output of numerical models expressed as annual averages, which concern very large areas and most often entire countries. Downscaling improves the simulation given by CTM models. This has been reported many times in the literature (Konovalov et al. 2009; Denby et al. 2008; Hamm et al. 2015; Mok et al. 2015). It can also be seen in our experiment. The $R^2$ in downscaling (0.23) improves the score obtained from the MOCAGE raw forecasts (0.15). However, another benefit of downscaling is the increase in the MOCAGE raw forecast values, which are generally quite low, in part because organic secondary aerosols are not yet implemented in that CTM. The annual average of the MOCAGE values is 10.7 µg.m$^{-3}$ compared to 19.8 µg.m$^{-3}$ with downscaling (a value very close to the annual average of monitored data (18.4 µg.m$^{-3}$).

### 4.2 Monitored $PM_{10}$ versus simulated $PM_{10}$ (MOCAGE)

There is a large difference in the quality of the estimates obtained from local regressions applied to monitored $PM_{10}$ concentration and MOCAGE. The values provided by the numerical forecasts of the MOCAGE model are not of the same nature as the values of monitored $PM_{10}$ concentrations. The CTM predicts the average concentrations in a cubic grid of several kilometers in the plane and tens of meters in height, while the monitored data concern a point concentration, very close to the ground where the boundary layer is often established (Zhang et al. 2017), in a much more complex environment. Part of

the difference between models and monitored data is therefore due to monitored data being much more precise than models in terms of their spatial representativeness. In addition, the MOCAGE data used in this study forecast for day+1, with several sources of uncertainty, such as meteorological forcing (Gillian et al. 2015), the emissions inventory (Frey and Zhao 2004), and the representation of the physicochemical processes. In addition, in the current parameterization, pollutants are injected vertically in a decreasing manner into the first five layers of the model regardless of the weather situation. In December 2016, however, since the boundary layer was very low, the injection was partially above the boundary layer, resulting in an increased negative bias.

### 4.3 Accuracy of inputs

Insufficient knowledge of pollutant sources and emission inventories can lead to significant bias and error in air quality estimations of mechanistic models (Chang and Anna, 2004; Borrego et al. 2008). Each monitored station is influenced by many sources (agricultural, industrial, road traffic, etc.), at varying distances, which contribute, in the form of different species (minerals, nitrates, organic carbon, etc.) to measured concentrations of $PM_{10}$. It may be considered futile to attempt to precisely locate the sources influencing each station using a specific emission inventory. Depending on the processes involved in aerosol evolution (transport, sedimentation, deposition, scavenging, transformation, etc.), the radius of influence of the sources varies greatly with the type of emission, the species emitted, the weather conditions, and the local geography.

But one can also consider that a source impacts the measured concentration as a direct function of its proximity. The use of a more detailed resolution issuing register than the one used (2 km resolution) would substantially improve the estimates. In geomatics, the location of data is a crucial point. Its quality determines the quality of the results.

### 4.4 Variation of $R^2$ and RMSE according to $PM_{10}$ concentration

| | **< 15** | **15 to 19.9** | **20 to 29.9** | **30 to 39.9** | **40 to 49.9** | **≥50** |
|---|---|---|---|---|---|---|
| **Nobs** | 75 | 123 | 109 | 36 | 17 | 6 |
| **$R^2$** | 0.71 | 0.73 | 0.75 | 0.76 | 0.80 | 0.81 |
| **RMSE** | 2.5 | 2.8 | 3.6 | 4.9 | 6.1 | 6.5 |
| **Amplitude** | 35.9 | 39.8 | 49.2 | 59.6 | 83.5 | 89.7 |

**Table 3**. Number of data in each class (Nobs), daily amplitude of monitored $PM_{10}$, $R^2$ and RMSE for six classes of the 366 daily $PM_{10}$ values (µg.m-3) of quartile 3; local regressions.

The 366 values of 3rd quartile of daily monitored $PM_{10}$ concentration (and the $R^2$ and RMSE that characterize them) were arranged into five classes: < 15 $\mu g.m^{-3}$, ..., ≥ 50 $\mu g.m^{-3}$. It appears that the $R^2$ values are perfectly well-ordered: they increase as the $PM_{10}$ concentration increases (tab. 3). Statistical results are therefore better when quartile 3 values exceed 40 $\mu g.m^{-3}$ than when they are lower than 15 $\mu g.m^{-3}$, whatever the variables included in the regressions. RMSEs are even more orderly.

Normally, RMSE values decrease when $R^2$ values increase. This is true for estimates relating to situations where the distribution of monitored data values is similar (mean, range of series). This is not the case here where the $R^2$ and RMSE characterize six ranges of non-homogeneous values. The estimates of low monitored data values (<15 $\mu g.m^{-3}$) have a relatively low RMSE despite their relatively low $R^2$ because the range of the series concerned is also low: 36 $\mu g.m^{-3}$. By comparison, the mean range of the

series with the highest observed values (> 40, > 50 µg.m$^{-3}$) is much larger (83, 90 µg.m$^{-3}$), resulting in a potential (and actual) higher error as well.

## 5 Conclusion

The LUR geomatic method used allows us to estimate $PM_{10}$ concentrations in a large country, the whole of continental France (550 000 km²), at a fine scale (with 250 m resolution). This method also allows us to respond to another problem: France is made up of very dissimilar areas: (1) rural zones where there are few if any air quality monitoring stations, (2) mountain areas where current spatial resolutions of 7 km are inadequate for displaying variations in pollution distribution due to the influence of topography, and (3) urban areas where the local emission sources are predominant in terms of $PM_{10}$ emissions. To achieve our objectives, disparate data were collected and homogenized in the form of tables and grids in a standardized format.

This study indicates that the explanatory variables we used (topography and emission inventories) provide very poor results, explaining respectively merely 6% and 12% of $PM_{10}$ distribution. In contrast, the use of local regressions increased these values significantly to 46% and 57%, respectively. This gain stems from the dispersion of $PM_{10}$ concentrations, which is primarily a local phenomenon. But this phenomenon is also of a regional nature, as indicated by the strong link between monitored $PM_{10}$ concentrations and the intercept of local regressions. Finally, application of the autoregressive process shows that spatial patterns very frequently persist from one day to the next, which greatly improves the estimates. These estimates at d+1 show that it is possible to significantly improve the quality of forecasts as to when the information and recommendation threshold (50 µg.m$^{-3}$) and the alert threshold (80 µg.m$^{-3}$) will be exceeded and at no additional costs for the public authorities responsible for pollution and health.

This study provides several useful lessons for the future. In particular, the initial emission inventory values at medium resolution (approximately 2 km) were transferred to high resolution matrices (250 m) by dilatation smoothing. However, it would be advisable to use emission inventories on a spatially finer scale than is used here. The use of regional inventories and their spatial compilation would be a preferable solution.

The same method could be applied to deeper, more detailed analysis on a smaller scale. For example, it would be very useful to explore the effect of local topography and built environment on the downscaling method. Improving aerosol modeling in numerical atmospheric chemistry models is a long-term task. Nevertheless, in the short term, downscaling $PM_{10}$ concentration forecasts can be improved by identifying (i) finer-scale emission variables and (ii) other explanatory variables such as atmospheric variables, which are widely used to reduce the biases introduced by numerical models.

**Acknowledgements**

The authors would like to thank INERIS for making the 0.025° resolution emission files available and the LCSQA for providing validated $PM_{10}$ monitored data for the year 2016.

**References**

Analitis A, Katsouyanni K, Dimakopoulou K et al (2006) Short-term effects of ambient particles on cardiovascular and respiratory mortality. Epidemiology 17:230-3

Arslan O, Akyürek Ö (2018) Spatial Modelling of Air Pollution from $PM_{10}$ and $SO_2$ concentrations during Winter Season in Marmara Region: 2013-2014. International Journal of Environment and Geoinformatics 5:1-16. doi: 1030897/ijegeo

Ayres-Sampaio D et al 2014) An investigation of the environmental determinants of asthma hospitalizations: An applied spatial approach. Appl Geogr 47:10-19

Baldasano J M, Soret A, Guevara M, Martínez F, Gassó S (2014) Integrated assessment of air pollution using observations and modelling in Santa Cruz de Tenerife (Canary Islands). Sci Total Environ 473-474:576–588

Barba-Vasseur M, Bernard N, Pujol S, Sagot P, Riethmuller D, Thiriez G, Houot H, Defrance J, Mariet A-S, Luu V-P, Barbier A, Benzenine E, Quantin C, Mauny F (2017) Does low to moderate environmental exposure to noise and air pollution influence preterm delivery in medium-sized cities? International Journal of Epidemiology 46:2017-2027

Barré J, El Amraoui L E, Ricaud P, Lahoz W A, Attié J L, Peuch V H et al (2013) Diagnosing the transition layer at extratropical latitudes using MLS O 3 and MOPITT CO analyses. Atmospheric Chemistry and Physics 13:7225-7240

Beelen R, Hoek G, Pebesma E, Vienneau D, de Hoogh K, Briggsd DJ (2009) Mapping of background air pollution at a fine spatial scale across the European Union. Science of The Total Environment 407:1852-1867

Beloconi A, Chrysoulakis N, Lyapustin A, Utzinger J, Vounatsou P (2018) Bayesian geostatistical modelling of $PM_{10}$ and $PM_{25}$ surface level concentrations in Europe using high-resolution satellite-derived products. Environment International 121, Part 1:57-70. doi.org/10.1016/j.envint.2018.08.041

Bentayeb M, Wagner V, Stempfelet M, Zins M, Goldberg M, Pascal M, Larrieu S, Beaudeau P, Cassadou S, Eilstein D, Filleul L, Le Tertre A, Medina S, Pascal L, Prouvost H, Quénel P, Zeghnoun A, Lefranc A (2015) Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. Environment International 85:5-14

Borrego C, Monteiro A, Ferreira J, Miranda AI, Costa AM, Carvalho AC, Lopes M (2008) Procedures for estimation of modelling uncertainty in air quality assessment. Environ International 34:613-620

Chang JC, Hanna SR (2004) Air quality model performance evaluation. Meteorol Atmos Phys 87:167-196

Chemel C, Arduini G, Staquet C, Largeron Y, Legain D, Tzanos D et al (2016) Valley heat deficit as a bulk measure of wintertime particulate air pollution in the Arve River Valley. Atmos Environ 128:208-215. doi: 10.1016/j.atmosenv.2015.12.058

Czarnecka M, Nidzgorska-Lencewicz J (2017) The impact of thermal inversion on the variability of $PM_{10}$ concentration in winter seasons in Tricity. Environment Protection Engineering 43:157-172

Dehghan A, Khanjani N, Bahrampour A, Goudarzi G, Yunesian M (2018) The relation between air pollution and respiratory deaths in Tehran, Iran- using generalized additive models BMC. Pulmonary Medicine 18, 49. doi: 10.1186/s12890-018-0613-9

de Hoogh K et al (2013) Development of Land Use Regression Models for Particle Composition in Twenty Study Areas in Europe. Environ Sci Technol 47:5778-5786

de Hoogh K, Gulliver J, Donkelaar AV, et al (2016) Development of West-European $PM_{25}$ and $NO_2$ land use regression models incorporating satellite-derived and chemical transport modelling data. Environ Res 151:1-10. doi: 10.1016/j.envres.2016.07.005. Epub 2016 Jul 20

Denby B, Horálek J, Walker SE, Eben K, Fiala J (2005) Interpolation and Assimilation Methods for European Scale Air Quality Assessment and Mapping Part I: Review and Recommendations ETC/ACC Technical Paper, 7
http://acm.eionet.europa.eu/docs/ETCACC_TechnPaper_2005_7_SpatAQ_Interpol_Part_I.pdf

Denby B, Schaap M, Segers A, Builtjes P, Horalek J (2008) Comparison of two data assimilation methods for assessing $PM_{10}$ exceedances on the European scale. Atmos Environ 42:7122-7134

Denby B, Cassiani M, de Smet P, de Leeuw F, and Horálek J (2011) Sub-grid variability and its impact on European wide air quality exposure assessment. Atmos Environ 45:4220-4229

Diaz-de-Quijano M, Joly D, Gilbert D, Bernard N (2014) A more cost-effective geomatic approach to modelling $PM_{10}$ dispersion across Europe. Appl Geogr 55:108-116

Diaz-de-Quijano M, Joly D , Gilbert D , Toussaint ML , Franchi M , Fallot JM , Bernard N (2016) Modelling and mapping trace element accumulation in *Sphagnum* peatlands at the European scale using a geomatic model of pollutant emissions dispersion. Envir Pollution 214:8-16

Dufour A, Amodei M, Ancellet G, Peuch VH (2005) Observed and modelled "chemical weather" during ESCOMPTE. Atmospheric Research 7:161-189

Eeftens M et al (2012) Development of Land Use Regression Models for $PM_{25}$, $PM_{25}$ Absorbance, $PM_{10}$ and PMcoarse in 20 European Study Areas, Results of the ESCAPE Project. Environ Sci Techno 46:11195-11205

El Amraoui L E, Attié J L et al (2010) Mid latitude stratosphere–troposphere exchange as diagnosed by MLS $O_3$ and MOPITT CO assimilated fields. Atmospheric Chemistry and Physics 10:2175-2194

El-Harbawi M (2013) Air quality modelling, simulation, and computational methods: a review. Environ Rev 21:149-179

Fischer PH, Marra K et al (2015) Air Pollution and Mortality in Seven Million Adults: The Dutch Environmental Longitudinal Study (DUELS). Environ Health Perspect 123:697-704

Frey H C and Zhao Y (2004) Quantification of Variability and Uncertainty for Air Toxic Emission Inventories With Censored Emission Factor Data. Environmental Science and Technology 38:6094-6100

Guevara M, Martínez F, Arévalo G, Gassó S, Baldasano JM (2013) An improved system for modelling Spanish emissions: HERMESv20. Atmospheric Environment 81:209-221.
doi.org/10.1016/j.atmosenv.2013.08.053

Guth J, Josse B, Marécal V, Joly M, Hamer P (2016) First implementation of secondary inorganic aerosols in the MOCAGE version R2150 chemistry transport model. Geosci Model Dev 9:137-160. doi:10.5194/gmd-9-137-2016

Hamm NSA, Finley AO, Schaap M, Stein A (2015) A spatially varying coefficient model for mapping $PM_{10}$ air quality at the European scale. Atmospheric Environment 102:393-405

Hair F, Anderson RE, Tatham RL, Black WC (2006) Multivariate Data Analysis. 6th ed Pearson Education Inc: Upper Saddle River, NJ, 2006

Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42:7561–7578

Hoek G, Beelen R, Kos G, Dijkema M, van der Zee SC, Fischer PH, Brunekreef B L (2011) and Use Regression Model for Ultrafine Particles in Amsterdam. Environ Sci Technol 45:622-628

Honoré C, Rouïl L et al (2008) Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. J Geophys Res 113, D04301. doi:10.1029/2007JD008761

Joly D, Brossard T, Cardot H, Cavailhès J, Hilal M, Wavresky P (2011) Temperature Interpolation by local information, the example of France. International Journal of Climatology 31:2141-2153

Joly D, Bois B, Zaksek K (2012) Rank-ordering of topographic variables correlated with temperature. Atmospheric and Climate Science 2:139-147. doi: 10.4236/acs.2012.22015

Joly D, Cardot H, Schaumberger A (2013) Improving spatial temperature estimates by resort to time autoregressive processes. International Journal of Climatology 33:2289-2448. doi: 10.1002/joc.3601

Josse B, Simon P, Peuch VH (2004) Radon global simulations with the multiscale chemistry and transport model MOCAGE. Tellus B 5 6:339-356

Konovalov IB, Beekmann M, Meleux F, Dutot A, Foret G (2009) Combining deterministic and statistical approaches for $PM_{10}$ forecasting in Europe. Atmospheric Environment 43:6425-6434

Kuklinska K, Wolska L, Namiesnik J (2015) Air quality policy in the US and the EU – a review Atmospheric. Pollution Research 6:129-137. doi.org/10.5094/APR.2015.015

Lacressonnière G, Peuch V H, Arteta J, Josse B, Joly M, Marécal V, (2012) Watson, L How realistic are air quality hindcasts driven by forcings from climate model simulations? Geoscientific Model Development 5:1565

Lamarque J F, Shindell D et al (2013) The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics. Geoscientific Model Development 6:179-206

Largeron Y, Staquet C (2016) The Atmospheric Boundary Layer during Wintertime Persistent Inversions in the Grenoble Valleys Frontiers. Earth Science 4:1-19. doi:10.3389/feart.2016.00070

Lecœur E, Seigneur C (2013) Dynamic evaluation of a multi-year model simulation of particulate matter concentrations over Europe Atmospheric. Chem Phys 13:4319–4337

Liu W, Li X et al (2015) Land use regression models coupled with meteorology to model spatial and temporal variability of NO2 and $PM_{10}$ in Changsha, China. Atmos Environ 116:272-280

Mailler S, Menut L et al. (2017) CHIMERE-2017: from urban to hemispheric chemistry-transport modeling. Geosci Model Dev 10:2397-2423. doi.org/10.5194/gmd-10-2397-2017

Mariet, A-S, Mauny, F et al. (2018) Multiple pregnancies and air pollution in moderately polluted cities: Is there an association between air pollution and fetal growth? Environment International 12:890-897

Makra L et al (2013) The effect of different transport modes on urban $PM_{10}$ levels in two European cities Sci Total Environ 458-460:36–46. doi: 10.1016/j.scitotenv.2013.04.021

Marco G, Bo X (2013) Air Quality Legislation and Standards in the European Union: Background, Status and Public Participation Advances. Climate Change Research 4:50-59 doi.org/10.3724/SP.J.1248.2013.050

Marécal V, Peuch V-H et al (2015) A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. Geoscientific Model Development 8:2777-2813. https://doi.org/10.5194/gmd-8-2777-2015

Martet M, Peuch V H, Laurent B, Marticorena B, Bergametti G (2009) Evaluation of long-range transport and deposition of desert dust with the CTM MOCAGE Tellus B 61:449-463

Meier-Girard D, Delgado-Eckert E et al (2019) Association of long-term exposure to traffic-related PM10 with heart rate variability and heart rate dynamics in healthy subjects. Environment International 125:07-116. doi: 10.1016/j.envint.2019.01.031

Menut L, Bessagnet B et al (2013) CHIMERE 2013: a model for regional atmospheric composition modelling. Geosci Model Dev 6:981-1028. doi.org/10.5194/gmd-6-981-2013

Milford C, Castell N et al (2013) Measurements and simulation of speciated $PM_{25}$ in south-west Europe. Atmos Environ 77:36–50

Miri M, Ghassoun Y, Dovlatabadi A, Ebrahimnejad A, Löwner MO (2019) Estimate annual and seasonal PM1, PM25 and PM10 concentrations using land use regression model. Ecotoxicology and Environmental Safety 174:137-1450. https://doi.org/10.1016/j.ecoenv.2019.02.070

Mok KM, Miranda AI, Yuen K V, Hoi K I, Monteiro A, Ribeiro I (2015) Selection of bias correction models for improving the daily $PM_{10}$ forecasts of WRF-EURAD in Porto, Portugal. Atmospheric Pollution Research 8:628-639

Monteiro A, Miranda A I, Borrego C, Vautard R, Ferreira J, Perez A(2007) Long-term assessment of particulate matter using CHIMERE model. Atmos Environ 41:7726-7738

Neal L S, Agnew P, Moseley S, Ordóñez C, Savage N H, Tilbee M (2014) Application of a statistical post-processing technique to a gridded, operational, air quality forecast. Atmospheric Environment 98:385-393. doi.org/10.1016/j.atmosenv.2014.09.004

Nopmongcol U, Koo B et al (2012) Modeling Europe with CAMx for the Air Quality Model Evaluation International Initiative (AQMEII). Atmos Environ 53:177-185

Paci, A, C Staquet, J et al (2016) The Passy-2015 field experiment: atmospheric dynamics and air quality in the Arve River Valley. Pollution Atmosphérique 2016:231-232. doi.org/10.4267/pollution-atmospherique.5903

Pascal M, Corso, M, Chanel (2013) Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project. Sci Total Environ 449:390-400. doi: 10.1016/j.scitotenv.2013.01.077. Epub 2013 Feb 26

Plutowski M, Sakata S, Hasenclever J (1994) Cross-validation estimates IMSE. JD Cowan, G Tesauro, J Alspector (Eds), Advances in Neural Information Processing Systems), San Mateo, CA 1994:391-398

Potier E, Waked A, Bourin A, Minvielle F, Péré JC, Perdrix E, Michoud V, Riffault V, Alleman LY, Sauvage S (2019) Characterizing the regional contribution to PM10 pollution over northern France using two complementary approaches: Chemistry transport and trajectory-based receptor models. Atmospheric Research 223:1-14

Riant M, Meirhaeghe A, Giovannelli J, Occelli F, Havet A, Cuny D, Amouyel P, Dauchet L (2018) Associations between long-term exposure to air pollution, glycosylated hemoglobin, fasting blood glucose and diabetes mellitus in northern France. Environment International 120:121-129

Riviere E, Bernard J et al (2019) Air pollution modeling and exposure assessment during pregnancy in the French Longitudinal Study of Children (ELFE). Atmospheric Environment, 205:103-114 https://doi.org/10.1016/j.atmosenv.2019.02.032

Ryan PH, LeMasters GK A (2007) Review of Land-use Regression Models for Characterizing Intraurban Air Pollution Exposure. Inhal Toxicol 19:127–133. doi: 10.1080/08958370701495998

Rouïl L, Honoré C et al (2009) PREV'AIR: an operational forecasting and mapping system for air quality in Europe. Bulletin of the American Meteorological Society 90:73-83

Sič B, El Amraoui L, Marécal V, Josse B, Arteta J, Guth J, Joly,M, Hamer PD (2015) Modelling of primary aerosols in the chemical transport model MOCAGE: development and evaluation of aerosol physical parameterizations. Geoscientific Model Development 8:381-408

Shahraiyni H T, Sodoudi S (2016) Statistical Modeling Approaches for $PM_{10}$ Prediction in Urban Areas, A Review of 21st-Century Studies Atmosphere 7 15 p doi:10.3390/atmos702001

Shtein A, Karnieli A, Katra I, Raz R, Levyd I, Lyapusti, A, Dorman M, Broday DM, Kloog I (2018) Estimating daily and intra-daily $PM_{10}$ and $PM_{25}$ in Israel using a spatio-temporal hybrid modeling approach. Atmospheric Environment 191:142-152

Simpson D, Benedictow A et al (2012) The EMEP MSC-W chemical transport model – technical description. Atmospheric Chem Phys 12:7825-7865

Stadlober E, Hörmann S, Pfeiler B (2008) Quality and performance of a $PM_{10}$ daily forecasting model .Atmospheric Environment 42:1098-1109

Stafoggia, M, Schwartz, J et al (2017) Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. Environment International 99:234-244

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Stat Soc Ser B Methodol 36:111-147

Teyssèdre H, Michou M, Clark H L, Josse B, Karche F, Olivié D, Nédélec PA (2007) New tropospheric and stratospheric Chemistry and Transport Model MOCAGE-Climat for multi-year studies: evaluation of the present-day climatology and sensitivity to surface processes. Atmospheric Chemistry and Physics 7:5815-5860. https://doi.org/10.5194/acp-7-5815-2007

Theobald MR, Simpson D, Vieno M (2016) Improving the spatial resolution of air-quality modelling at a European scale – development and evaluation of the Air Quality Re-gridder Model (AQR v11). Geosci Model Dev 9:4475-4489, 2016. doi:10.5194/gmd-9-4475-2016

Valavanidis A, Fiotakis K, Vlachogiann, T (2008) Airborne Particulate Matter and Human Health: Toxicological Assessment and Importance of Size and Composition of Particles for Oxidative Damage and Carcinogenic Mechanisms J Environ Sci Health Part C-Environ. Carcinog Ecotoxicol Rev 26:339-362

Vienneau D, de Hoogh K, Beelen,R, Fischer P, Hoek G, Briggs D (2010) Comparison of land-use regression models between Great Britain and the Netherlands. Atmos Environ 44:688-696

Vienneau D, de Hoogh K, Bechle M J, Beelen R, van Donkelaar A, Martin RV, Millet DB, Hoek G, Marshall JD (2013) Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO2 and $PM_{10}$. Environ Sci Technol 47:13555-13564

Waked A, Bourin A et al (2018) Investigation of the geographical origins of PM10 based on long, medium and short-range air mass back-trajectories impacting Northern France during the period 2009–2013. Atmospheric Environment 193:143-152. https://doi.org/10.1016/j.atmosenv.2018.08.015

Zhang Z, Gong D, Mao R, Kim SJ, Xu J, Zhao X, Ma Z (2015) Cause and predictability for the severe haze pollution in downtown Beijing in November–December 2015. Science of the Total Environment 592:627-638. doi: 10.1016/j.scitotenv.2017.03.009

Figure 1. Location of the 325 PM10 monitoring stations in continental France

Click here to access/download;Figure;Fig1.tif

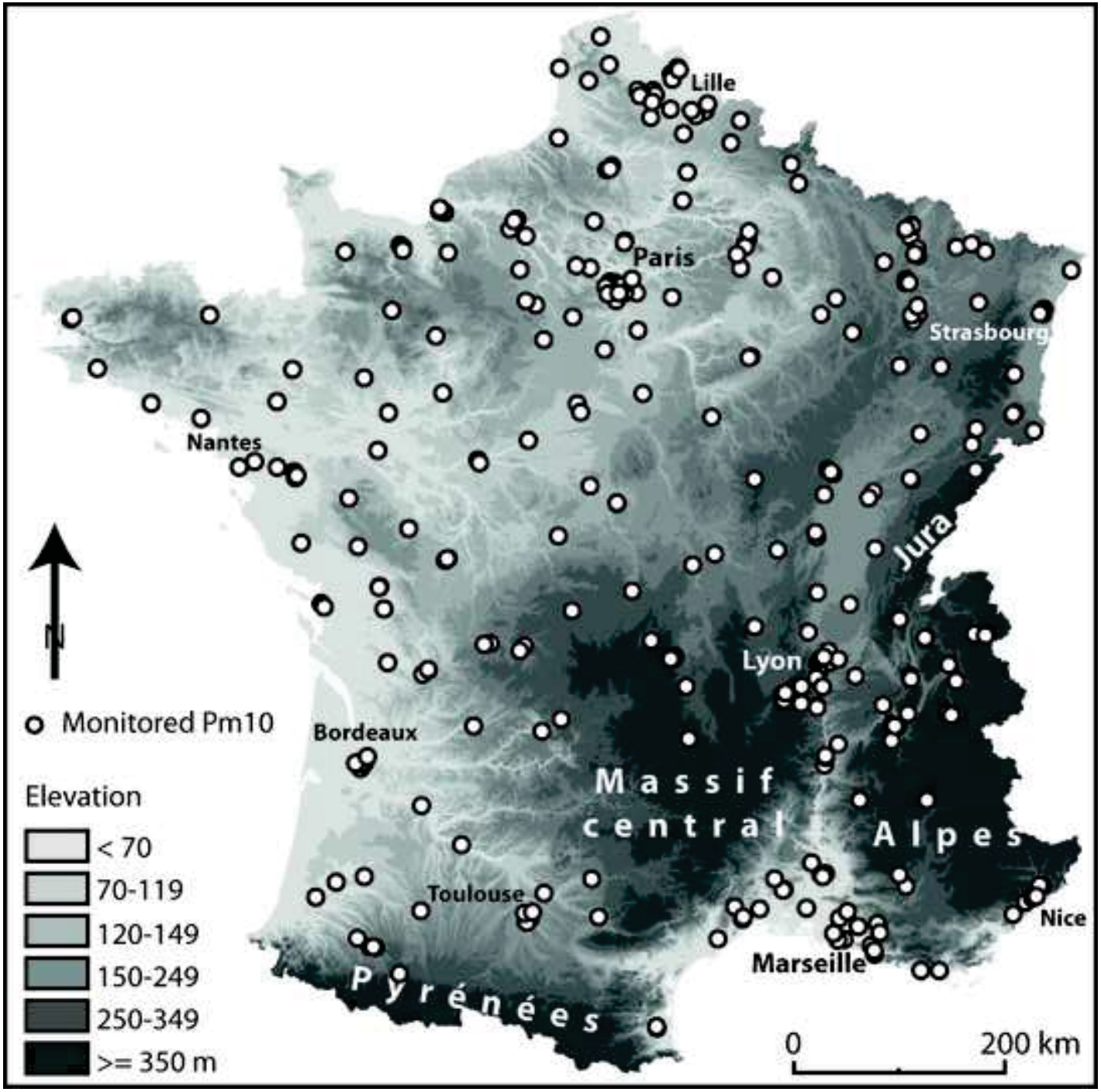


Figure 2. Available data and their transformation for further statistical processing

Click here to access/download;Figure;Fig2.tif

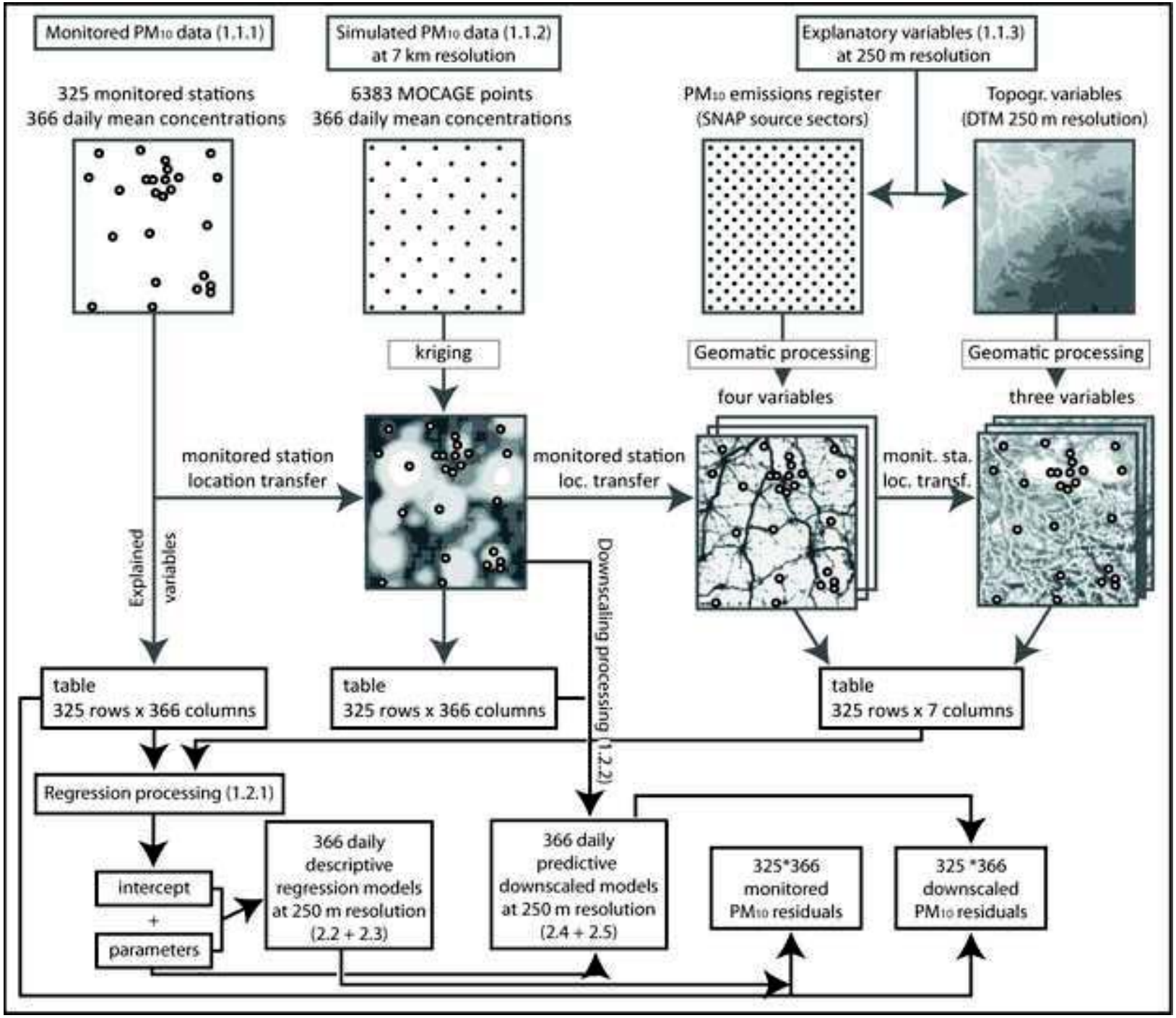


Figure 3. Scatter plot between the 325 monitored PM10 versus the values simulated by Mocage and the values predicted by downscaling; case of December 1, 2016

Click here to access/download;Figure;Fig3.tif

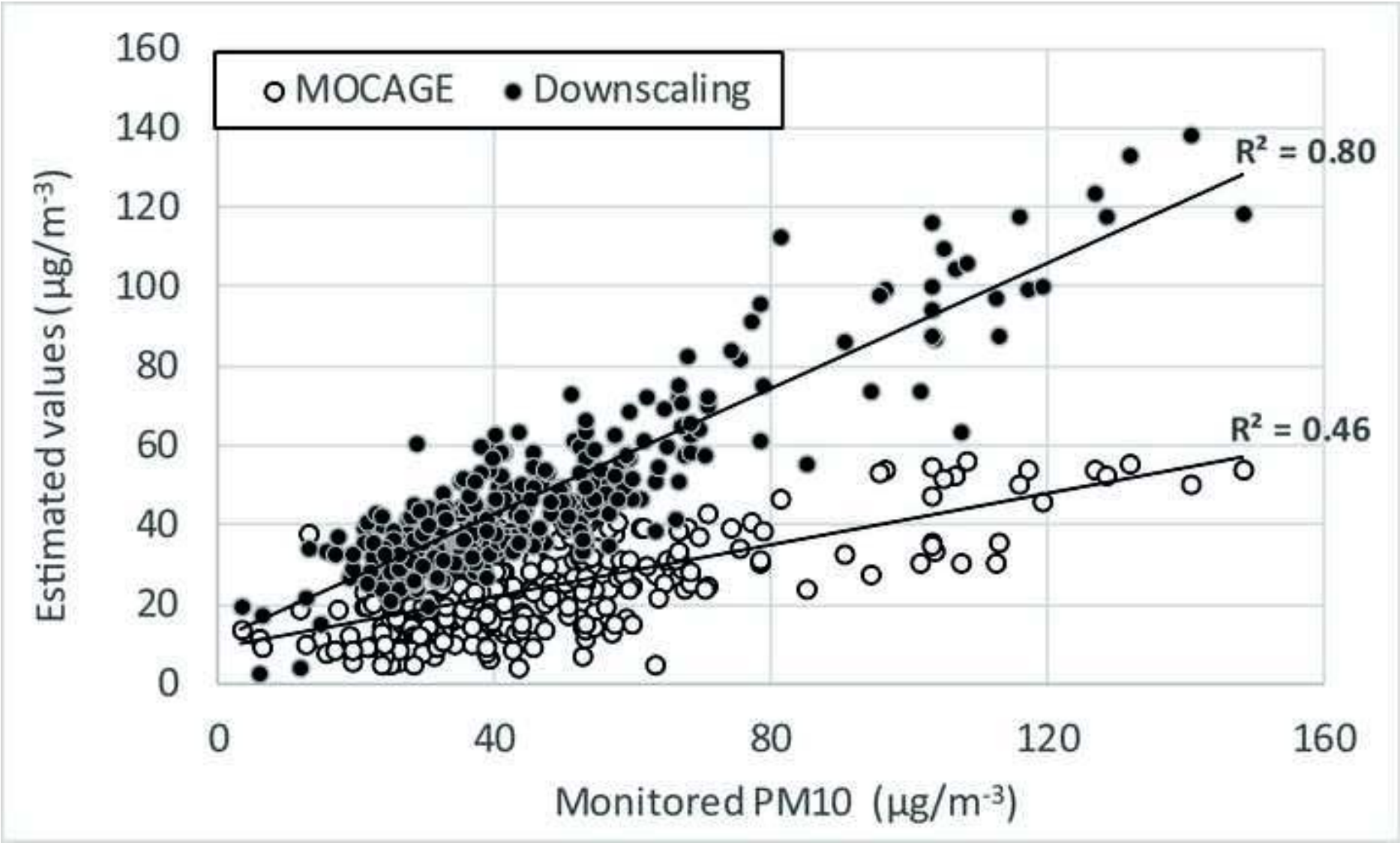
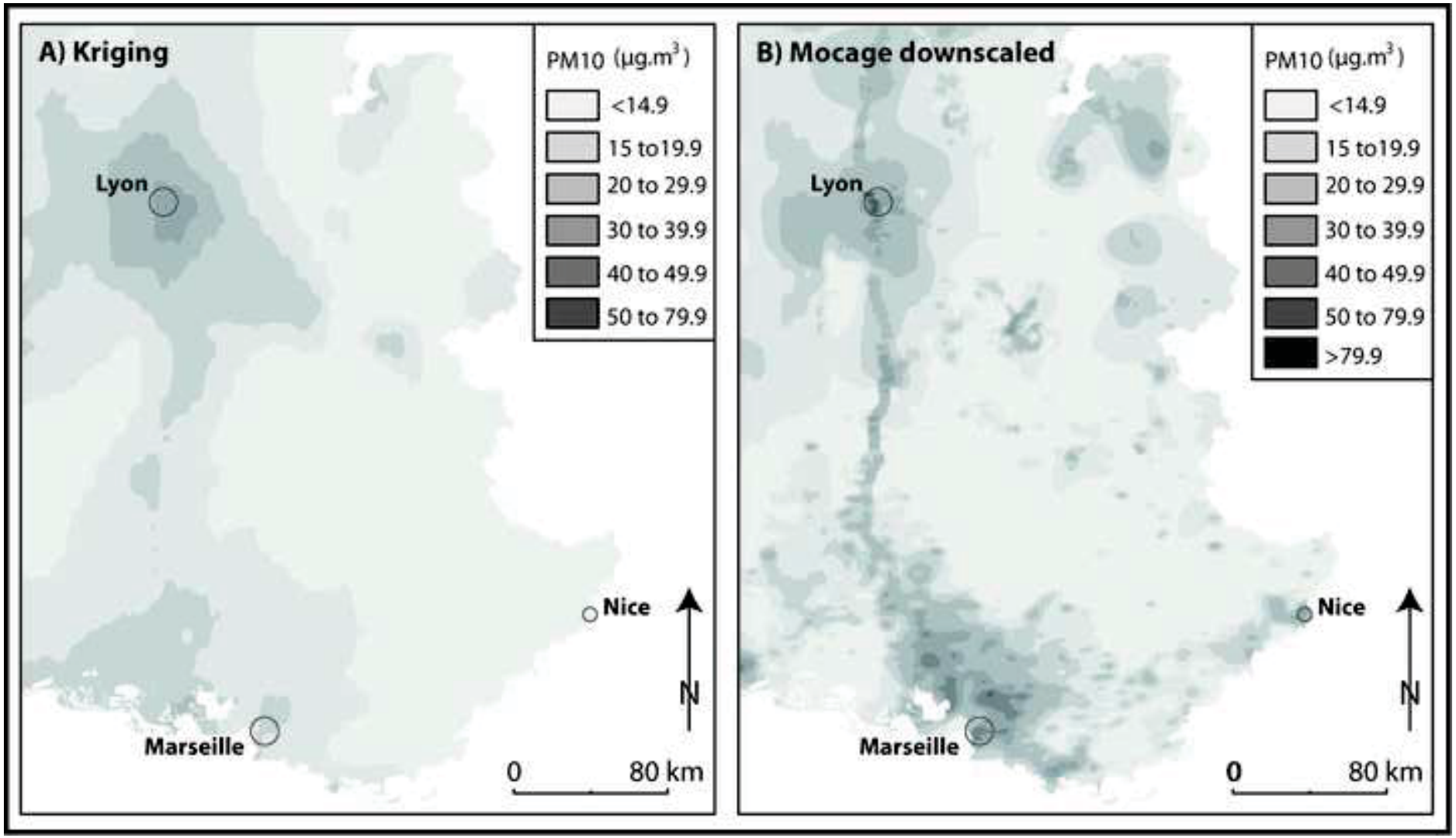


Figure 4. Restitution of PM10 simulated by MOCAGE for December 1, 2016 by (A) kriging and (B) downscaling.he values predicted by downscaling; case of December 1,

Click here to access/download;Figure;Fig4.tif