A multi-resolution ensemble model of three decision-tree-based algorithms to predict daily NO2 concentration in France 2005–2022

Guillaume Barbalat, Ian Hough, Michael Dorman, Johanna Lepeule, Itai Kloog

Environmental Research 257 (2024) 119241

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

A multi-resolution ensemble model of three decision-tree-based algorithms to predict daily NO_2 concentration in France 2005–2022

Guillaume Barbalat [a,b,*], Ian Hough [c], Michael Dorman [d], Johanna Lepeule [a,1,**], Itai Kloog [d,e,1]

[a] University Grenoble Alpes, Inserm, CNRS, Team of Environmental Epidemiology Applied to Development and Respiratory Health, Institute for Advanced Biosciences (IAB), Grenoble, France
[b] Centre Ressource de Réhabilitation Psychosociale et de Remédiation Cognitive, Hôpital Le Vinatier, Pôle Centre Rive Gauche, UMR, 5229, CNRS & Université Claude Bernard Lyon 1, France
[c] Université Grenoble Alpes, CNRS, INRAE, IRD, INP-G, IGE (UMR 5001), Grenoble, France
[d] The Department of Environmental, Geoinformatics and Urban Planning Sciences, Ben-Gurion University of the Negev, Israel
[e] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

ARTICLE INFO

Keywords:
Nitrogen dioxide
200 m resolution
Daily predictions
Spatio-temporal modeling
Decision-tree
Spatio-temporal blocking

ABSTRACT

Understanding and managing the health effects of Nitrogen Dioxide (NO_2) requires high resolution spatiotemporal exposure maps. Here, we developed a multi-stage multi-resolution ensemble model that predicts daily NO_2 concentration across continental France from 2005 to 2022. Innovations of this work include the computation of daily predictions at a 200 m resolution in large urban areas and the use of a spatio-temporal blocking procedure to avoid data leakage and ensure fair performance estimation. Predictions were obtained after three cascading stages of modeling: (1) predicting NO_2 total column density from Ozone Monitoring Instrument satellite; (2) predicting daily NO_2 concentrations at a 1 km spatial resolution using a large set of potential predictors such as predictions obtained from stage 1, land-cover and road traffic data; and (3) predicting residuals from stage 2 models at a 200 m resolution in large urban areas. The latter two stages used a generalized additive model to ensemble predictions of three decision-tree algorithms (random forest, extreme gradient boosting and categorical boosting). Cross-validated performances of our ensemble models were overall very good, with a ten-fold cross-validated R^2 for the 1 km model of 0.83, and of 0.69 for the 200 m model. All three basis learners participated in the ensemble predictions to various degrees depending on time and space. In sum, our multi-stage approach was able to predict daily NO_2 concentrations with a relatively low error. Ensembling the predictions maximizes the chance of obtaining accurate values if one basis learner fails in a specific area or at a particular time, by relying on the other learners. To the best of our knowledge, this is the first study aiming to predict NO_2 concentrations in France with such a high spatiotemporal resolution, large spatial extent, and long temporal coverage. Exposure estimates are available to investigate NO_2 health effects in epidemiological studies.

1. Introduction

According to The Lancet Commission on pollution and health, air pollution caused 6.7 million deaths in 2019, with increasing figures for deaths attributable to ambient air pollution (Fuller et al., 2022). Nitrogen dioxide (NO_2), one of the most noxious gaseous pollutants, is thought to affect health both directly and indirectly via its role in fine particulate matter 2.5 (PM2.5) and ozone formation. Exposure to NO_2 has been demonstrated to raise the risk of a wide range of major disorders, from asthma to low birth weight or reduced cognitive function (Dominski et al., 2021; Mainka and Żak, 2022). Based on such a negative evaluation of the impact of NO_2 on health, the WHO recommends a maximum annual concentration of 10 $\mu g/m^3$ (World Health Organization, 2021).

Having a lifetime of just a few hours (Pommier, 2023), NO_2 shows relatively large spatial and temporal gradients. This is especially

* Corresponding author. University Grenoble Alpes, Inserm, CNRS, Team of Environmental Epidemiology Applied to Development and Respiratory Health, Institute for Advanced Biosciences (IAB), Grenoble, France.
** Corresponding author.
E-mail addresses: guillaume.barbalat@univ-grenoble-alpes.fr (G. Barbalat), johanna.lepeule@univ-grenoble-alpes.fr (J. Lepeule).
[1] Co-senior authors.

https://doi.org/10.1016/j.envres.2024.119241
Received 19 January 2024; Received in revised form 13 May 2024; Accepted 26 May 2024
Available online 27 May 2024

important for urban areas (Wang et al., 2020), where a large proportion of individuals live around some of the major sources of NO_2. In epidemiological studies, imprecise NO_2 concentrations may lead to a number of issues. For both time-series and long-term exposure studies, large uncertainties in effect estimates have been reported when low spatial and temporal resolution concentrations were used (Richmond-Bryant and Long, 2020; Sellier et al., 2014). Spatial error in particular may negatively bias effect estimates, and positively bias effect estimates when the variance of error correlates with the exposure (Richmond-Bryant and Long, 2020). Therefore, to reduce measurement error in epidemiological studies and better understand health effects related to NO_2, it is paramount to obtain NO_2 concentration maps at a high spatio and temporal resolution. However, monitoring stations are scarce, even in developed countries, and monitors tend to be clustered in cities, which makes it difficult to estimate NO_2 exposure for suburban and rural populations. Even in cities, NO_2 monitors are often too sparse to properly capture the spatial variation in NO_2 concentration.

One popular approach to estimate NO_2 concentrations outside monitoring stations is to use geostatistical learning. Recently, some have used flexible machine learning algorithms such as tree-based methods (e.g. random-forest and extended gradient boosting procedures) to learn the spatio-temporal relationships between pollutants at monitoring stations and related variables, and in turn predict concentrations of ambient air pollutants outside monitor areas, with overall good predictive accuracy. A hybrid model ensembling predictions from individual learners could potentially improve model performance. Previous studies have obtained good performance accuracy after employing these procedures to calculate relatively highly resolved predictions of NO_2 concentration in countries (de Hoogh et al., 2019: $R^2 = 0.58–0.73$; Di et al., 2020: $R^2 = 0.79$), large regions (Pan et al., 2021: $R^2 = 0.72$) or cities (He et al., 2023: $R^2 = 0.66–0.87$; Zhang et al., 2021: $R^2 = 0.77$), with relatively high spatial and temporal resolution, and with good performance accuracy.

Despite these notable advances, only a few exposure models have provided high-quality NO2 concentrations over large spatial areas. First, less than a handful of studies have calculated predictions of NO2 concentrations at a spatial resolution of 200 m or lower (de Hoogh et al., 2019; He et al., 2023; Zhang et al., 2021). But estimating intra-urban variation in NO_2 with precision would be paramount given the large proportion of the population living in cities and the strong spatial and temporal gradients of NO_2. Second, some exposure models only produced annual predictions (de Hoogh et al., 2018). Such models are not well suited for research investigating critical windows of exposure to air pollution, as is the case in mother child cohort studies (Guilbert et al., 2023; Hough et al., 2023). Third, very few studies have reported the use of a specific procedure for model performance evaluation that takes into account the spatio-temporal dependencies of monitor data (Hough et al., 2021; Just et al., 2020; Shtein et al., 2020). Failure to do so is a potential source of overfitting, the artificial inflation of a model's performance (Meyer and Pebesma, 2022; Roberts et al., 2017).

In an effort to improve exposure assessment methods for their application in environmental epidemiology, we predicted daily NO_2 concentration at a 1 km resolution in France from 2005 to 2022, and improved predictions to a 200 m resolution in large urban areas. We used an ensemble of decision-tree-based algorithms (random forest, extreme gradient boosting and categorical boosting), and a large set of potential predictors of NO_2 concentrations. We computed the predictive accuracy of our models using a blocking scheme that takes into account spatio-temporal dependencies.

2. Material and methods

2.1. Data

2.1.1. Study domain and model grids

Continental France covers a roughly hexagonal area of 542,973 km^2 in Western Europe. The population is about 64.5 million, where about 20% of the population is rural, and 37% live in towns and small cities with fewer than 500,000 residents (Insee, 2023).

For this study, we ran one model per year, based on raster grids covering continental France. All years could not be modeled together due to incurring computational costs. Our 1 km grid was derived from the 2022 MOD09GA product (Vermote and Wolfe, 2015), covering 591869 km^2, to which we added a third temporal dimension of 365 days (or 366 for leap years). To obtain a 200 m resolution grid covering large urban areas, we first selected the 118 French towns with a population larger than 50,000 inhabitants (Hough et al., 2020) as per the 2020 census (Insee, 2023) (obtained from https://www.observatoire-des-territoires.gouv.fr/outils/cartographie-interactive/). The location of each area was obtained from https://www.data.gouv.fr/fr/datasets/decoupage-administratif-communal-francais-issu-d-openstreetmap/, a government database using data reconstructed from OpenStreetMap. This dataset was subsequently merged with our 1 km grid, and downscaled to a finer 200 m grid.

2.1.2. Outcome variable: NO_2 concentration at monitor stations

NO_2 concentrations were obtained as vector points at monitor stations from The French National Institute for Industrial Environment and Risks (INERIS, n.d.). Monitors report hourly NO_2 concentrations everyday. NO_2 concentrations were determined by the chimiluminescence method or the cavity attenuated phase shift (CAPS) technique. The detection limit was 3.824 $\mu g/m^3$. Values higher than 499.032 $\mu g/m^3$ were subject to further investigations to confirm their validity. Finally, all values that were equal to or higher than the negative detection limit were accepted as such, and other values were invalidated (LCSQA, n.d.). Over the course of the study period, the average number of NO_2 monitors was of 423 per year (range: 357 to 487) and tended to decrease over time. After indexing each monitor to the containing 1 km and 200 m grid cells, we found that the number of grid cells that include NO_2 observations was of 414 per year (range: 350–478) for the 1 km grid, and of 197 per year (range: 177–227) for the 200 m grid. We indexed each monitor to the containing 1 km and 200 m grid cell and calculated daily mean NO_2 for days with at least 18 hourly observations. To limit the impact of instrument malfunctions and rare events, and given measurement distribution, we considered daily NO_2 concentrations >210 $\mu g/m^3$ as outlier values.

2.1.3. Predictors

All variables below were added as raster layers to our originalHough et al., 2023 1 km raster grids. Where appropriate, we also state which variables were added as raster layers to our 200 m grids. Each raster layer was resampled to fit our grids (using nearest neighbor resampling).

2.1.3.1. Proxys of NO_2 concentrations. To improve model prediction accuracy, we used NO_2 concentrations proxies, such as NO_2 satellite data obtained from Ozone Monitoring Instrument products, previously modeled NO_2 data from the Copernicus Atmospheric Data Service, and estimations of NOx emissions.

We used total column densities of NO_2 from earthshine radiances measured by the Ozone Monitoring Instrument (OMI), aboard the EOS-Aura satellite (level 3) (Bucsela et al., 2006). The temporal resolution of OMI data product is of one day, and its spatial resolution is of 0.25° latitude x 0.25° longitude (equivalent to 27.75 km near the equator). Measurement values are in number of molecules per cm^2. Negative measurement values were removed from the original dataset.

We acquired NO_2 column simulations from Copernicus Atmosphere Monitoring Service (CAMS), a reanalysis data set (Inness et al., 2015). The CAMS data for NO_2 rely on observations from multiple satellites, combined with state-of-the-art computer models. CAMS NO_2 columns were obtained daily with a spatial resolution of 0.75° latitude × 0.75° latitude (equivalent to 83.25 km near the equator), and at a 3 h interval.

Averaged daily values were calculated for each grid cell.

We extracted total NOx emissions for each French town, averaged for each trimester of the years 2004, 2007 and 2012 (Inventaire National Spatialisé, n.d.). NOx emission datasets cover various domains such as agriculture, household, industry, traffic and wood smoke emissions. We used Ordinary Kriging to fill in space between towns' centroids. We assumed values to be constant from one measurement date to the next, and from 2012 to 2022.

2.1.3.2. Meteorological variables. Meteorological parameters such as wind, rain, and temperature affect surface NO_2 concentrations. We obtained hourly meteorological parameters at approximately 30 km spatial resolution from the Copernicus Climate Change Service ERA5 reanalysis (Hersbach et al., 2020). For each day, we used 13 meteorological predictors: boundary layer height at 0:00 (blh 00) and 12:00 (blh 12) UTC, total precipitation (tp), mean and standard deviation of 2 m air temperature (t2m mean and t2m sd), mean 2 m dewpoint temperature (d2m), mean surface pressure (sp), mean u- and v-components of 10 m wind speed (u10 and v10), mean total cloud cover (tcc), mean total evaporation (e), mean surface solar radiation (ssr) and mean snow albedo (asn).

2.1.3.3. *Elevation.* Topography can determine the dispersion and deposition of NO_2. We obtained elevation data from the Global Multi-Resolution Terrain Elevation Dataset 2010 (Earth Resources Observation And Science (EROS) Earth Resources Observation And Science Center, 2017), with an original 7.5-arc-second spatial resolution (approximately 225 m). Elevation was added as a raster layer to both our 1 km and 200 m grids.

2.1.3.4. NDVI. Normalized difference vegetation index (NDVI) is a widely used indicator for vegetation coverage. Vegetation can represent the absence of traffic or the presence of agriculture, and can influence the deposition and dispersion of air pollutants, including NO_2. We obtained NDVI data from the MODIS data product MOD13A3 every 30 days at a 1 km resolution from the USGS EarthExplorer website (Didan et al., 2015). We also extracted NDVI from the MODIS data product MOD13Q1 every 16 days at a 250 m resolution. NDVIs from MOD13A3 were added as raster layers to our 1 km grid, and NDVIs from MOD13Q1 were added to our 200 m grids.

2.1.3.5. *Corine land cover.* The 100 m European Corine Land Cover file (Bossard et al., 2000)was released in 2020 as a first update of final version 20, and downloaded for the reference years 2000, 2006, 2012, 2018. Each file covers 6 years of land use. From the 44 land classes available, six main groups were extracted: residential (Corine class = 1 & 2; RES), industry or commercial (3; IND), urban green (10; URBGR), total built up (1–9; BUILT), agriculture (12–22; AGR) and semi-natural and forest (23–41; NAT). The percentage of each land use variable was calculated within each cell of our 1 km and 200 m grids.

2.1.3.6. Population density. NO_2 concentration directly results from anthropogenic behaviours, which was our rationale to include population density as a predictor of NO_2 concentration. A map of population density was obtained from the gridded population of the world website from an approximately 1 km grid for the years 2005, 2010, 2015, 2020 (Doxsey-Whitfield et al., 2015).

2.1.3.7. Distance to the nearest highly populated cell. As per the World Bank, a threshold of 1500 inhabitants per 1 km^2 identifies cities with a high degree of urbanization (Dijkstra et al., 2020). We reasoned that a short distance to dense areas may be linked to higher NO_2 concentrations (e.g. because of highly congested traffic).

2.1.3.8. Road density. Road density is a proxy for total traffic volume, and thus is related to traffic emissions. We obtained road networks for four types of roads from the Institut National de l'Information Géographique et Forestière (IGN): Autoroutes, Nationales, Départementales, Other; i.e. Motorways, Primary roads, Secondary roads and Tertiary roads, respectively (IGN, 2021). The original data sets were line vectors obtained on a yearly basis. From these, we calculated the total length of roads in each grid cell of our 1 km and 200 m grids.

2.1.3.9. Intersections. Road intersections are proxies for cars idling, and again should be related to traffic emissions. These were obtained from the IGN database and defined as a node where 3 or more roads connect (IGN, 2021). From these, we calculated the total number of intersections in each cell of our 1 km and 200 m grids.

2.1.3.10. Distance to the nearest road and intersection. We adopted the same reasoning as for population density (see above): a short distance to a road or intersection may be linked to greater NO_2 concentrations because of a more congested traffic. For each grid cell, we therefore calculated the shortest distance from the cell centroid to the nearest road (for each road type), and to the nearest road node.

2.1.3.11. Railway tracks. The original datasets were line vectors obtained on a yearly basis from the IGN data service (IGN, 2021). From these, we calculated the length of electrified and non-electrified railway tracks in each cell of our 1 km and 200 m grids.

2.1.3.12. Train stations. Location of both passenger and freight train stations were obtained from the IGN data service on a yearly basis (IGN, 2021). From these, we calculated the number of train stations in each cell of our 1 km and 200 m grids.

2.1.3.13. Night-time light. Night-time light serves as a proxy for overall economic activity (Mellander et al., 2015), but also provides a measure of night-time traffic, and thus is indirectly related to NO_2 emissions. We acquired night-time light from the Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) before 2013 and from the Visible Infrared Imaging Radiometer Suite (VIIRS) from 2014 onward. These are generated by the Earth Observation Group, and were obtained from the U.S. National Oceanic and Atmosphere Administration (NOAA) website (Earth Observation Group, n.d.). These data products are cleaned prior to averaging by removing stray light, lightning, lunar illumination and cloud-cover. VIIRS data are also cleaned to screen out lights from aurora, fires, boats, and other temporal lights, and background (non-light) values, as well as outliers. We used annual average nighttime light, with a spatial resolution of 1 km (DMSP), and 500 m (VIIRS).

2.1.3.14. Time and space. Latitudes and longitudes were extracted at the centroids of each cell of our 1 km and 200 m grids, and added as raster layers to both grids. Ordinal date (to capture long-term trends), day of the week (to capture commuting and business activity) and month of the year (to capture seasonal trends) were also included as raster layers for both the 1 km and 200 m grids.

Supplementary Table 1 summarizes data availability, spatial and temporal resolution, and missingness-filling strategy for all of our predictors.

2.2. Modeling stages

We implemented our modeling strategy in three stages: (1) filling gaps in OMI data; (2) predicting daily NO_2 concentration at a 1 km spatial resolution based on a large set of predictors, including predicted OMI data from stage 1; and (3) increasing spatial resolution to 200 m in large urban areas. We chose to use three decision-tree-based algorithms,

namely random forest (RF) (Kamińska, 2018), extreme gradient boosting (XGB) (Just et al., 2020), and categorical gradient boosting (catB) (Bouguerra et al., 2023), based on their better performance on pre-tests compared to other models such linear mixed models, Gaussian Markov Random Fields (GMRF), support vector machines and neural networks. A description of the decision-tree-based algorithms and their main hyperparameters is provided in the Supplementary Materials (**Supplementary Methods**).

2.2.1. Stage 1: filling gaps in OMI data

Missing OMI data occurred due to cloud cover or instrument malfunction. For each year of the study period, we estimated the daily density of NO_2 measured by OMI satellite at a 1 km spatial resolution, using a RF model with the equation:

$$OMI_{st}^{Y} = f\left(CAMS_{st}^{Y}, x_s, y_s, wday_t^{Y}, yday_t^{Y}, month_t^{Y}\right) + \varepsilon_{st}^{Y} \quad (1)$$

where for each year Y (2005, …, 2022), OMI_{st}^{Y} is the NO_2 density at OMI observed at 1 km grid cell s (1, …, 591869) on day t (1, …, number of days in year Y); $CAMS_{st}^{Y}$ is the NO_2 density reconstructed with the CAMS methodology; x_s and y_s are spatial coordinates of cell s; $wday_t^{Y}$, $yday_t^{Y}$, $month_t^{Y}$, are, respectively, the day of the week, the day of the year, and month of year on day t; and ε_{st}^{Y} is the error at cell s on day t.

The main parameters of the random forest model were num.trees = 500, mtry = 2, min.node.size = 5, splitrule = variance, which are the default parameters of the *ranger* R function. A random subset of 80% of data was used for training and the remaining 20% of data were used for testing the model. Finally, we predicted OMI data over the entire dataset using the above-mentioned set of predictors and fitted model.

2.2.2. Stage 2: predicting daily NO_2 concentrations at a 1 km spatial resolution

We then estimated daily NO_2 concentrations at a 1 km spatial resolution across France. We trained a catB, a XGB, and a RF, with the equation:

$$NO2_{st}^{Y} = f\left(OMI_{st}^{Y}, X_{1st}^{Y}, \ldots, X_{38st}^{Y}, x_s, y_s, wday_t^{Y}, yday_t^{Y}, month_t^{Y}\right) + \varepsilon_{st}^{Y} \quad (2)$$

where for each year Y (2005, …, 2022), $NO2_{st}^{Y}$ is the NO_2 concentration at 1 km grid cell s (1, …, 591869) on day t (1, …, number of days in year Y); OMI_{st}^{Y}, x_s, y_s, $wday_t^{Y}$, $yday_t^{Y}$, $month_t^{Y}$, are as in equation (1); X_{1st}^{Y}, …, X_{38st}^{Y} are the value for the NOx predictor, and each of the predictors in chapter 1.1.3.2. to 1.1.3.13. (inc. meteorological, land-cover, anthropogenic predictors etc …) at cell s on day t; and ε_{st}^{Y} is the error at cell s on day t.

We trained each basis learner (catB, XGB, RF) yearly, yielding 54 base models (3 models × 18 years; mean observations per year = 141,358; range 123,881 to 161,172). Predictors were scaled to have a similar range. We used the base models to predict NO_2 for the 591,869 km × 365 days (or 366 for leap years) 1 km grid cell-days of the study domain (i.e. 216,032,185 cell-days, or 216,624,054 for leap years).

To minimize computational time to search for best hyperparameters of the three basis learners, we used a resampling "adaptive" scheme that was found effective at finding reasonable values of tuning parameters in a more computationally effective way (Kuhn, 2014).

We then used an ensemble of the above-mentioned algorithms to predict NO_2 concentrations. Such an ensemble modeling procedure usually enables a slight gain in performance accuracy, but mostly ensures model robustness and independence by adding a cross-validation step, while maximizing the chances that one of the models would capture variations that the others had not. We used annual generalized additive models (GAM) that weight the basis learners according to spatiotemporal variations in their performance. Specifically, we used smooths that allowed the weight for each base learner's predictions to vary smoothly over space and time, using a tensor product of penalized cubic regression splines of the spatial coordinates and the temporal index of each grid cell:

$$NO2_{st}^{Y} = te(x_s, y_s, t)catB_{st}^{Y} + te(x_s, y_s, t)XGB_{st}^{Y} + te(x_s, y_s, t)RF_{st}^{Y} + \varepsilon_{st}^{Y} \quad (3)$$

where for each year Y (2005, …, 2022), $NO2_{st}^{Y}$ is the NO_2 concentration at 1 km grid cell s (1, …, 591869) on day t (1, …, number of days in year Y); $te(x_s, y_s, t)$ is the tensor product of penalized cubic regression splines of the spatial coordinates of cell s (x_s and y_s) and the temporal index t; $catB_{st}^{Y}, XGB_{st}^{Y}, RF_{st}^{Y}$ are, respectively, the predicted NO_2 concentration at cell s on day t from a catB, XGB, and RF; and ε_{st}^{Y} is the error at cell s on day t.

2.2.3. Stage 3: increasing spatial resolution to 200 m across large urban areas

Third, we increased the spatial resolution of our predictions to 200 m over large urban areas by leveraging predictions obtained at the 1 km stage. We started by associating each 200 m grid cell with 1 km NO_2 predictions obtained from stage 2 by interpolating the 1 km predictions to the 200 m grid centroids. Next, we calculated the residuals for all 200 m grid cell-days with a monitoring station. The averaged number of observations per year was of 66,651 (range 61,364 to 74,988). We trained a catB, a XGB, and a RF to predict residuals at monitor stations with the equation:

$$R_{ij}^{Y} = f\left(NO2p_{ij}^{Y}, Elevation_i, NDVI_{ip}^{Y}, CLC_{ily}^{Y}, Traffic_{ig}^{Y}, x_i, y_i, wday_j^{Y}, yday_j^{Y}, month_j^{Y}\right) + \varepsilon_{ij}^{Y} \quad (4)$$

where for each year Y (2005, …, 2022), R_{ij}^{Y} is the residual of the 1 km ensemble model associated with 200 m grid cell i on day j; $NO2p_{ij}^{Y}$ is the 1 km NO_2 ensemble prediction associated with 200 m grid cell i on day j; $Elevation_i$ is the elevation of cell i; $NDVI_{ip}^{Y}$ is the NDVI of cell i for the time period p in which day j falls; CLC_{ily}^{Y} is the fraction of cell i occupied by each of the 6 corine land cover groups l in the CLC inventory year y, preceding and closest to day j; $Traffic_{ig}^{Y}$ is the density of each traffic-related group g (motorways, primary, secondary, tertiary roads, road nodes, train stations, electrified and non-electrified railway tracks) in cell i, x_i and y_i are the spatial coordinates of cell i; $wday_j^{Y}$, $yday_j^{Y}$, $month_j^{Y}$, are, respectively, the day of the week, the day of the year, and month of year on day j; and ε_{ij}^{Y} is the error for cell i on day j.

We used the base models to predict the residuals for the 230,904 cells × 365 days (or 366 for leap years) 200 m grid cell-days of the study domain (i.e. 84,279,960 cell-days, or 84,510,864 for leap years).

Predictors were scaled to have similar range. We trained each basis learner yearly, yielding again 54 base models. To minimize computational time due to the search of best hyperparameters, again we used a resampling "adaptive" scheme (Kuhn, 2014).

We then ensembled the predictions using annual generalized additive models (GAM) weighing the base learners according to spatiotemporal variations in their performance. Specifically, we used smooths that allowed the coefficient for each base learner's predictions to vary smoothly over space and time, using a tensor product of penalized cubic regression splines:

$$R_{ij}^{Y} = te(x_i, y_i, j)catB_{ij}^{Y} + te(x_i, y_i, j)XGB_{ij}^{Y} + te(x_i, y_i, j)RF_{ij}^{Y} + \varepsilon_{ij}^{Y} \quad (5)$$

where for each year Y (2005, …, 2022), R_{ij}^{Y} is the residual of the 1 km ensemble model associated with 200 m grid cell i (1, …, 230904) on day j (1, …, number of days in year Y); $te(x_i, y_i, j)$ is the tensor product of penalized cubic regression splines of the spatial coordinates of cell i (x_i and y_i) and the temporal index j; $catB_{ij}^{Y}, XGB_{ij}^{Y}, RF_{ij}^{Y}$ are, respectively, the predicted residual at cell i on day j from a catB, XGB, and RF; and ε_{ij}^{Y} is

the error at cell *i* on day *j*.

2.3. Cross-validation scheme for stages 2 and 3

One of the main drawbacks of flexible, complex machine learning algorithms such as gradient boosting and random forest is that they are prone to overfitting. Fair performance accuracy is ensured by preventing any information leakage between data used to train the model and data used to test the model. This implies that the model is evaluated on data that is not spatially and temporally correlated with the training data (Just et al., 2020; Meyer and Pebesma, 2022; Roberts et al., 2017). Special care must be taken in case of air pollution monitors which are highly clustered. In such a case, evaluating the model with a sample-based cross-validation scheme, where training and testing observations are picked up randomly, risks information leakage and tends towards error underestimation.

A "blocking" procedure forces testing on data that are spatially and temporally distant from training data. In time-based cross-validation, information leakage is prevented by removing training observations that are temporally close to testing observations. In site-based cross-validation, information leakage is prevented by removing training observations that belong to the same monitoring site as testing observations. In city-based cross-validation, information leakage is prevented by removing training observations that are spatially close to testing observations that belong to the same city. Methods can also be combined. For instance, with spatial and temporal buffer cross-validation, for each testing observation, some past and future training observations at that location are removed, and observations within a certain distance from the location are also removed. Blocking however presents the notable risk of reducing the length of the training dataset, thereby restricting the learning abilities of the algorithm. Roberts et al. recommend using as many cross-validation folds as computationally feasible, and blocks that are "no larger than necessary considering the grain and extent of analysis and the spatial scale of patterning of environment" (Roberts et al., 2017). To predict daily NO_2 concentrations at a 1 km spatial resolution (stage 2) and when increasing spatial resolution to 200 m across large urban areas (stage 3), we chose a spatio-temporal blocking scheme of 2 kms and 5 days. That is, we excluded observations from the training set if they were within a radius of 2 kms and a temporal distance of 5 days to any observations in the testing set (Spatial and temporal buffer cross-validation, Supplementary Fig. 1). With this procedure, training occurred on 52–84% of the available data (across folds).

An advantage of using an ensemble procedure is that the cross-validation relies on a further training step, which makes the predictive accuracy more robust to extrapolation. We used a 10 (outer) folds cross-validation procedure, to which we added a further 5 (inner) folds cross-validation. This procedure ensured that the two sets of cross-validated predictions from each basis learner, used to subsequently train vs. test the ensemble, are obtained from training on independent data. Details on the cross-validation procedure are reported in Supplementary Fig. 2.

2.4. Assessment of model performance

To compare the performance between the basis learners and ensembles over the study period and across each area (urban vs. peri-urban vs. rural areas), we used the following metrics.

- root mean squared error (RMSE), the square root of the mean observed differences between the observed and predicted values of NO_2, considered a summary measure of prediction error;
- coefficient of determination (R^2), which reflects the fraction of spatiotemporal variation in NO_2 captured by a model;
- mean absolute error (MAE), which reflects the typical difference between a model's predictions and measured concentration of NO_2.

We also evaluated whether each algorithm was better at predicting the spatial vs. temporal variability of NO_2 concentrations. To do so, we extracted spatial and temporal metrics as described by Kloog et al. (2011).

Overall temporal metrics are calculated by regressing *Delta Observed* against *Delta Predicted*, where: *Delta Observed* is the difference between the observed NO_2 concentration at place *p* and day *t* and the annual mean NO_2 at place *p*; and *Delta predicted* is defined similarly for the predicted NO_2 concentrations. Overall spatial metrics are calculated by regressing the annual mean observed NO_2 concentration at place *p* against the annual mean predicted NO_2 concentration at place *p*. We report the spatial and temporal RMSE and R^2, as well as the spatial and temporal slope and intercept. The slope is the coefficient from the linear regression between observed and predicted NO_2, and represents the multiplicative bias of the model; the intercept of the linear regression between observed and predicted NO_2 represents the additive bias.

2.5. Variable importance

Variable importance is a measure of the importance of a predictor in a model's predictive performance. Variable importance was computed for each of the predictors in stages 1, 2 and 3. For catB, we used the "PredictionValuesChange" metrics which measures by how much on average the prediction changes if the feature value changes. For XGB, we used the "gain" metrics, which is based on the training error reduced by each split across all trees. For RF, we used the "impurity" metrics, which is based on the sum of all the differences in response variance between nodes pre vs. post split where the feature of interest is used.

2.6. Final predictions of NO2 concentrations

In stages 2 and 3, we refitted a single model on all available data, and used this model to get predictions for the entire grid (Kuhn and Johnson, 2013). With this approach, the models are trained using all available data. Another approach would have not refitted a model on all available data, but instead would have used the individual models from the cross-validation procedure to get the final predictions (Hastie et al., 2009). Yet, with such an approach, (1) training would not have occurred on the complete dataset, compromising the sufficiency principle that all available data have been used for training (Roberts et al., 2017); (2) the computation time would have increased dramatically.

For large urban areas, we obtained our final ensemble 200 m predictions by adding stage 2 predictions with stage 3 predictions. For non-urban areas, our final ensemble 1 km predictions were simply stage 2 predictions. We computed maps of NO_2 concentrations averaged over example years (2005, 2013, 2022). We also computed maps on example days (on the 18th of February in 2005, 2013 and 2022). Finally, we computed a, map showing areas where predicted NO_2 concentrations, averaged over year 2022, exceeded the WHO threshold of 10 $\mu g/m^3$.

For this analysis, we used R version 4.1, and packages sf (Pebesma et al., 2023a), stars (Pebesma et al., 2023b), caret (Kuhn et al., 2023), ranger (Wright et al., 2023), xgboost (Chen et al., 2023), catboost (Prokhorenkova et al., 2018), and mgcv (Wood, 2023).

3. Results

3.1. Stage 1: filling gaps in NO_2 OMI data

Overall, total column density of NO_2 was present for 70% of the 1 km cell-days in the study domain (range over years 2005–2022: 60%–92%). Mean total column density of NO_2 from earthshine radiances measured by OMI over continental France for the years 2005–2022 was 2.7×10^{15} molecules/cm^2 (range: 2.2 to 3.6 $\times$ 10^{15} molecules/cm^2). R^2 obtained on the testing sets for the RF ranged from 0.93 in 2022 to 0.97 in 2013. Mean RMSE was of 0.66×10^{15} molecules/cm^2 (range: 0.48 to 1.06 $\times$ 10^{15} molecules/cm^2), and mean MAE was of 0.37×10^{15} molecules/cm^2 (range: 0.31 to 0.50 $\times$ 10^{15} molecules/cm^2). Impurity-based variable

importance assessment revealed that NO_2 predicted by CAMS methodology was the strongest predictor, followed by the day of the year, longitude and latitude (Supplementary Table 2).

3.2. Stage 2: predicting daily 1 km NO2 with three basis learners and ensembling the predictions

3.2.1. Data at monitor stations

NO_2 monitors were highly clustered (Fig. 1A). Over the study period, mean NO_2 concentration was of 22.8 $\mu g/m^3$ (SD = 17.7) and tended to decrease over years, with fluctuations in warm vs. colder seasons (Supplementary Fig. 3). Over the study period, mean NO_2 concentrations were lower in rural areas compared to periurban and urban areas (mean NO_2 concentration = 9.6 $\mu g/m^3$ (SD = 9.5) in rural areas, 20.2 $\mu g/m^3$ (15.2) in periurban areas, and 25.5 $\mu g/m^3$ (17.8) in urban areas; Fig. 1B). Note however, that the number of monitors was significantly lower in rural areas compared to periurban and urban areas (Fig. 1C).

3.2.2. Predictions of NO_2 concentration

Supplementary Fig. 4 shows NO_2 concentrations at a 1 km resolution, averaged over 2005–2022, predicted by each basis learner and the GAM ensemble. The GAM ensemble predictions resemble those of the basis learners. NO_2 concentrations are high in cities and on major road axes. NO_2 concentrations decreased over time, both in cities and major road axes (Supplementary Fig. 5).

Supplementary Fig. 6 shows the weights of each basis learner in the GAM ensemble when predicting NO_2 concentrations across France on the 30th of January in three different years (2005, 2013 and 2022). The relative importance of the basis learners in the GAM ensemble varied over space and time, but XGB predictions usually had the highest weights (mean weight: 0.55), followed by RF (0.31), and catB (0.19) (Supplementary Table 3). However, there were areas where XGB was

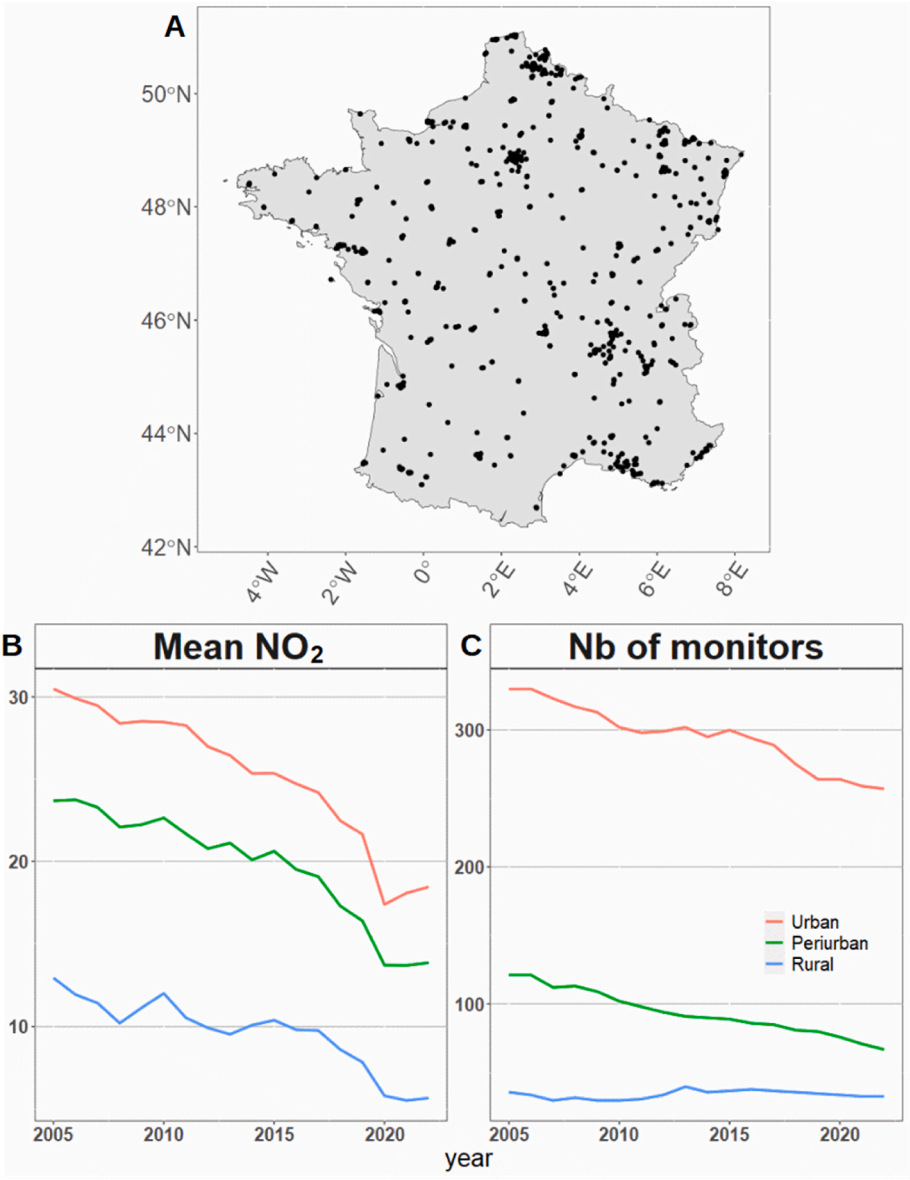

Fig. 1. Observed NO_2 concentration across France, 2005–2022.
A. Distribution of NO_2 monitoring stations across the country. Note that monitoring stations are highly clustered.
B. Mean NO_2 concentrations ($\mu g/m^3$), stratified by type of area (urban vs. periurban vs. rural). Note the decrease in mean NO_2 concentrations in urban, periurban and rural areas
C. Number of monitors per year, stratified by type of area (urban vs. periurban vs. rural). Note the decrease in the number of monitors in urban and periurban areas. Also note the small number of monitors in rural areas.

only slightly involved in the GAM ensemble (range of weights: 0.03 to 0.92). Great variations in weights were also observed for catB (range: −0.31 to 0.81) and RF (range: −0.40 to 1.14) (Supplementary Table 3).

3.2.3. Cross-validated performance

Table 1 shows the average cross-validated performance of the basis learners and GAM ensemble for the 1 km predictions. Overall, XGB was the most accurate basis learner (RMSE = 7.04 μg/m^3; R^2 = 0.82; MAE = 5.01 μg/m^3), followed by RF, with catB the least accurate (Table 1). The GAM ensemble slightly improved overall performance (RMSE = 6.98 μg/m^3; R^2 = 0.83; MAE = 4.92 μg/m^3). The basis learners and GAM ensemble captured variation in NO_2 concentration between locations better than day-to-day variation at individual locations (GAM ensemble spatial R^2 = 0.97, temporal R^2 = 0.62; GAM ensemble spatial RMSE = 1.94 μg/m^3, temporal RMSE = 9.89 μg/m^3).

Model performance varied according to the location of monitors (urban vs. periurban vs. rural). Overall, performance accuracy as measured by the R^2 was higher in urban areas than in periurban or rural areas (Fig. 2). RMSE and MAE decreased over years (Fig. 2). RF performed better than XGB and the GAM ensemble in rural areas (Fig. 2). In periurban and urban areas, the GAM ensemble performed better than RF and catB, and slightly better than XGB (Fig. 2).

Over the study period, yearly performances fluctuated around an averaged R^2 of about 0.80, while RMSE and MAE decreased over time (Supplementary Table 4). Each model yielded better spatial than temporal performance (Supplementary Table 4).

3.2.4. Variable importance

Important variables in predicting 1 km NO_2 concentrations differed according to the basis learners (Fig. 3). Most important predictors for catB were standard deviation of 2 m air temperature (t2m sd), tertiary road density, and percentage of agriculture areas (AGR); most important predictors for XGB were light at night, NDVI, and population density; and most important predictors for RF were t2m sd, elevation, and tertiary road density. Least important predictors for catB were: day of the year, distance to nearest secondary roads, and boundary layer height at 12:00 UTC (blh 12); least important predictors for XGB were percentage of urban green areas (URBGR), non-electric railways, and train strations; and least important predictors for RF were mean total cloud cover (tcc), RES, and percentage of industry or commercial areas (IND).

3.3. Stage 3: predicting daily residuals at a 200 m resolution over large urban areas

3.3.1. Data at monitor stations

There were 338 monitors in large urban areas over the study period (Supplementary Fig. 7). Mean NO_2 concentration was of 28.6 μg/m^3 (SD = 18.6).

Table 1
Cross-validated performance (averaged over 2005–2022) of the basis learners and GAM ensemble, predicting daily 1 km **NO_2** concentration across France (stage 2).

Metrics	catB	XGB	RF	ens
RMSE	7.89	7.04	7.53	6.98
R^2	0.78	0.82	0.80	0.83
MAE	5.72	5.01	5.38	4.92
Sp RMSE	2.46	1.72	2.06	1.94
Sp R^2	0.96	0.98	0.97	0.97
Sp intercept	−0.70	−0.43	−1.34	0.54
Sp slope	1.02	1.01	1.05	0.96
Tp RMSE	7.68	6.99	7.30	6.89
Tp R^2	0.53	0.61	0.57	0.62
Tp intercept	0.00	0.00	0.00	0.00
Tp slope	0.91	0.95	1.14	0.98

Legend. RMSE: root mean square error; MAE: mean absolute error; Sp: spatial; Tp: temporal; catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble.

3.3.2. Predictions of NO_2 concentration at a 200 m resolution

For all 200 m grid cell-days over the study period, the average of predicted residuals in large urban areas were of 1.13 μg/m^3 (SD = 2.93; range over 2005–2022: −39.2 to 33.0 μg/m^3).

Fig. 4 shows NO_2 concentrations over Paris predicted by the GAM ensemble and the three basis learners on three similar days in 2005, 2013 and 2022. Fig. 5 shows the final predictions of NO_2 concentrations over the whole country on three similar days in 2005, 2013 and 2022. Again, note that NO_2 concentrations decreased considerably over time, both in cities and major road axes.

The relative importance of the basis learners in the GAM ensemble varied over space and time. Over the study period, RF predictions had the highest weights (mean weight: 0.74), followed by catB (0.13), and XGB (0.10) (Supplementary Table 5). There were areas however, where RF was negatively involved in the GAM ensemble (range of weights: −0.15 to 1.26). Great variations in weights were also observed for catB (range: −0.29 to 0.64) and XGB (range: −0.24 to 0.70) (Supplementary Table 5).

Supplementary Fig. 8 presents the areas that exceeded the WHO threshold for NO_2 concentration of 10 μg/m^3 as of 2022. These were mostly located in urban areas and some of the major road axes of the country. Using a 1 km gridded map from the gridded population of the world database, we estimate that more than 30 million people live in areas with annual NO_2 concentrations above the WHO threshold.

3.3.3. Cross-validated performance

Table 2 presents the cross-validated performance of the stage 3 models predicting daily 200 m residuals. These models also performed relatively well, with overall RMSE of 4.32 μg/m^3, R^2 of 0.69, and MAE of 3.03 μg/m^3 (residual scale). This time however, RF outperformed catB and XGB. Just as for stage 2, performance tended to increase over time, and spatial performance was better than temporal performance (Supplementary Table 6).

3.4. Variable importance

Most important predictors for catB were percentage of agriculture areas (AGR), percentage of urban green areas (URBGR), percentage of industry or commercial areas (IND); most important predictors for XGB were longitude, latitude and elevation; and most important predictors for RF were longitude, AGR, and URBGR (Fig. 6). Least important predictors for catB were secondary roads density, train stations, and motorways; least important predictors for XGB were non-electric railways, motorways density, and AGR; and least important predictors for RF were secondary roads density, train stations and electric railways (Fig. 6).

Supplementary Fig. 9 shows the spatiotemporal distribution of predicted NO_2 concentrations in urban vs. non-urban areas for each year (stage 2 model). Supplementary Fig. 9 also shows the distribution of the stage 2 residuals in urban areas at a 200 m resolution (stage 3 results).

4. Discussion

High-resolution spatiotemporally-resolved air pollution datasets are essential to understanding and managing the health effects of pollutants, a pressing issue in a warming, urbanizing world. Here, we provide daily predictions of NO_2 concentration across continental France for years 2005–2022 at a high spatial resolution. Predictions were obtained after three cascading stages of modeling: (1) predicting NO_2 total column density from OMI satellite data; (2) predicting NO_2 concentrations at a 1 km spatial resolution; and (3) predicting residuals from stage 2 models at a 200 m resolution in large urban areas. The latter two stages used an ensemble of three decision-tree algorithms (random forest, extreme gradient boosting and categorical boosting), and a large set of potential predictors.

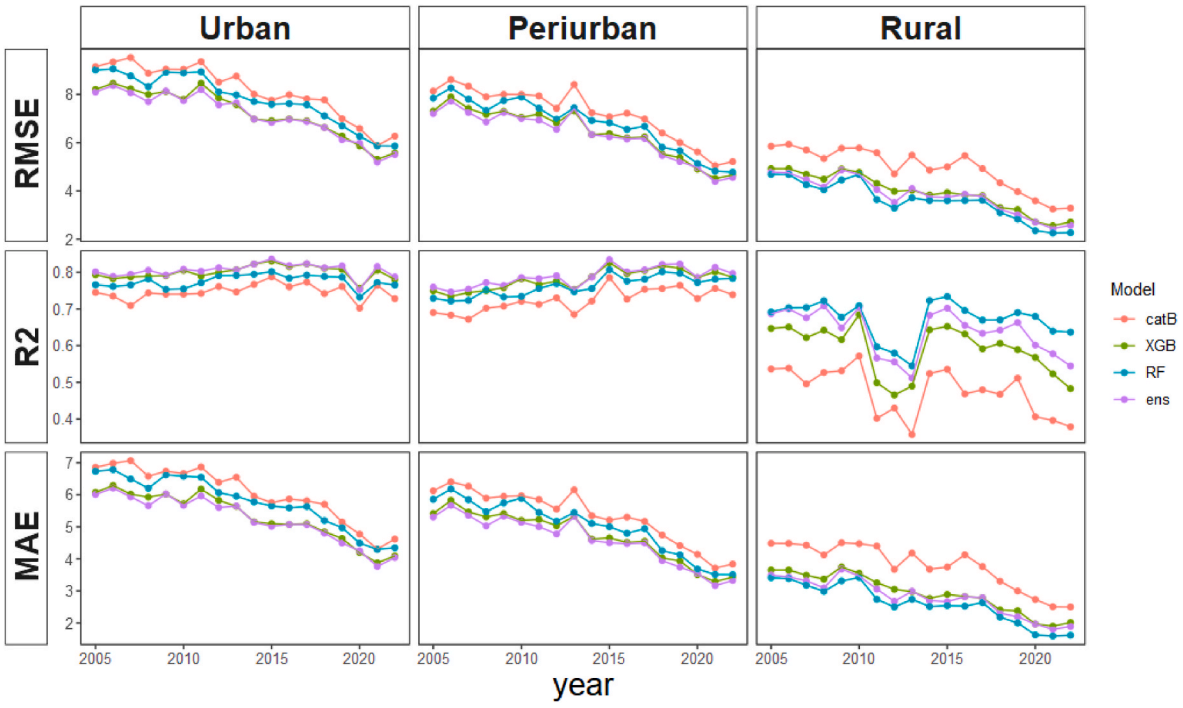

Fig. 2. Monthly cross-validated RMSE (top), R^2 (middle) and MAE (bottom) of the basis learners (catB, XGB, RF) and the GAM ensemble, predicting daily 1 km NO_2 concentrations (stage 2) in urban (left), periurban (middle), and rural areas (right).
Legend. catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble.
Note the sudden collapse in rural R^2 in 2011–2013. This could in part be due to the lower variability of NO_2 concentration in rural areas (Lower variance increases R^2). Variance of rural NO_2 dropped by more than 20% in year 2011 compared to 2010; and by more than 40% in year 2013 compared to year 2014.

Previous studies have used an ensemble of machine learning algorithms to produce NO_2 concentrations at a high spatial and temporal resolution. Yet, to our knowledge, there have been no publications applying a hybrid methodology to model daily, residence level (sub 1 km) NO_2 concentration across a very complex geo-climate region with very different spatial characteristics spanning alpine mountains and Mediterranean coasts. Likewise, other NO_2 studies have applied a spatial blocking cross-validation scheme (Kim et al., 2020; Young et al., 2016), but these studies did not produce NO_2 concentrations at a high spatial (sub 1 km) and temporal (daily) resolution. Finally, our blocking procedure considers both spatial and temporal units, which we believe provides a more robust method of cross-validation.

Predictive performance of the basis learners and the ensemble was very good. Overall, our model explained 83% of variation in NO_2 concentration at a 1 km resolution, and 69% of variation in the residuals at a 200 m resolution. Comparison with previous studies is difficult because previous reports have not always detailed their cross-validation methods and their blocking schemes. In addition, to our knowledge there are no precise guidelines regarding spatio-temporal blocking. Taking into account those limitations, our model compares well with other machine learning studies aiming at predicting NO_2 concentration over large spatial areas. For instance, de Hoogh et al. found a R^2 of 0.58 when predicting daily NO_2 at a 1 km resolution across Switzerland, and of 0.73 when predicting residuals at a 100 m resolution (de Hoogh et al., 2019 – no blocking scheme reported); Di et al. found a R^2 of 0.79 when predicting at a 1 km daily resolution in the USA (spatial R^2 of 0.84, and temporal R^2 of 0.73) (Di et al., 2020 – no blocking scheme reported); Zhan et al. found a spatial R^2 of 0.73 and a temporal R^2 of 0.62 when predicting at a 0.1° (about 11.1 km) daily resolution in China (Zhan et al., 2018 – no blocking scheme reported); and He et al. found a R^2 of 0.87 when predicting at a 1 km daily resolution and of 0.66 when predicting residuals at a 200 m resolution in Mexico city (He et al., 2023 – no blocking scheme reported). Finally, compared with a previous study aiming to predict daily NO_2 concentration over France using a chemical transport model and spatio-temporal kriging (Real et al., 2022), our study provides a better spatial resolution and improves predictive accuracy (with our model, the range of RMSE was 4–8 $\mu g/m^3$ in urban and periurban areas and 2–5 $\mu g/m^3$ in rural areas; in the previous study, the RMSE was around 10 $\mu g/m^3$ or over – no blocking scheme reported).

Our predictions are readily available to any researcher willing to carry out epidemiological research based on NO_2 concentrations to further inform policy-makers on air quality issues. In response to the continued threat of air pollution to public health, the 2021 update of the WHO air quality guidelines provides an ambitious threshold for NO_2 of 10 $\mu g/m^3$ per annum (World Health Organization, 2021). Based on our models, as of 2022, several urban regions and major road axes in France saw their annual NO_2 concentrations exceed the WHO threshold. Using a 1 km gridded map from the gridded population of the world database, we estimate that more than 30 million people lived in areas with ambient annual NO_2 concentrations above the WHO guideline. Therefore, the current study adds to the increasing body of evidence showing that a sizeable proportion of the population is at risk of health issues due to NO_2 (European Environment Agency, 2023). While NO_2 concentrations have declined over the past decade, our findings urge policymakers to amplify their efforts to address high levels of NO_2 and accelerate the transition to a carbon-free society, all the more since a slight increase in NO_2 concentrations has been observed since 2020, probably as a rebound post-COVID.

Another goal of the current study was to provide insights regarding some of the most widely used decision-tree algorithms in geospatial modeling. There were interesting differences among the basis learners in terms of predictive accuracy. Specifically, XGB models performed better than RF and catB to predict 1 km NO_2 concentrations in urban and periurban areas, but RF models performed better to predict residuals of NO_2 concentrations in large urban areas at a higher, 200 m resolution, and in rural areas at a 1 km resolution. Ensembling the predictions maximizes the chance of obtaining accurate values if one basis learner fails in a specific area or at a particular time, by relying on the other learners and

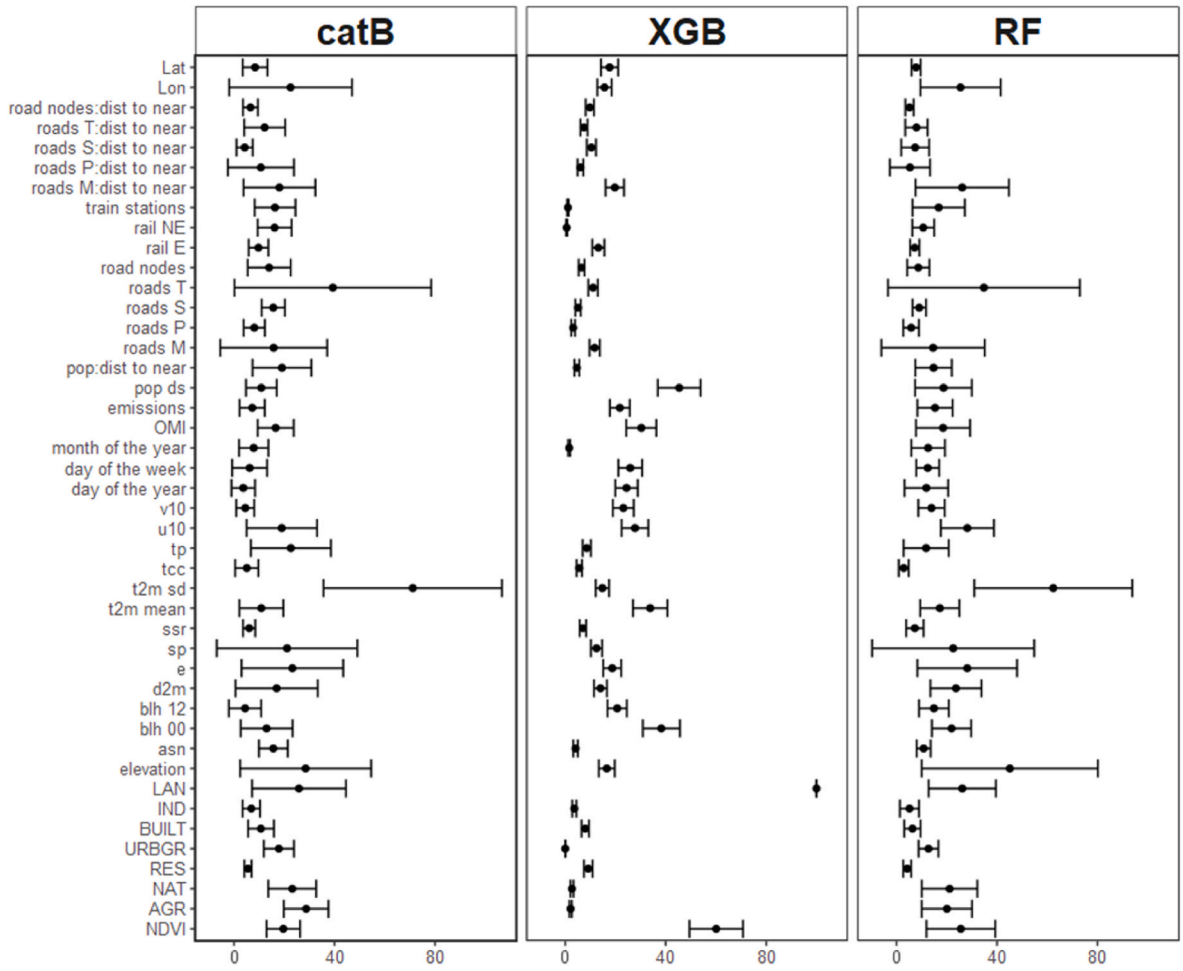

Fig. 3. Variable importance in predicting 1 km NO_2 concentrations (stage 2), stratified by basis learners (catB, XGB, and RF).
Legend. catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble; Lat: latitude; Lon: longitude; dist to near: distance to nearest; NE: non-electrified; E: electrified; roads T: tertiary roads; roads S: secondary roads; roads P: primary roads; roads M: motorways; pop: population; ds: density; emissions: NO_2 emissions; OMI: total column density of NO_2 from OMI (obtained from stage 1); v10: mean v-component of 10 m wind speed; u10: mean u-component of 10 m wind speed; tp: total precipitation; tcc: mean total cloud cover; t2m sd: standard deviation of 2 m air temperature; t2m mean: mean of 2 m air temperature; ssr: mean surface solar radiation; sp: mean surface pressure; e: mean total evaporation; d2m: mean 2 m dewpoint temperature; blh 12: boundary layer height at 12:00 UTC; blh 00: boundary layer height at 0:00 UTC; asn: mean snow albedo; LAN: light at night; IND: percentage of industry or commercial areas; BUILT: percentage of total built up areas; URBGR: percentage of urban green areas; RES: percentage of residential areas; NAT: percentage of semi-natural and forest areas; AGR: percentage of agriculture areas; NDVI: normalized difference of vegetation index. Note that importance values are scaled to the best predictor for each year.

their properties. All three basis learners were involved in the final predictions, though to various degrees, depending on the year and location at stake. For instance, though catB was the least accurate model overall when predicting NO_2 concentrations at a 1 km resolution, it played a key role in the final predictions in some areas and at given dates. Variable importance also varied across the three basis learners and over time. Some models had their performance rely on land cover such as vegetation or corine land cover predictors, others on light at night or road traffic data. Again, such variability is likely to improve the predictive accuracy of the ensemble procedure.

Differences between the 1 km vs. 200 m NO_2 distributions in urban regions were not negligible, meaning that using the 200 m predictions in epidemiological studies may decrease the risk of measurement bias (Sellier et al., 2014). Such disparities may be accounted for by influential variables of our stage 3 model (predicting residuals from the 1 km model at a 200 m resolution). Most important variables were three groups from the Corine Land Cover data product: industry/commercial buildings, agriculture and urban green. The former two are important sources of NO_2, and urban greening has been shown to decrease pollutants' concentration due to dispersion processes (Xing and Brimblecombe, 2019). Other important predictors were latitude, longitude and elevation, reflecting the fact that NO_2 has a relatively large spatial gradient, especially in urban regions (Wang et al., 2020).

4.1. Strengths and limitations

Strengths of the current study include the use of powerful machine learning algorithms, which we ensembled to get the most accurate and trustful predictions, and the use of a cross-validation scheme suited for extrapolating predictions in spatio-temporal analysis, ensuring fair estimation of performance accuracy. Very good performances were observed overall, and in urban, periurban and rural areas. In particular, our model's spatial accuracy was excellent, underlying its ability to capture spatial variation in NO_2. Overall, our predictions of daily NO_2 concentrations at a 1 km resolution over a large spatial extent, improved to a higher spatial resolution (200 m) over large urban areas, will be particularly helpful to reduce measurement error bias in future epidemiological studies.

Our study also has some limitations. First, accuracy was lower in rural areas, which could in part be due to the lower variability of rural NO_2. In addition, predictors that work well in urban areas may not work well in rural areas. Other studies had run different models in urban vs.

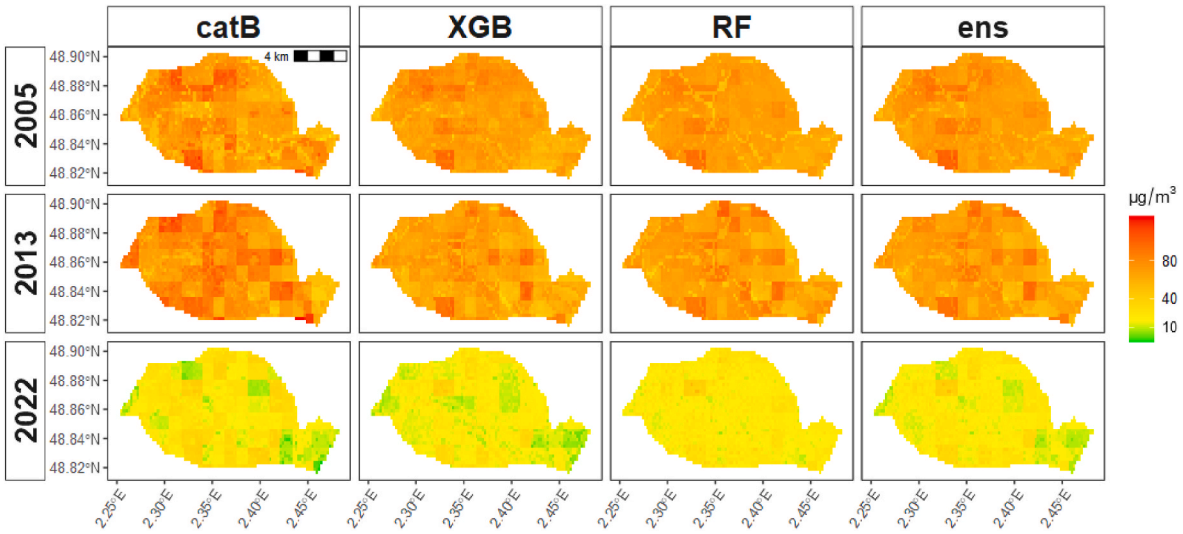

Fig. 4. Mean 24 h concentrations of NO_2 over Paris predicted by the GAM ensemble and the three basis learners at a 200 m resolution on the 18th of February in 2005, 2013 and 2022.
Legend. catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble.
Note. These predictions were obtained by downscaling 1 km NO_2 predictions (from stage 2 models) to which we added 200 m predictions of the residuals (from stage 3 models).

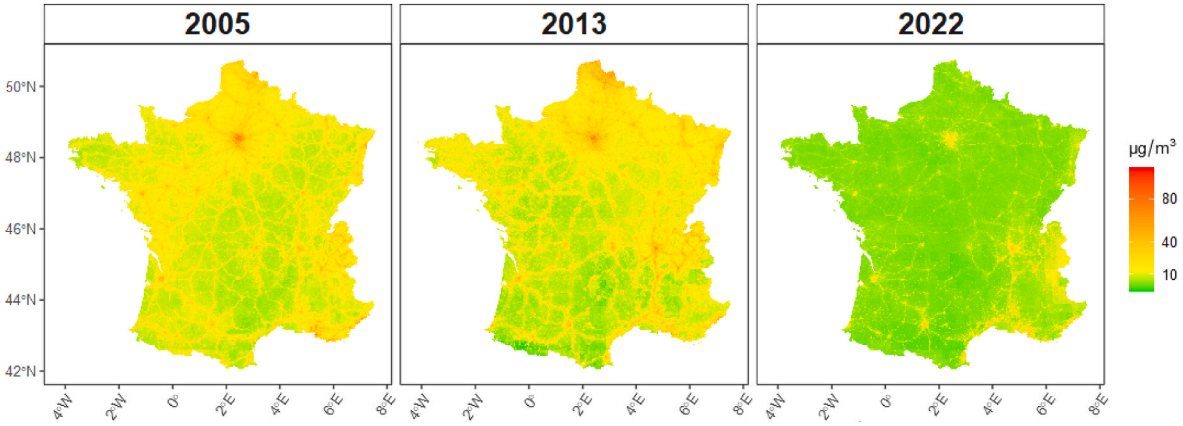

Fig. 5. Final NO_2 predictions obtained with the GAM ensemble on the 18th of February in 2005, 2013 and 2022.
Note. These predictions are at a 1 km resolution in non-urban areas and at a 200 m resolution in large urban areas. The latter were obtained by rescaling 1 km predictions (from stage 2 models) to which we added 200 m predictions of the residuals (from stage 3 models).

Table 2
Cross-validated performance (averaged over 2005–2022) of the basis learners and the GAM ensemble, predicting daily 200 m residuals at monitoring stations in large urban areas (stage 3).

Metrics	catB	XGB	RF	ens
RMSE	4.73	4.66	4.33	4.32
R^2	0.63	0.64	0.69	0.69
MAE	3.34	3.30	3.03	3.03
Sp RMSE	1.04	0.76	0.72	0.77
Sp R^2	0.97	0.98	0.98	0.98
Sp inter	−0.03	−0.01	0.00	0.01
Sp slope	1.05	1.01	1.01	1.02
Tp RMSE	4.61	4.46	4.30	4.29
Tp R^2	0.08	0.14	0.20	0.21
Tp inter	0.00	0.00	0.00	0.00
Tp slope	0.63	0.59	0.82	0.84

Legend. RMSE: root mean square error; MAE: mean absolute error; Sp: spatial; Tp: temporal; catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble.

rural regions, which was not possible here because of data scarcity in rural areas.

Second, satellite data (OMI products) are produced at a very coarse resolution. This might explain why OMI NO_2 was not in the top three predictors for any basis learner. TROPOMI products have a better resolution but are only available from year 2019.

Third, another issue with OMI data is their lack of vertical profile. NO_2 vertical profiles may have been a better contributor to ground-level NO_2 than total column measurement. However, satellite-based retrievals of tropospheric NO_2 vertical column densities are based on a monthly mean climatology of NO_2 profile shapes constructed from a chemical transport model, with a rather coarse spatial resolution (NASA, n.d.). Tropospheric column measurements obtained from satellites correlate almost perfectly with total vertical column density and therefore are subject to the same uncertainties (Travis et al., 2016). NO_2 vertical profiles are available in other satellite products (such as SCIAMACHY or LIMS data product), but they do not cover our period of interest, and/or have a coarser temporal or spatial resolution. To make up for the lack of vertical profile in our analysis, and calibrate satellite data to

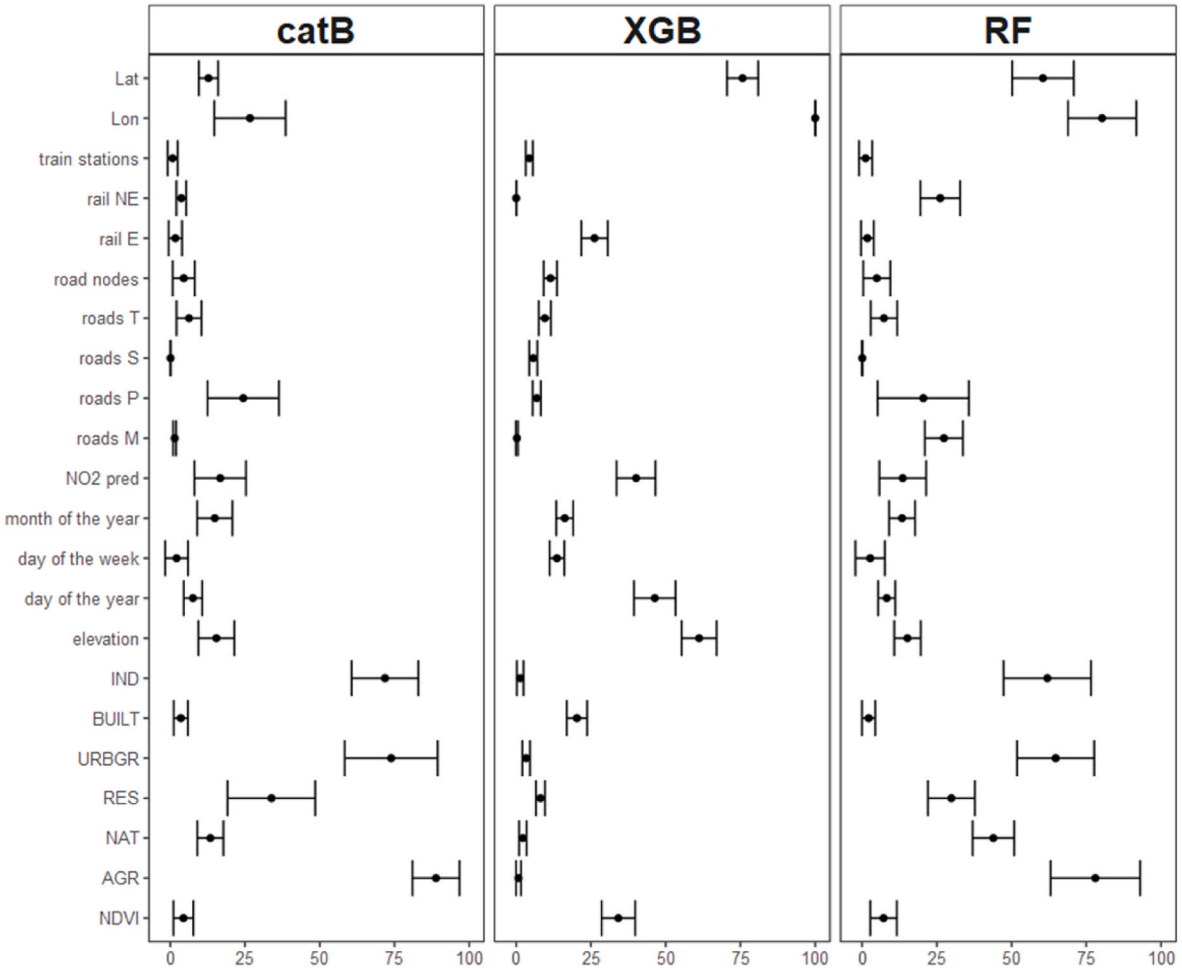

Fig. 6. Variable importance for models predicting daily 200 m residuals at monitoring stations in large urban areas (stage 3), stratified by basis learners (catB, XGB, and RF).
Legend. catB: categorical gradient boosting; XGB: extreme gradient boosting; RF: random forest; ens: ensemble; Lat: latitude; Lon: longitude; NE: non-electrified; E: electrified; roads T: tertiary roads; roads S: secondary roads; roads P: primary roads; roads M: motorways; NO_2 pred: NO_2 predictions; IND: percentage of industry or commercial areas; BUILT: percentage of total built up areas; URBGR: percentage of urban green areas; RES: percentage of residential areas; NAT: percentage of semi-natural and forest areas; AGR: percentage of agriculture areas; NDVI: normalized difference of vegetation index.
Note that importance values are scaled to the best predictors for each year.

ground-level NO_2, we instead used a variety of spatial and temporal predictors in modeling stages 2 and 3, and found that our model performed well in predicting NO_2 concentrations at breathing level. Different methods to derive surface NO_2 concentrations from satellite measurements have been tested recently (Heue et al., 2022; Wagner et al., 2021), and may be used in future studies aiming to estimate ground-level concentrations of NO_2.

Fourth, our models demonstrated very high spatial performance, but poorer temporal performance. This signs the fact that none of our predictors was able to capture the large temporal variability of NO_2. Future studies would gain from including such temporal information, e.g. daily mean road traffic data.

Fifth, our final predictions only concern ambient air pollution. These do not take into account time spent indoors, as well as citizens' behaviours regarding heating.

4.2. Conclusion

Our multi-stage approach was able to predict daily NO_2 concentrations across France at a 1 km resolution, and at a 200 m resolution across large urban areas, for the years 2005–2022, with a low error. We improved accuracy and robustness by ensembling the predictions of three basis learners. We demonstrated that all basis learners contributed to the final predictions. To the best of our knowledge, this is the first study conducted in France with such a high spatiotemporal resolution, large spatial extent, and long temporal coverage. The predictions are available to be used by health and ecosystems researchers in France and may inform policy makers on air quality issues.

Funding

This work was supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), through the IRGA 2021 PAPex project.

Data sharing

Final NO_2 predictions can be made available after the manuscript publication. Data requests needs to be formally addressed to Dr Johanna Lepeule (johanna.lepeule@univ-grenoble-alpes.fr). The code of the method is accessible by contacting the corresponding author.

CRediT authorship contribution statement

Guillaume Barbalat: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ian Hough:** Writing – review & editing, Validation, Methodology. **Michael Dorman:** Writing – review & editing, Visualization, Methodology. **Johanna Lepeule:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Itai Kloog:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

Most data processing and analyses were performed using the GRI-CAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Equip@Meso (ANR-10-EQPX-29-01).

We thank the LCSQA for providing validated NO_2 monitor data (Laurent Letinois), INERIS for providing a spatial emissions inventory, and NASA, ECMWF, Copernicus, IGN, Observatoire des Territoires, and OpenStreetMap for their open datasets.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2024.119241.

References

Bossard, M., Feranec, J., Otahel, J., 2000. CORINE Land Cover Technical Guide: Addendum 2000. European Environment Agency, Copenhagen.

Bouguerra, H., Tachi, S.E., Bouchehed, H., Gilja, G., Aloui, N., Hasnaoui, Y., Aliche, A., Benmamar, S., Navarro-Pedreño, J., 2023. Integration of high-accuracy geospatial data and machine learning approaches for soil erosion susceptibility mapping in the mediterranean region: a case study of the macta basin, Algeria. Sustainability 15, 10388. https://doi.org/10.3390/su151310388.

Bucsela, E.J., Celarier, E.A., Wenig, M.O., Gleason, J.F., Veefkind, J.P., Boersma, K.F., Brinksma, E.J., 2006. Algorithm for NO2 vertical column retrieval from the ozone monitoring instrument. IEEE Trans. Geosci. Remote Sens. 44 https://doi.org/10.1109/TGRS.2005.863715.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2023. Xgboost: Extreme Gradient Boosting implementation), Xgb. contributors (base Xgb.

de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., Hoek, G., 2018. Spatial PM2.5, NO2, O3 and BC models for western europe - evaluation of spatiotemporal stability. Environ. Int. 120, 81–92. https://doi.org/10.1016/j.envint.2018.07.036.

de Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E.A., Strassmann, A., Puhan, M., Röösli, M., Stafoggia, M., Kloog, I., 2019. Predicting fine-scale daily NO2 for 2005–2016 incorporating OMI satellite data across Switzerland. Environ. Sci. Technol. 53, 10279–10287. https://doi.org/10.1021/acs.est.9b03107.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2020. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environ. Sci. Technol. 54, 1372–1384. https://doi.org/10.1021/acs.est.9b03358.

Didan, K., Munoz, A.B., Solano, R., Huete, A., 2015. MODIS vegetation index user's guide (MOD13 series). Univ. Ariz. Veg. Index Phenol. Lab 35, 2–33.

Dijkstra, L., Hamilton, E., Lall, S., Wahba, S., 2020. How do we define cities, towns, and rural areas? [WWW Document]. URL. https://blogs.worldbank.org/sustainablecities/how-do-we-define-cities-towns-and-rural-areas, 1.2.24).

Dominski, F.H., Lorenzetti Branco, J.H., Buonanno, G., Stabile, L., Gameiro da Silva, M., Andrade, A., 2021. Effects of air pollution on health: a mapping review of systematic reviews and meta-analyses. Environ. Res. 201, 111487 https://doi.org/10.1016/j.envres.2021.111487.

Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O., Baptista, S.R., 2015. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. Pap. Appl. Geogr. 1, 226–234. https://doi.org/10.1080/23754931.2015.1014272.

Earth Observation Group, n.d. DMSP & VIIRS Data Download [WWW Document]. URL https://www.ngdc.noaa.gov/eog/download.html (accessed 1.2.24).

Earth Resources Observation And Science (EROS) Center, 2017. Global Multi-Resolution Terrain Elevation Data 2010. https://doi.org/10.5066/F7J38R2N. GMTED2010).

European Environment Agency, 2023. Europe's air quality status 2023 [WWW Document]. URL. https://www.eea.europa.eu/publications/europes-air-quality-status-2023, 1.3.24).

Fuller, R., Landrigan, P.J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., Lichtveld, M., Martin, K., Mustapha, A., Sanchez-Triana, E., Sandilya, K., Schaefli, L., Shaw, J., Seddon, J., Suk, W., Téllez-Rojo, M.M., Yan, C., 2022. Pollution and health: a progress update. Lancet Planet. Health 6, e535–e547. https://doi.org/10.1016/S2542-5196(22)00090-0.

Guilbert, A., Bernard, J.Y., Peyre, H., Costet, N., Hough, I., Seyve, E., Monfort, C., Philippat, C., Slama, R., Kloog, I., Chevrier, C., Heude, B., Ramus, F., Lepeule, J., 2023. Prenatal and childhood exposure to ambient air pollution and cognitive function in school-age children: examining sensitive windows and sex-specific associations. Environ. Res. 235, 116557 https://doi.org/10.1016/j.envres.2023.116557.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7.

He, M.Z., Yitshak-Sade, M., Just, A.C., Gutiérrez-Avila, I., Dorman, M., de Hoogh, K., Mijling, B., Wright, R.O., Kloog, I., 2023. Predicting fine-scale daily NO2 over Mexico city using an ensemble modeling approach. Atmospheric Pollut. Res. 14, 101763 https://doi.org/10.1016/j.apr.2023.101763.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Heue, K.-P., Loyola, D., Romahn, F., Zimmer, W., Chabrillat, S., Errera, Q., Ziemke, J., Kramarova, N., 2022. Tropospheric ozone retrieval by a combination of TROPOMI/S5P measurements with BASCOE assimilated data. Atmospheric Meas. Tech. 15, 5563–5579. https://doi.org/10.5194/amt-15-5563-2022.

Hough, I., Just, A.C., Zhou, B., Dorman, M., Lepeule, J., Kloog, I., 2020. A multi-resolution air temperature model for France from MODIS and Landsat thermal data. Environ. Res. 183, 109244 https://doi.org/10.1016/j.envres.2020.109244.

Hough, I., Rolland, M., Guilbert, A., Seyve, E., Heude, B., Slama, R., Lyon-Caen, S., Pin, I., Chevrier, C., Kloog, I., Lepeule, J., 2023. Early delivery following chronic and acute ambient temperature exposure: a comprehensive survival approach. Int. J. Epidemiol. 52, 761–773. https://doi.org/10.1093/ije/dyac190.

Hough, I., Sarafian, R., Shtein, A., Zhou, B., Lepeule, J., Kloog, I., 2021. Gaussian Markov random fields improve ensemble predictions of daily 1 km PM2.5 and PM10 across France. Atmos. Environ. 264, 118693 https://doi.org/10.1016/j.atmosenv.2021.118693.

IGN, 2021. ROUTE 500® | Géoservices [WWW Document]. URL. https://geoservices.ign.fr/route500, 1.2.24).

INERIS, n.d. Institut national de l'environnement industriel et des risques [WWW Document]. URL https://www.ineris.fr/en (accessed 1.2.24).

Inness, A., Blechschmidt, A.-M., Bouarar, I., Chabrillat, S., Crepulja, M., Engelen, R.J., Eskes, H., Flemming, J., Gaudel, A., Hendrick, F., Huijnen, V., Jones, L., Kapsomenakis, J., Katragkou, E., Keppens, A., Langerock, B., de Mazière, M., Melas, D., Parrington, M., Peuch, V.H., Razinger, M., Richter, A., Schultz, M.G., Suttie, M., Thouret, V., Vrekoussis, M., Wagner, A., Zerefos, C., 2015. Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS. Atmospheric Chem. Phys. 15, 5275–5303. https://doi.org/10.5194/acp-15-5275-2015.

Insee, 2023. Estimation de la population au 1er janvier 2023 [WWW Document]. URL. https://www.insee.fr/fr/statistiques/1893198, 1.2.24).

Inventaire National Spatialisé, n.d. Carte - Inventaire National Spatialisé [WWW Document]. URL http://emissions-air.developpement-durable.gouv.fr/map.html?name=metropole (accessed 1.2.24).

Just, A.C., Arfer, K.B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., Kloog, I., 2020. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. Atmospheric Environ. Oxf. Engl. 239, 117649 https://doi.org/10.1016/j.atmosenv.2020.117649, 1994.

Kamińska, J.A., 2018. The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław. J. Environ. Manage. 217, 164–174. https://doi.org/10.1016/j.jenvman.2018.03.094.

Kim, S.-Y., Bechle, M., Hankey, S., Sheppard, L., Szpiro, A.A., Marshall, J.D., 2020. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: role of prediction model parsimony in integrated empirical geographic regression. PLOS ONE 15, e0228535. https://doi.org/10.1371/journal.pone.0228535.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos. Environ. 45, 6267–6275. https://doi.org/10.1016/j.atmosenv.2011.08.066.

Kuhn, M., 2014. Futility analysis in the cross-validation of machine learning models. ArXiv Prepr. ArXiv14056974.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Core Team, R., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2023. Caret: Classification and Regression Training.

LCSQA, n.d. Accueil | LCSQA [WWW Document]. URL https://www.lcsqa.org/fr (accessed 1.2.24).

Mainka, A., Żak, M., 2022. Synergistic or antagonistic health effects of long- and short-term exposure to ambient NO2 and PM2.5: a review. Int. J. Environ. Res. Public. Health 19, 14079. https://doi.org/10.3390/ijerph192114079.

Mellander, C., Lobo, J., Stolarick, K., Matheson, Z., 2015. Night-time light data: a good proxy measure for economic activity? PLOS ONE 10, e0139779. https://doi.org/10.1371/journal.pone.0139779.

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat. Commun. 13, 2208. https://doi.org/10.1038/s41467-022-29838-9.

NASA, n.d. GES DISC Dataset: OMI/Aura NO2 Total and Tropospheric Column Daily L2 Global Gridded 0.25 degree x 0.25 degree V3 (OMNO2G 003) [WWW Document]. URL https://disc.gsfc.nasa.gov/datasets/OMNO2G_003/summary (accessed 5.13.24).

Pan, Y., Zhao, C., Liu, Z., 2021. Estimating the daily NO2 concentration with high spatial resolution in the beijing–tianjin–hebei region using an ensemble learning model. Remote Sens. 13, 758. https://doi.org/10.3390/rs13040758.

Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pedersen, T.L., Baston, D., Dunnington, D., 2023a. Sf: Simple Features for R.

Pebesma, E., Sumner, M., Racine, E., Fantini, A., Blodgett, D., 2023b. Stars: Spatiotemporal Arrays, Raster and Vector Data Cubes.

Pommier, M., 2023. Estimations of NOx emissions, NO2 lifetime and their temporal variation over three British urbanised regions in 2019 using TROPOMI NO2 observations. Environ. Sci. Atmospheres 3, 408–421. https://doi.org/10.1039/D2EA00086E.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 31.

Real, E., Couvidat, F., Ung, A., Malherbe, L., Raux, B., Gressent, A., Colette, A., 2022. Historical reconstruction of background air pollution over France for 2000–2015. Earth Syst. Sci. Data 14, 2419–2443. https://doi.org/10.5194/essd-14-2419-2022.

Richmond-Bryant, J., Long, T.C., 2020. Influence of exposure measurement errors on results from epidemiologic studies of different designs. J. Expo. Sci. Environ. Epidemiol. 30, 420–429. https://doi.org/10.1038/s41370-019-0164-z.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881.

Sellier, Y., Galineau, J., Hulin, A., Caini, F., Marquis, N., Navel, V., Bottagisi, S., Giorgis-Allemand, L., Jacquier, C., Slama, R., Lepeule, J., EDEN Mother–Child Cohort Study Group, 2014. Health effects of ambient air pollution: do different methods for estimating exposure lead to different results? Environ. Int. 66, 165–173. https://doi.org/10.1016/j.envint.2014.02.001.

Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., Just, A.C., Stafoggia, M., 2020. Estimating daily PM2.5 and PM10 over Italy using an ensemble model. Environ. Sci. Technol. 54, 120–128. https://doi.org/10.1021/acs.est.9b04279.

Travis, K.R., Jacob, D.J., Fisher, J.A., Kim, P.S., Marais, E.A., Zhu, L., Yu, K., Miller, C.C., Yantosca, R.M., Sulprizio, M.P., Thompson, A.M., Wennberg, P.O., Crounse, J.D., St Clair, J.M., Cohen, R.C., Laughner, J.L., Dibb, J.E., Hall, S.R., Ullmann, K., Wolfe, G. M., Pollack, I.B., Peischl, J., Neuman, J.A., Zhou, X., 2016. Why do models overestimate surface ozone in the Southeast United States? Atmospheric Chem. Phys. 16, 13561–13577. https://doi.org/10.5194/acp-16-13561-2016.

Vermote, E., Wolfe, R., 2015. MOD09GA MODIS/terra surface reflectance daily L2G global 1kmand 500m SIN grid V006. https://doi.org/10.5067/MODIS/MOD09GA.006.

Wagner, A., Bennouna, Y., Blechschmidt, A.-M., Brasseur, G., Chabrillat, S., Christophe, Y., Errera, Q., Eskes, H., Flemming, J., Hansen, K.M., Inness, A., Kapsomenakis, J., Langerock, B., Richter, A., Sudarchikova, N., Thouret, V., Zerefos, C., 2021. Comprehensive evaluation of the Copernicus Atmosphere monitoring service (CAMS) reanalysis against independent observations: reactive gases. Elem. Sci. Anthr. 9, 00171 https://doi.org/10.1525/elementa.2020.00171.

Wang, Y., Bechle, M.J., Kim, S.-Y., Adams, P.J., Pandis, S.N., Pope, C.A., Robinson, A.L., Sheppard, L., Szpiro, A.A., Marshall, J.D., 2020. Spatial decomposition analysis of NO2 and PM2.5 air pollution in the United States. Atmos. Environ. 241, 117470 https://doi.org/10.1016/j.atmosenv.2020.117470.

Wood, S., 2023. Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.

World Health Organization, 2021. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide [WWW Document]. URL. https://www.who.int/publications-detail-redirect/9789240034228, 1.2.24).

Wright, M.N., Wager, S., Probst, P., 2023. Ranger: A Fast Implementation of Random Forests.

Xing, Y., Brimblecombe, P., 2019. Role of vegetation in deposition and dispersion of air pollution in urban parks. Atmos. Environ. 201, 73–83. https://doi.org/10.1016/j.atmosenv.2018.12.027.

Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO2 and model validation in a national prediction model based on universal kriging and land-use regression. Environ. Sci. Technol. 50, 3686–3694. https://doi.org/10.1021/acs.est.5b05099.

Zhan, Y., Luo, Y., Deng, X., Zhang, K., Zhang, M., Grieneisen, M.L., Di, B., 2018. Satellite-based estimates of daily NO2 exposure in China using hybrid random forest and spatiotemporal kriging model. Environ. Sci. Technol. 52, 4180–4189. https://doi.org/10.1021/acs.est.7b05669.

Zhang, X., Just, A.C., Hsu, H.-H.L., Kloog, I., Woody, M., Mi, Z., Rush, J., Georgopoulos, P., Wright, R.O., Stroustrup, A., 2021. A hybrid approach to predict daily NO2 concentrations at city block scale. Sci. Total Environ. 761, 143279 https://doi.org/10.1016/j.scitotenv.2020.143279.