Gaussian Markov random fields improve ensemble predictions of daily 1 km $PM_{2.5}$ and PM_{10} across France

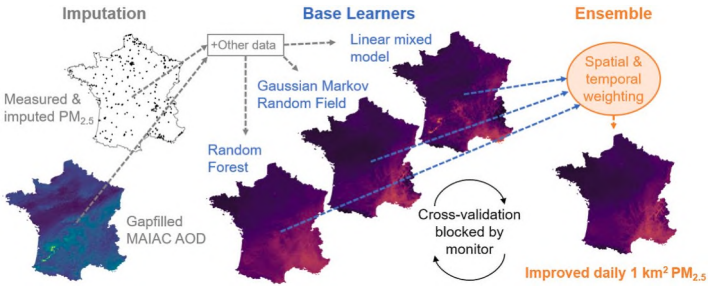
Ian Hough [a,b,*], Ron Sarafian [b,c], Alexandra Shtein [b], Bin Zhou [b,d], Johanna Lepeule [a,1], Itai Kloog [b,1]

[a] Univ. Grenoble Alpes, Inserm, CNRS, IAB, La Tronche, France
[b] Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[c] Department of Industrial Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[d] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

HIGHLIGHTS

- We modelled daily 1 km $PM_{2.5}$ and PM_{10} concentrations in France 2000–2019.
- We ensembled random forests, Gaussian Markov random fields, and mixed models.
- Imputing $PM_{2.5}$ at more common PM_{10} monitors increased the ensemble's accuracy.
- Gaussian Markov random fields were the most accurate component of the ensemble.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Particulate matter
Exposure assessment
Aerosol optical depth
Ensemble model
Epidemiology

ABSTRACT

Understanding the health impacts of particulate matter (PM) requires spatiotemporally continuous exposure estimates. We developed a multi-stage ensemble model that estimates daily mean $PM_{2.5}$ and PM_{10} at 1 km spatial resolution across France from 2000 to 2019. First, we alleviated the sparsity of $PM_{2.5}$ monitors by imputing $PM_{2.5}$ at more common PM_{10} monitors. We also imputed missing satellite aerosol optical depth (AOD) based on modelled AOD from atmospheric reanalyses. Next, we trained three base learners (mixed models, Gaussian Markov random fields, and random forests) to predict daily PM concentrations based on AOD, meteorology, and other variables. Finally, we generated ensemble predictions using a generalized additive model with spatio-temporally varying weights that exploit the strengths and weaknesses of each base learner. The Gaussian Markov random field dominated the ensemble, outperforming mixed models and random forests at most locations on most days. Rigorous cross-validation showed that the ensemble predictions were quite accurate, with mean absolute error (MAE) of 2.72 μg/m^3 and R^2 of 0.76 for $PM_{2.5}$; PM_{10} MAE was 4.26 μg/m^3 and R^2 0.71. Our predictions are available to improve epidemiological studies of acute and chronic PM exposure in urban and rural France.

* Corresponding author. Univ. Grenoble Alpes, Inserm, CNRS, IAB, La Tronche, France.
E-mail address: ian.hough@univ-grenoble-alpes.fr (I. Hough).
[1] Co-senior authors.

1. Introduction

Ambient particulate matter (PM) air pollution is a leading environmental health risk and the 7th overall risk factor for death and disease worldwide (Murray et al., 2020). To protect public health, the World Health Organization sets guideline limits (WHO, 2006) for both chronic (e.g. annual mean) and acute (e.g. daily mean) exposure to PM with a diameter of 10 μm or less (PM_{10}) and 2.5 μm or less ($PM_{2.5}$), and many countries maintain air quality monitoring networks to ensure compliance with national limits. But these networks are not always sufficient for epidemiological studies: monitors tend to be clustered in cities, which makes it difficult to estimate exposure for suburban and rural populations, and are often too sparse to capture spatial variation in PM concentrations within a city.

Recently, methods have been developed to improve exposure estimates for epidemiological studies by modelling the continuous spatiotemporal distribution of PM. One approach is to use chemical transport models that simulate the formation, dispersion, and deposition of PM based on emissions, meteorology, and atmospheric chemistry. These are effective at large and small scales and can be used to forecast future PM concentrations (Zhang et al., 2012a), but are limited by the accuracy of the input emissions and meteorological data and the completeness of their representation of atmospheric physics and chemistry (Zhang et al., 2012b). They are also computationally intensive, although advances in computing have allowed increasing resolutions to be used over larger areas and timescales. For example, a recent study estimated PM concentration in France in 2010 and 2011 by combining a national model at 4 km-hourly resolution, 7 regional models at 3–4 km resolution, and 43 urban models at 10–200 m resolution (Riviere et al., 2019).

An alternative approach is to calibrate a statistical model that relates measured PM concentration to variables such as aerosol optical depth (AOD), a measure of the absorption and scattering of light by particles suspended in the atmosphere. AOD can be retrieved by satellite instruments across large areas at fairly high spatiotemporal resolution (e.g. 1 km-daily), making it a useful proxy for the spatial and temporal distribution of ground-level PM. However, satellite AOD is often missing due to cloud cover, glint on snow or water, or instrument malfunction. Early studies accommodated this by calibrating two relationships: one to estimate PM based on AOD and another to estimate PM when AOD was not available (Hu et al., 2014; Kloog et al., 2011). Recently, methods have been developed to fill gaps in satellite AOD based on modelled AOD from atmospheric reanalyses, allowing AOD-based prediction for all days and locations (Di et al, 2016, 2019, 2016; Stafoggia et al., 2019).

Another challenge is that the relationship between PM and AOD varies over both time and space. Studies in the United States (Chudnovsky et al, 2012, 2014; Hu et al., 2014; Kloog et al, 2011, 2014; Lee et al., 2011, 2016), Europe (Beloconi et al., 2016; de Hoogh et al., 2018; Nordio et al., 2013; Stafoggia et al., 2017), China (Liang et al., 2018; Xiao et al., 2017; Xie et al., 2015; Zhang et al., 2018; Zheng et al., 2016), Mexico (Just et al., 2015), and Israel (Kloog et al., 2015; Shtein et al., 2018) have used mixed models to allow the PM – AOD relationship to vary from day to day and between regions. This approach performs well and is computationally cheap but implies sharp changes in the daily PM – AOD relationship at the borders of predefined regions, which may be unrealistic. Smooth spatiotemporal variation is possible for large datasets with a Gaussian Markov random field (GMRF) solved via integrated nested Laplace approximations (INLA) (Lindgren et al., 2011; Rue et al., 2009). Recently, Sarafian et al. (2019) showed that GMRFs with spatially smooth daily random effects predicted daily $PM_{2.5}$ more accurately than mixed models both near to and far from monitors in the northeastern United States. GMRFs have also performed well predicting daily PM_{10} in northwest Italy (Cameletti et al., 2013) and annual mean PM across Europe (Beloconi et al., 2018).

Other studies have used various statistical approaches, including geographically weighted regression (Hu et al., 2013; Ma et al., 2014; Song et al., 2014; Van Donkelaar et al., 2016), geographically and temporally weighted regression (Guo et al., 2017; He and Huang, 2018; Liu et al., 2020), and machine learning algorithms such as random forests (Chen et al., 2019; Hu et al., 2017; Schneider et al., 2020; Stafoggia et al, 2019, 2020), gradient boosting (Chen et al., 2019; Just et al., 2020), and neural networks (Chen et al., 2019; Di et al., 2016; Park et al., 2020; Yan et al., 2020). Machine learning algorithms perform particularly well as they can capture complex nonlinear relationships, and recent work has shown that performance can be slightly improved by ensembling predictions from multiple base learners (Di et al., 2019; Murray et al., 2019; Shtein et al., 2019; Zhai and Chen, 2018). However, the flexibility of machine learning algorithms makes them vulnerable to overfitting, so it is important to evaluate their accuracy on independent data to ensure they can generalize to unmonitored locations (Just et al., 2020; Sarafian et al., 2019). A common approach is cross-validation (CV): data are repeatedly split into training and test sets, the model is calibrated using only the training data, and its predictions are compared to the held-out test data. Since PM concentrations are often spatiotemporally autocorrelated, the splitting must be done in a way that ensures test data are far in space and time from training data. Recent studies have done this using spatial blocking, holding out all data from a group of monitors (Just et al., 2020; Meng et al., 2021; Murray et al., 2019; Park et al., 2020; Pu and Yoo, 2021; Schneider et al., 2020; Shtein et al., 2019; Stafoggia et al., 2019; Xiao et al., 2020) or temporal blocking, holding out all data from one year (He et al., 2021; Meng et al., 2021; Pu and Yoo, 2021; Xiao et al., 2020; Yan et al., 2020). Ensemble models require special care: data held out to test the ensemble should not have been used to train the base learners, and ensembles should be calibrated using base learner CV predictions (predictions for held-out test data), as these reflect each base learner's ability to generalize to new areas (Shtein et al., 2019).

The goal of this study was to estimate $PM_{2.5}$ and PM_{10} concentrations across continental France from 2000 through 2019. To the best of our knowledge, this is the first study to estimate daily 1 km PM concentration in France over two decades using satellite data and geostatistical models. We incorporated recent methodological advances to mitigate the sparsity of $PM_{2.5}$ monitors and fill gaps in satellite AOD (Stafoggia et al., 2019), and developed the first ensemble model of daily PM concentration that incorporates a GMRF. We used a rigorous cross-validation scheme to estimate accuracy and provide the first evidence that GMRFs may predict daily PM concentration more accurately than random forests.

2. Materials

2.1. Study domain

Continental France covers a roughly hexagonal area of 542,973 km^2 in western Europe bounded by the Atlantic Ocean to the west and the Mediterranean Sea to the southeast (Fig. 1). Most of the terrain is at low elevation, but the Pyrenees in the southwest rise to over 3000 m and in the southeast the Alps reach 4809 m. Annual mean temperature ranges from about 0 °C at high elevations to about 17 °C in the Mediterranean southeast (Hough et al., 2020). The population is approximately 64.5 million, of which 12.5 million (20%) live in the greater Paris metropolis. About 20% of the population is rural, and 37% live in towns and small cities with fewer than 500,000 residents (Insee, 2020). For this study, we defined a grid of 632,571 approximately 1 km^2 cells covering continental France coincident with the pixels of the satellite AOD data (section 2.3). We considered the 7245 days from 1 March 2000 through 31 December 2019, giving a total study domain of 4.58×10^9 cell-days.

2.2. Air quality monitoring data

We obtained hourly $PM_{2.5}$ and PM_{10} ($\mu g/m^3$) measurements from 12 regional air quality monitoring networks (federated by ATMO France) through the French Central Air Quality Monitoring Laboratory. Monitors

were mostly clustered in urban areas; the number of PM_{10} monitors increased from 222 to 330 and the number of $PM_{2.5}$ monitors increased from 9 to 142 over the course of the study period (Fig. 1). To limit the impact of instrument malfunctions and rare events, we excluded hourly $PM_{2.5}$ concentrations >200 μg/m^3 and hourly PM_{10} concentrations >300 μg/m^3 (0.003% of all observations). We indexed each monitor to the containing 1 km grid cell and calculated daily mean PM for days with at least 18 hourly observations.

2.3. Aerosol optical depth

We obtained satellite-derived 0.469 μm AOD at approximately 1 km spatial resolution from the Moderate Resolution Imaging Spectroradiometer (MODIS) Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product (MCD19A2v006) (Lyapustin et al., 2018). MCD19A2v006 provides AOD up to four times per day (between 9:00 and 15:00 UTC in France). We used the quality assurance band to identify all "best quality" observations; we also included "land; research quality" and "clear; within 2 km of coast" as these represent potentially useable observations over urban areas and coasts where there are few "best quality" observations. We indexed these observations to the 1 km grid (whose cells were defined to coincide with the MAIAC AOD pixels) and calculated daily mean AOD.

To fill gaps in MAIAC AOD (mostly due to cloud cover), we obtained modelled 3-hourly 0.469 μm AOD at approximately 80 km spatial resolution from the Copernicus Atmospheric Monitoring Service EAC4 Reanalysis (Inness et al., 2019). Since EAC4 begins on 1 March 2003, for 1 March 2000 to 28 February 2003 we obtained modelled hourly 0.55 μm AOD for 08:30 to 15:30 UTC at approximately 60 km spatial resolution from the MERRA2 reanalysis (Randles et al., 2017). We bilinearly interpolated EAC4 and MERRA2 AOD to the 1 km grid, giving 8 values per cell-day (0 UTC, 3 UTC, …, 21 UTC for EAC4; 08:30 UTC, 09:30 UTC, …, 15:30 UTC for MERRA2).

2.4. Meteorology

Meteorological parameters such as wind, rain, temperature, and the height of the planetary boundary layer affect surface PM concentrations and indicate the extent to which AOD represents aerosols near the surface or higher in the atmosphere. We obtained hourly meteorological parameters at approximately 30 km spatial resolution from the Copernicus Climate Change Service ERA5 reanalysis (Hersbach et al., 2020). We bilinearly interpolated the parameters to the 1 km grid and calculated 10 daily values: boundary layer height at 0:00 and 12:00 UTC, total precipitation, mean and standard deviation of 2m air temperature, mean 2m dewpoint temperature, mean surface pressure, mean u- and v-components of 10m wind speed, and mean cloud cover.

2.5. Normalized difference vegetation index

Vegetation may influence PM dispersion and the density of PM sources. We obtained monthly composite normalized difference vegetation index (NDVI) at approximately 1 km^2 spatial resolution from the MODIS MOD13A3v006 product (Didan et al., 2015), which is spatially coincident with the MAIAC AOD data. We indexed NDVI to the 1 km grid, filled rare missing values with the Gaussian kernel mean of nearby cells, and used the same value for every day of each month.

2.6. Spatial predictors

In addition to the previous spatiotemporal predictors, we used impervious surfaces, land cover, road and railway density, elevation, population, climatic region, distance to coast, and $PM_{2.5}$ and PM_{10} emissions as indicators of the typical spatial distribution of PM. Since

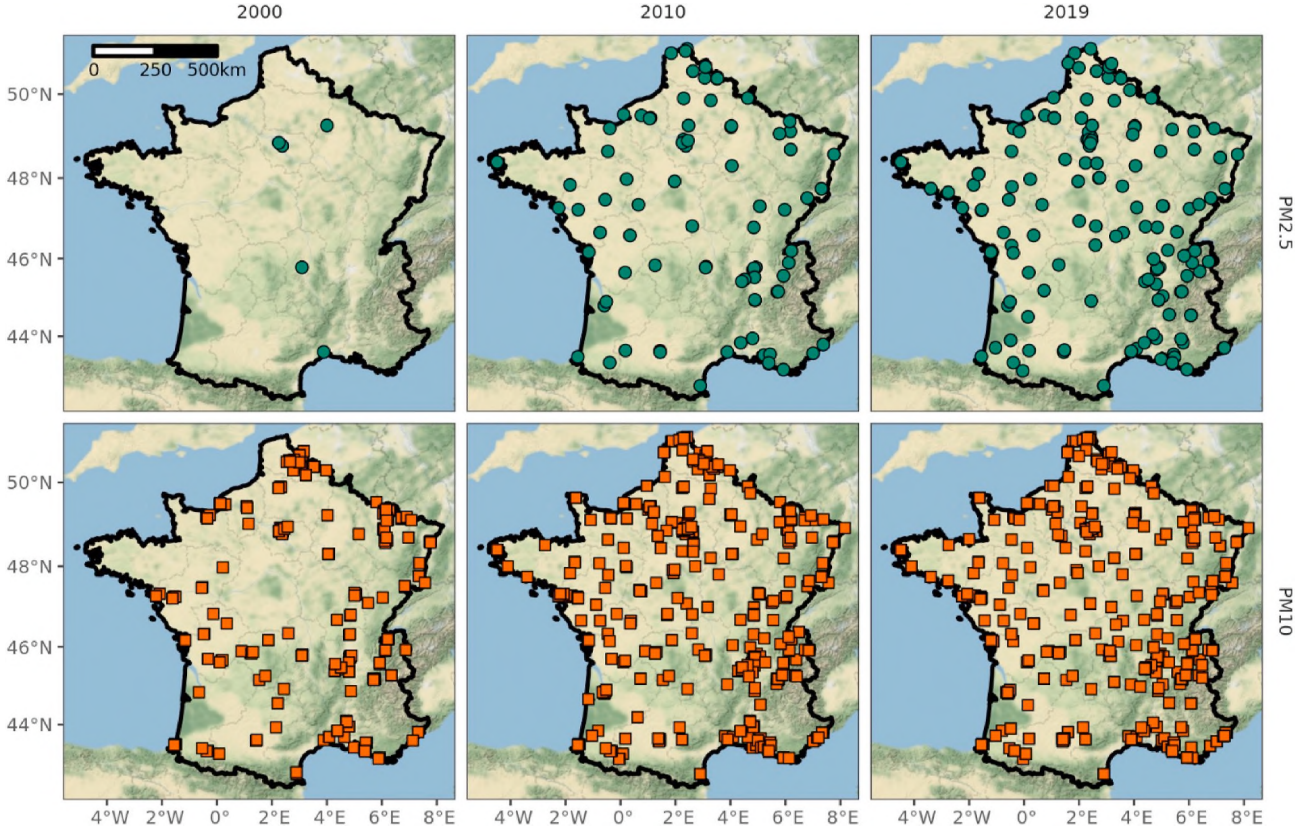


Fig. 1. Spatial distribution of $PM_{2.5}$ monitors (top row) and PM_{10} monitors (bottom row) in continental France from 2000 to 2019. Basemap by Stamen Design, under CC BY 3.0.

these data are time invariant, we used the value from the closest reference year for every day of each year. Supplemental Table S1 describes how we indexed these data to the 1 km grid and derived 19 spatial predictors.

3. Methods

We used a four-stage process to predict $PM_{2.5}$ and PM_{10} for the 4.58 $\times$ 10^9 1 km grid cell-days in the study domain (Fig. 2). Briefly, we: 1) alleviated the sparsity of $PM_{2.5}$ monitors by training a random forest (RF) to impute daily $PM_{2.5}$ at monitors that only measured PM_{10}; 2) filled gaps in MAIAC AOD data by training monthly RFs to impute missing MAIAC AOD based on co-located EAC4 or MERRA2 AOD; 3) trained three base learners for each year (linear mixed models [LMM], Gaussian Markov random fields [GMRF], and random forests [RF]) to predict daily 1 km PM based on gap-filled AOD, meteorology, NDVI, and spatial predictors; 4) increased accuracy by ensembling the base learner predictions with annual generalized additive models (GAM) that weight the base learners according to spatiotemporal variations in their performance. We performed all data processing and statistical analyses in R 3.6.3 (R Core Team, 2020) using the packages lme4 for LMM (Bates et al., 2015), R-INLA for GMRF (Bakka et al., 2018), ranger for RF (Wright and Ziegler, 2017) with mlr and mlrMBO for tuning via model-based optimization (Bischl et al, 2016, 2017), and mgcv for GAM (Wood, 2017).

3.1. Stage 1: imputing $PM_{2.5}$ at PM_{10} monitors

Most monitors in France measured PM_{10} but not $PM_{2.5}$. To mitigate the sparsity of $PM_{2.5}$ monitors, we used all co-located daily measures of $PM_{2.5}$ and PM_{10} (n = 474,761) to tune and train a RF of 500 trees to predict $PM_{2.5}$ based on measured PM_{10} and monitor characteristics:

$$PM_{2.5_{mt}} = f\left(PM_{10_{mt}}, vol_{mt}, loc_{mt}, infl_{mt}, lat_m, lon_m, wday_t, yday_t, date_t\right) + \varepsilon_{mt} \quad (1)$$

where $PM_{2.5_{mt}}$ and $PM_{10_{mt}}$ are, respectively, the $PM_{2.5}$ and PM_{10} measured by monitor m (1, …, 205) on day t (1, …, 7245); vol_{mt} indicates whether on day t monitor m excluded, included, or included an estimate of the semi-volatile fraction of PM_{10}; loc_{mt} and $infl_{mt}$ are, respectively, the locale (rural, suburban, or urban) and predominant influence (traffic, industrial, or background) of monitor m on day t; lat_m and lon_m are the latitude and longitude of monitor m; $wday_t$, $yday_t$, and $date_t$ are, respectively, the day of week (to capture trends related to commuting or business activity), day of year (to capture seasonal trends), and date (to capture long-term trends); and ε_{mt} is the error at monitor m on day t. To reduce bias in the variable importance estimates, we sampled 63.2% of observations without replacement for each tree and estimated importance by permutation.(Strobl et al., 2007) We tuned mtry (the number of variables to consider at each split) to minimize mean absolute error via model-based optimization and estimated accuracy using 5-fold CV blocked by monitor (section 3.5). We used the RF to impute $PM_{2.5}$ for the 1.71 $\times$ 10^6 monitor-days where only PM_{10} was measured.

3.2. Stage 2: filling gaps in MAIAC AOD

Clouds and snow cover often prevented MAIAC AOD retrieval over part of the study area. To fill these gaps, we trained RFs to predict MAIAC AOD based on co-located modelled AOD from atmospheric reanalysis. For computational reasons, we used 96 trees per forest, tuned using one spatiotemporally blocked 50% subsample of the data (section 3.5), and trained one RF for each month in the study period (mean observations per month $\approx$ 4.36 $\times$ 10^6):

$$AOD_{st}^{M} = f(R_{1_{st}}, \ldots, R_{8_{st}}, x_s, y_s, wday_t, yday_t) + \varepsilon_{st} \quad (2)$$

where for each month M (1, …, 238), AOD_{st}^{M} is the MAIAC AOD observed at 1 km grid cell s (1, …, 632571) on day t (1, …, number of days in month M); $R_{h_{st}}$ is the AOD from atmospheric reanalysis (MERRA2 before 1 January 2003; EAC4 otherwise) at cell s on day t at each of eight times (8:30 UTC, 9:30 UTC, …, 15:30 UTC for MERRA2; 0 UTC, 3 UTC, …, 21 UTC for EAC4); x_s and y_s are the spatial coordinates of cell s; $wday_t$ and $yday_t$ are, respectively, the day of week and day of year; and ε_{st} is the error at cell s on day t. We estimated accuracy using 5-fold CV with spatiotemporal blocking (see section 3.5) and used the RFs to predict AOD for the 3.54 $\times$ 10^9 1 km grid cell-days without MAIAC AOD.

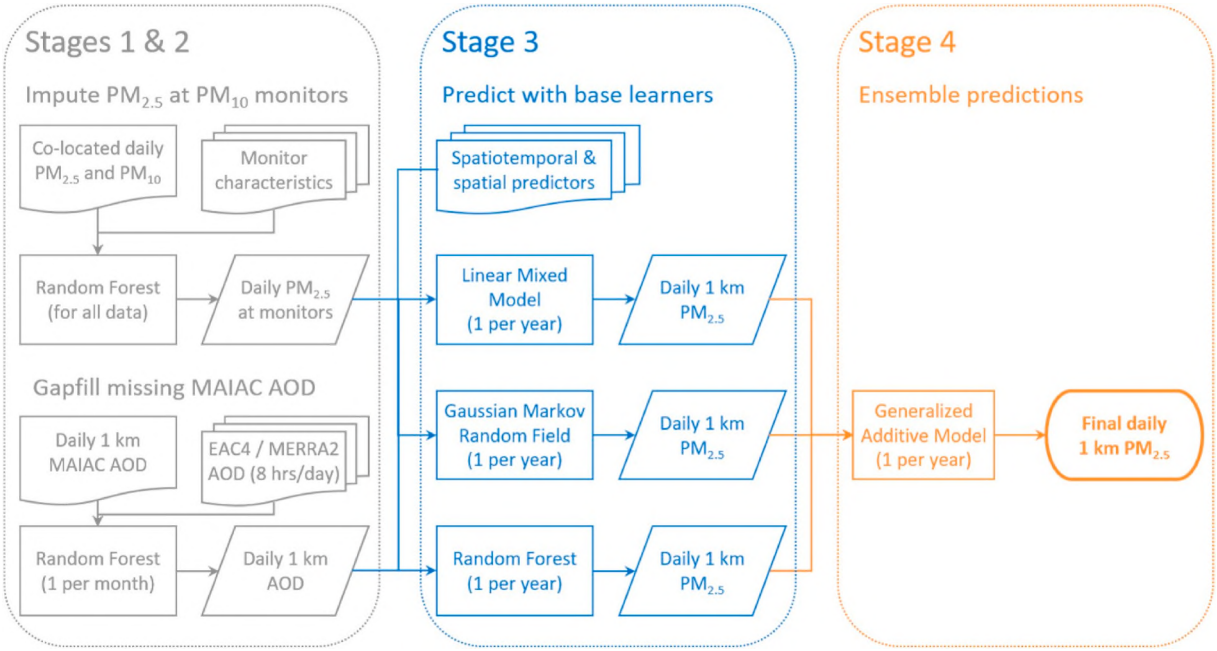


Fig. 2. Flowchart of four-stage process to predict daily 1 km $PM_{2.5}$.

3.3. Stage 3: predicting daily 1 km PM using three base learners

In stage 3, we trained LMMs, GMRFs, and RFs to predict $PM_{2.5}$ (from stage 1) and PM_{10} based on gap-filled MAIAC AOD (from stage 2), 11 spatiotemporal predictors (sections 2.4 and 2.5), and 19 spatial predictors (section 2.6). We scaled the predictors to have similar range, and for the LMMs and GMRFs we log-transformed PM to approximate normality and prevent negative predictions. We trained each base learner for each of $PM_{2.5}$ and PM_{10} in each year, yielding 120 base models (mean observations per year = 111,610; range 54,353 to 126,544). We estimated the accuracy of each base model using multi-stage CV blocked by monitor (section 3.5) and used the base models to predict PM for the 4.58×10^9 1 km grid cell-days of the study domain.

3.3.1. Linear mixed models

For each PM size fraction in each year, we calibrated a LMM with a random effect that allowed the PM-AOD relationship to vary daily for each of 8 climatic regions:

$$\log\left(PM_{st}^{FY}\right) = \left(\alpha^{FY} + \mu_{tr}^{FY}\right) + \left(\beta_{AOD}^{FY} + \nu_{tr}^{FY}\right)AOD_{st} + \sum_{p=1}^{32}\left(\beta_p^{FY} X_{p_{st}}\right) + \varepsilon_{st}^{FY} \quad (3)$$

where for each PM size fraction F (2.5 or 10) and year Y (2000, …, 2019), $\log(PM_{st}^{FY})$ is the log-transformed PM concentration at 1 km grid cell s (1, …, 632571) on day t (1, …, number of days in year Y); α^{FY} is the fixed intercept and μ_{tr}^{FY} is the random intercept on day t for the climatic region r that contains cell s; β_{AOD}^{FY} is the fixed slope of AOD and ν_{tr}^{FY} is the random slope of AOD on day t for the climatic region r that contains cell s; AOD_{st} is the AOD at cell s on day t. β_p^{FY} is the coefficient and $X_{p_{st}}$ the value at cell s on day t for each of the 11 spatiotemporal predictors, 19 spatial predictors, and sine and cosine transforms of the day of week; and ε_{st}^{FY} is the error at cell s on day t.

3.3.2. Gaussian Markov random fields

For each PM size fraction in each year, we calibrated a GMRF with a spatiotemporal random effect that varied smoothly over space on each day:

$$\log\left(PM_{st}^{FY}\right) = \alpha^{FY} + \beta_{AOD}^{FY}AOD_{st} + \sum_{p=1}^{32}\left(\beta_p^{FY} X_{p_{st}}\right) + \omega_{st}^{FY} + \varepsilon_{st}^{FY} \quad (4)$$

where F, Y, s, t, $\log(PM_{st}^{FY})$, α^{FY}, β_{AOD}^{FY}, AOD_{st}, β_p^{FY}, $X_{p_{st}}$, and ε_{st}^{FY} are as in equation (3), and ω_{st}^{FY} is the spatiotemporal random effect at cell s on day t. We assumed that the error was independent and identically distributed (i.i.d.) following $\mathcal{N}(0, \sigma_\varepsilon^2)$ and the spatiotemporal random effect was temporally i.i.d. with Matérn spatial covariance, i.e.:

$$Cov(\omega_{st}, \omega_{s't'}) = \begin{cases} 0, & t \neq t' \\ \sigma_\omega^2 C(d_{ss'}; \rho_\omega), & t = t' \end{cases} \quad (5)$$

where σ_ω^2 is the variance of the spatiotemporal random effect, $\mathscr{C}$ is the Matérn function, $d_{ss'}$ is the Euclidean distance between locations s and s', and ρ_ω is a hyperparameter that governs the range (distance at which the correlation falls to less than about 10%). We assigned penalized complexity priors to σ_ε^2, σ_ω^2, and ρ_ω that shrank the spatiotemporal random effect towards the null (Fuglstad et al., 2019; Simpson et al., 2017) and fit the model using INLA.

3.3.3. Random forests

For each PM size fraction in each year, we trained a RF with the equation:

$$PM_{st}^{FY} = f\left(AOD_{st}, X_{1_{st}}, \ldots, X_{30_{st}}, x_s, y_s, wday_t, yday_t\right) + \varepsilon_{st}^{FY} \quad (6)$$

where F, Y, s, t, PM_{st}^{FY}, AOD_{st}, and ε_{st}^{FY} are as in equation (3); $X_{1_{st}}, \ldots, X_{30_{st}}$ are, respectively, the value for each of the 11 spatiotemporal predictors and 19 spatial predictors at cell s on day t; x_s and y_s are the spatial coordinates of cell s; and $wday_t$ and $yday_t$ are the day of week and day of year. We used 250 trees and fixed mtry at 5 because exploratory tuning suggested that mtry >5 provided little benefit and risked overfitting.

3.4. Stage 4: ensembling daily 1 km PM predictions to improve accuracy

In stage 4, we calibrated a GAM to ensemble the predictions of the stage 3 base learners. We used predictions for held-out monitors to calibrate the GAMs because these reflect accuracy at unmonitored locations (section 3.5). We fit a GAM for each PM size fraction in each year (20 GAMs total; mean observations per year = 111,610; range 54,353 to 126,544) using tensor product smooths that allowed the coefficient for each base learner's predictions to vary smoothly over space and time:

$$PM_{st}^{FY} = te(x_s, y_s, t)LMM_{st}^{FY} + te(x_s, y_s, t)GMRF_{st}^{FY} + te(x_s, y_s, t)RF_{st}^{FY} + \varepsilon_{st}^{FY} \quad (7)$$

where for each PM size fraction F (2.5 or 10) and year Y (2000, …, 2019), PM_{st}^{FY} is the PM concentration at 1 km grid cell s (1, …, 632571) 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; LMM_{st}^{FY}, $GMRF_{st}^{FY}$, and RF_{st}^{FY} are, respectively, the CV prediction at cell s on day t from a LMM, GMRF, and RF that were trained while holding out all data from the fold that contains cell s; and ε_{st}^{FY} is the error at cell s on day t. We estimated accuracy using multi-stage CV blocked by monitor (section 3.5) and used the GAMs to predict AOD for the 4.58×10^9 1 km grid cell-days of the study domain.

3.5. Cross-validation

To limit bias due to spatiotemporal autocorrelation and avoid information leakage between the base learners and the ensemble, we extended the blocked CV scheme described by Shtein et al. (2019).

In stage 1 (imputing $PM_{2.5}$ at PM_{10} monitors), we estimated accuracy at monitors that never measured $PM_{2.5}$ using 5-fold CV blocked by monitor: we randomly assigned each monitor that measured both $PM_{2.5}$ and PM_{10} to one of 5 folds. This ensured that no observations from test monitors were in the training set.

In stage 2 (filling gaps in MAIAC AOD data), we estimated accuracy at locations far from same-day MAIAC AOD observations (because MAIAC AOD tends to be missing in spatial clumps). We used 5-fold CV with spatiotemporal blocking: we split the study area into 50 regions and randomly assigned MAIAC AOD in each day-region to one of 5 folds (Fig. S1). This ensured that no same-day observations from test regions were in the training set.

In stages 3 and 4 (predicting daily 1 km PM with three base learners and ensembling the predictions), we estimated accuracy at unmonitored locations. We also trained the stage 4 ensembles on base learner predictions for held-out monitors (to emulate performance at unmonitored locations) and ensured that data held out to test the ensemble had not been used to train the base learners. We used multi-stage CV blocked by monitor (Fig. S2). In stage 3, we randomly assigned each monitor to one of 5 folds, and for each combination of three of the folds (10 combinations total) we trained the base learners and predicted for the two held-out test folds. In stage 4, we used the test predictions from all base learners that held out the same one of the five folds to train an ensemble and predict for the held-out fold. This ensured that the ensemble was trained on base learner predictions for held-out monitors and evaluated at monitors that were held out from both the base learners and the ensemble.

3.6. Performance metrics

We evaluated the models using mean absolute error (MAE), which reflects the typical difference between a model's predictions and measured PM, R^2, which reflects the fraction of spatiotemporal variation in PM captured by a model, and root mean squared error (RMSE), which can be compared with the standard deviation (SD) of measured PM to see by how much a model improves upon a naïve prediction of the mean. We also split each of these metrics into a spatial and temporal component as described by Kloog et al. (2011).

4. Results

4.1. Stage 1: imputing $PM_{2.5}$ at PM_{10} monitors

Mean $PM_{2.5}$ was 13.7 μg/m^3 (standard deviation [SD] 9.9 μg/m^3) and mean PM_{10} was 21.6 μg/m^3 (SD 12.5 μg/m^3) across all monitors in continental France for 2000–2019 (Fig. S3). PM concentrations declined over the study period and were generally highest in winter and lowest in summer (Fig. S4). The cross-validated predictions of the stage 1 RF showed good correspondence with observed $PM_{2.5}$ at monitors that were not used to train the RF ($R^2 = 0.87$, MAE = 2.48 μg/m^3) with little bias (mean error = −0.157) but a tendency to underestimate very high concentrations (Fig. 3). Performance was good even in early years when there were few $PM_{2.5}$ monitors ($R^2 \geq 0.82$ in every year except 2008; MAE <2.5 μg/m^3 in most years) (Table S2). The drop in performance in 2007 and 2008 coincided with a change in monitor technology that increased measured PM_{10} concentrations and likely complicated the relationship between PM_{10} and $PM_{2.5}$. PM_{10} concentration was by far the most important predictor of $PM_{2.5}$ concentration (Fig. S5).

4.2. Stage 2: filling gaps in MAIAC AOD

Mean MAIAC AOD over continental France for 2000–2019 was 0.126 (SD 0.084); MAIAC AOD was missing for 77% of the 1 km cell-days in the study domain (Fig. S6), similar to other areas (Di et al., 2019; Schneider et al., 2020; Stafoggia et al., 2019). Cross-validated R^2 for the stage 2 RFs typically ranged from about 0.55 in winter to about 0.78 in summer (Fig. S7), coinciding with fewer MAIAC observations in winter and more in summer. MAE typically ranged from about 0.025 in fall to about 0.034 in summer, coinciding with lower AOD in fall and higher AOD in summer. Performance was similar between periods that used modelled AOD from MERRA2 vs. EAC4 as predictors. There was a slight tendency to overestimate high AOD (slope = 0.94), but average performance was good (mean $R^2 = 0.70$; mean MAE = 0.030) (Table S3). Prior to 2003, the most important predictors of MAIAC AOD were modelled AOD at 12 UTC, the day of year, and the spatial y coordinate (Fig. S8). From 2003 on, modelled AOD at 15 UTC was the second most important predictor. This likely related to the mid-2002 launch of the Aqua satellite, which passes over continental France around 13 UTC; previously, MAIAC AOD was only available around 11 UTC from the Terra satellite. Fig. S9 shows an example of gapfilled AOD.

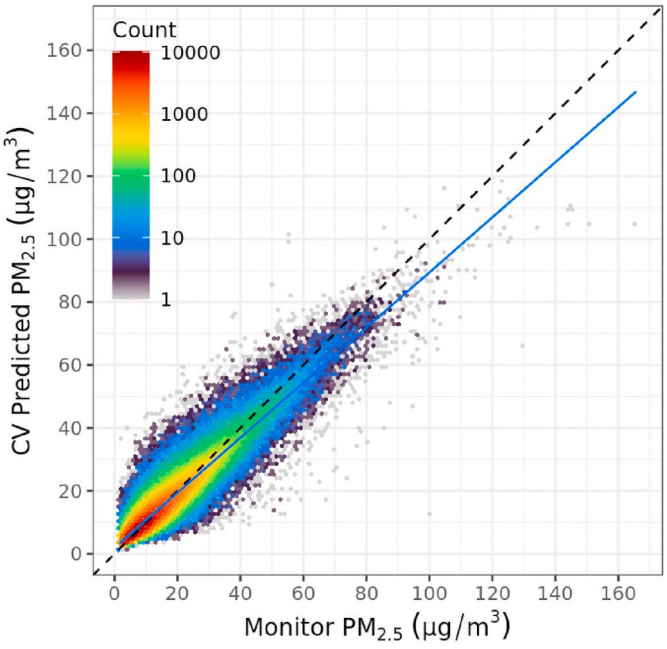


Fig. 3. Cross-validation (CV) predicted vs. observed daily $PM_{2.5}$ concentrations from the stage 1 random forest. Dashed black line shows 1:1 relationship; solid blue line shows actual relationship ($R^2 = 0.873$; MAE = 2.48 μg/m^3; mean error = −0.157). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.3. Stages 3 and 4: predicting daily 1 km PM_{10} and $PM_{2.5}$ with three base learners and ensembling the predictions to increase accuracy

Table 1 shows the average cross-validated performance of the stage 3 base learners and stage 4 GAM ensemble. GMRF was the most accurate base learner (mean $PM_{2.5}$ $R^2 = 0.75$, MAE = 2.72 μg/m^3; mean PM_{10} $R^2 = 0.70$, MAE = 4.26 μg/m^3), followed by RF, with LMM the least accurate. The stage 4 GAM ensembles slightly improved performance (mean $PM_{2.5}$ $R^2 = 0.76$, MAE = 2.72 μg/m^3; mean PM_{10} $R^2 = 0.71$, MAE = 4.26 μg/m^3), almost eliminating the GMRFs' slight bias and increasing spatial R^2 compared to both the GMRFs and RFs. The relative importance of the base learners in the GAM ensemble varied over space and time (Fig. S10), but GMRF predictions usually had the highest weight, consistent with their high cross-validated accuracy.

R^2 for all models increased in early years with the number of monitors and remained high from 2009 to 2019; MAE covaried with mean observed PM (Fig. 4). The sharp increase in PM_{10} MAE in 2007 coincided with a change in monitor technology: in 2007, all PM_{10} monitors were modified to measure semi-volatile particles in addition to non-volatile particles, increasing observed PM_{10} concentrations. $PM_{2.5}$ monitors were modified in 2008 and 2009, corresponding to the increase in $PM_{2.5}$ MAE in 2008 and 2009. R^2 was highest in winter and spring and lowest in summer; MAE was highest in winter and lowest in summer, corresponding to typically higher PM concentrations in winter and lower concentrations in summer (Fig. S11).

The base learners and GAM ensemble captured day-to-day variation in PM concentration at individual locations better than between-locations differences in annual mean PM concentration (GAM ensemble spatial $R^2 \approx 0.46$, temporal $R^2 \approx 0.80$). This is in part because PM concentration varies more over time than space; spatial MAE was lower than temporal MAE ($PM_{2.5}$ GAM ensemble spatial MAE = 1.6, temporal MAE = 2.2), indicating that predicted annual mean PM concentrations were quite accurate. It may also reflect difficulty capturing spatial variation in PM concentrations in urban areas. The lowest spatial R^2 and highest spatial MAE were in Île-de-France, the densely populated region that contains Paris, which also had the highest and most variable PM concentrations (Table S4, Fig. S12). There may not have been enough monitors for the model to capture complex spatial variation in PM concentration over greater Paris.

Since the majority of our $PM_{2.5}$ data consisted of imputed $PM_{2.5}$ concentration at PM_{10} monitors from the stage 1 RF, we performed a sensitivity analysis comparing the cross-validated predictions of the GAM ensemble to only observed $PM_{2.5}$ concentration at $PM_{2.5}$ monitors. Apart from 2000 (when there were only 5 $PM_{2.5}$ monitors), performance was similar at $PM_{2.5}$ monitors (mean $R^2 = 0.77$, mean MAE = 2.98) and across all monitors (mean $R^2 = 0.77$, mean MAE = 2.71). We also constructed an alternate model by retraining the base learners and ensemble using only $PM_{2.5}$ monitors for all years except 2000. This alternate model was less accurate (mean $R^2 = 0.66$, mean MAE = 3.62)

Table 1
Cross-validated (CV) performance (averaged over 2000–2019) of the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensemble (GAM) predicting daily 1 km PM ($\mu g/m^3$).

	Observed PM			Multi-stage CV Performance								
			Model	Total					Spatial		Temporal	
	Mean	SD[a]		RMSE	Bias[b]	Slope[c]	R^2	MAE	R^2	MAE	R^2	MAE
$PM_{2.5}$	13.8	8.5	LMM	5.03	0.67	0.63	0.63	3.35	0.38	1.82	0.68	2.92
			GMRF	4.09	0.46	0.75	0.75	2.72	0.45	1.68	0.81	2.24
			RF	4.52	−0.14	0.63	0.70	3.18	0.47	1.69	0.74	2.68
			GAM	4.02	−0.01	0.76	0.76	2.72	0.49	1.63	0.81	2.23
PM_{10}	21.5	11.9	LMM	7.65	1.04	0.60	0.58	5.21	0.32	2.98	0.64	4.39
			GMRF	6.40	0.73	0.72	0.70	4.26	0.39	2.79	0.78	3.36
			RF	7.07	−0.19	0.57	0.64	5.00	0.41	2.80	0.70	4.11
			GAM	6.28	−0.02	0.72	0.71	4.26	0.43	2.71	0.78	3.34

[a] Standard deviation.
[b] Mean error.
[c] Slope of regression of CV predicted PM on observed PM.

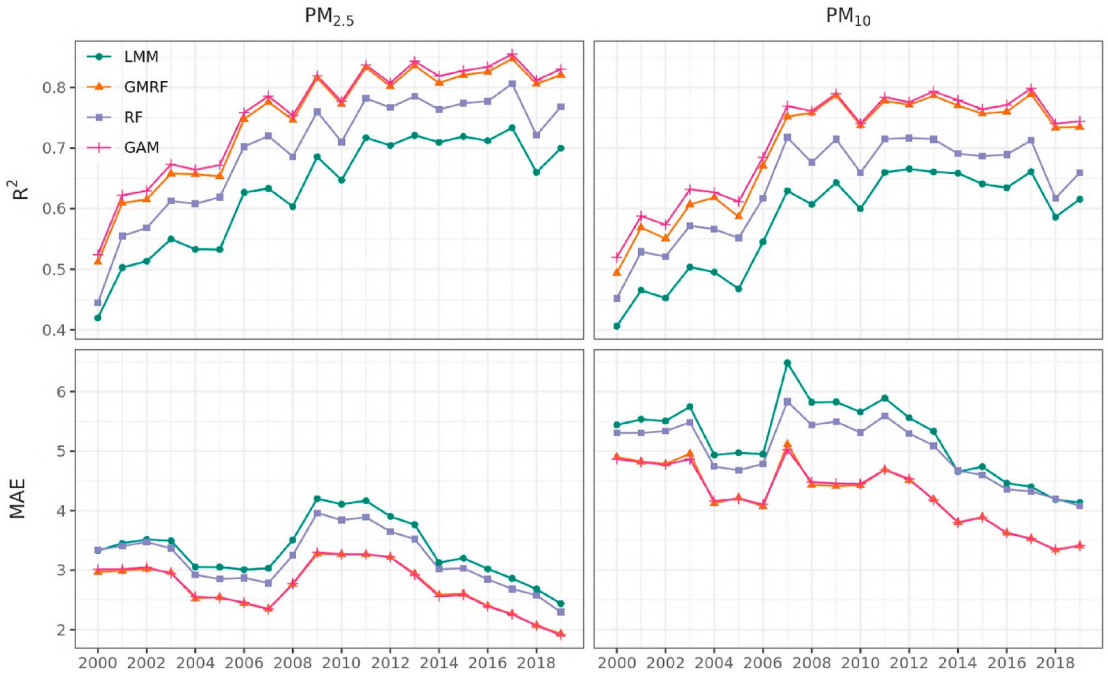


Fig. 4. Annual cross-validated R^2 (top) and MAE (bottom; $\mu g/m^3$) of the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensemble (GAM) predicting daily 1 km $PM_{2.5}$ (left) and PM_{10} (right).

than the stage 4 GAM ensemble evaluated only at $PM_{2.5}$ monitors (Fig. 5), indicating that increasing the quantity of training data by imputing $PM_{2.5}$ at PM_{10} monitors resulted in more accurate final $PM_{2.5}$ predictions than if we had relied solely on $PM_{2.5}$ monitors.

Fig. 6 shows the mean 2000–2019 $PM_{2.5}$ and PM_{10} concentration predicted by each base learner and GAM ensemble. The LMM and GMRF predictions are similar; the RF predictions are slightly higher in rural areas. The GAM ensemble predictions resemble those of the GMRF with some contribution from the RF in the southeast and southwest. PM concentrations are high in the north with a hotspot over greater Paris. In the southeast, high concentrations extend south down the Rhône river valley from the hotspot of greater Lyon, east into alpine valleys, and along the Mediterranean coast. The lowest concentrations are over the sparsely populated south centre, the Pyrenees in the southwest, and the peninsulas of Bretagne and Cotentin in the northwest. $PM_{2.5}$ concentrations show less contrast between urban and rural areas than PM_{10} concentrations.

Fig. 7 shows $PM_{2.5}$ and PM_{10} concentrations over greater Paris predicted by the GAM ensemble on three days. PM concentrations are higher over built-up areas; on two days major roads stand out as lines of elevated PM.

5. Discussion

Our finding that GMRFs predicted daily 1 km PM concentration more accurately than LMMs is consistent with results in the northeastern United States (Sarafian et al., 2019). Our finding that GMRFs were also more accurate than RFs is novel and of note, as RFs and other tree-based machine learning algorithms performed well in several recent studies (Di et al., 2019; Just et al., 2020; Schneider et al., 2020; Stafoggia et al, 2019, 2020). We emphasize the importance of careful performance evaluation when using flexible machine learning algorithms: GMRFs had the best cross-validated performance (corresponding to accuracy at unmonitored locations), but RFs performed better than GMRF on non-independent training data (corresponding to accuracy at monitors). Evaluation methods that do not ensure independence between training and testing data risk mistaking good performance at monitors for good performance everywhere.

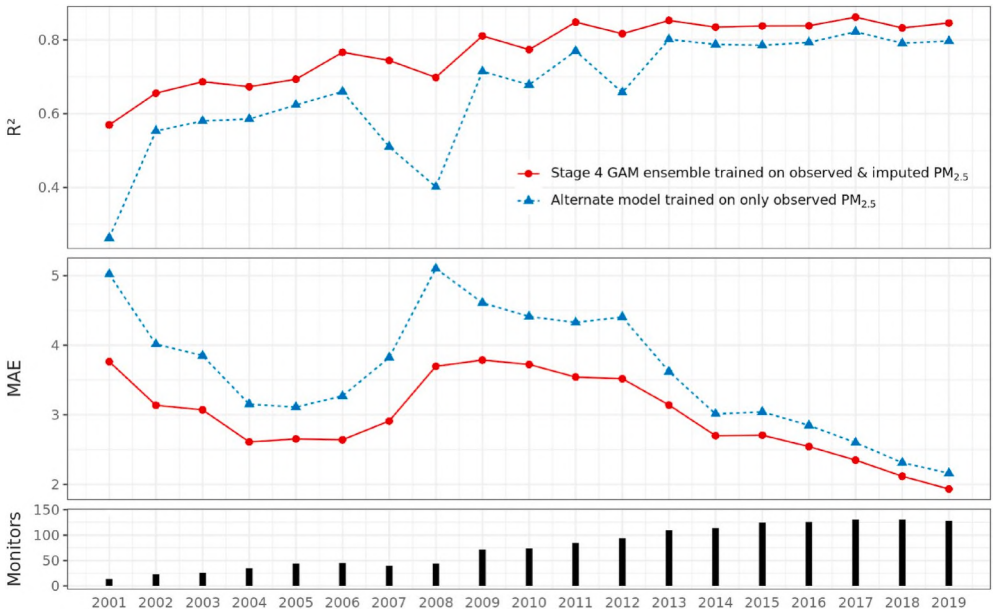


Fig. 5. Annual cross-validated performance at $PM_{2.5}$ monitors of the stage 4 GAM ensemble (red circles) and an alternate model (blue triangles) trained on only observed $PM_{2.5}$. Top: R^2; middle: MAE ($\mu g/m^3$); bottom: number of $PM_{2.5}$ monitors. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

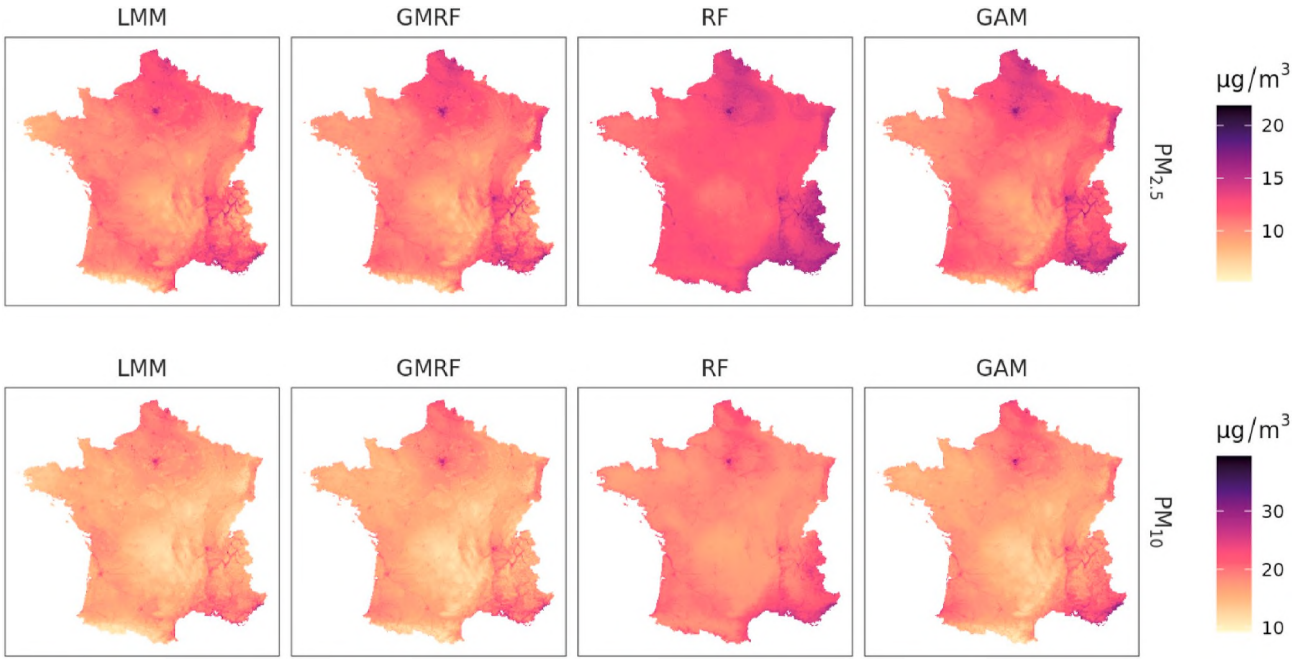


Fig. 6. Mean $PM_{2.5}$ (top) and PM_{10} (bottom) concentration predicted by the stage 3 base learners (LMM, GMRF, RF) and stage 4 ensembles (GAM) for 2000–2019.

Our GAM ensemble captured temporal variation in PM concentration better than spatial variation. We attempted to improve spatial performance in urban areas by downscaling the residuals of the GAM ensemble using RFs trained on the high spatial resolution predictors listed in Table S1, as was done in a few previous studies (Di et al., 2019; Kloog et al., 2014; Stafoggia et al., 2019). Unlike previous studies, we used 5-fold cross-validation blocked by monitor to assess whether the downscaling improved accuracy: the downscaled predictions were less accurate than the 1 km GAM ensemble predictions (higher MAE, lower R^2). Downscaling over cities is an area for further research, as epidemiological studies would benefit from better estimates of differences in PM exposure within a city.

Despite good overall performance, our approach has some limitations. First, the sparsity of the monitoring network limited performance in early years. Even in later years, most monitors were clustered in cities, making it difficult to evaluate accuracy in smaller towns and rural areas and risking overreliance on predictors that work well in urban areas but may not work well elsewhere. The clustering of monitors near cities means our model is roughly weighted by population density, which may or may not be appropriate depending on the intended use for the predictions (Sarafian et al., 2020). New low-cost PM sensors might complement the existing monitoring network, particularly since our model's weaker spatial performance suggests that a few PM observations at new locations might be more useful than a long timeseries of measurements

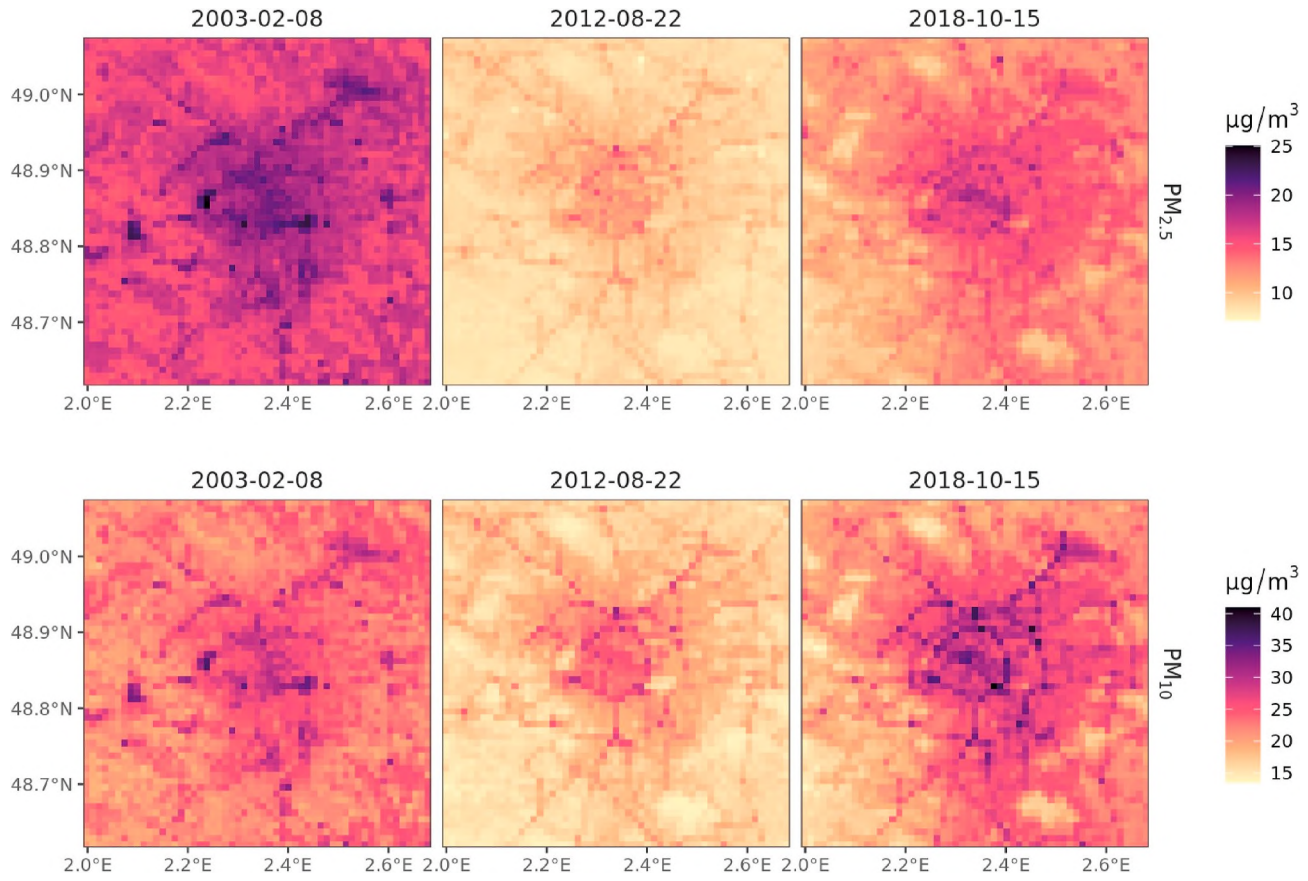


Fig. 7. Mean 24-h $PM_{2.5}$ (top) and PM_{10} (bottom) concentration over greater Paris predicted by the stage 4 GAM ensemble on three example days.

at a single location.

Second, MAIAC AOD was the only predictor with both a high spatial (1 km) and temporal (daily) resolution, but since it is based on a few daytime observations of the entire atmospheric column, it is both vertically and temporally misaligned with surface-level daily mean PM concentration. We included planetary boundary layer height to help distinguish between surface-level vs. high-altitude aerosols, but it had a much coarser spatial resolution than MAIAC AOD. Our model might have benefitted from AOD at coarser spatial but higher temporal resolution, such as from geostationary weather satellites, or from considering longer time periods and giving greater weight to rare observations when filling gaps in MAIAC AOD.

Despite these limitations, our multi-stage ensemble approach was able to predict daily 1 km $PM_{2.5}$ and PM_{10} with low error across a large area over 20 years. To the best of our knowledge, this is the first work conducted in France with such a high spatiotemporal resolution (1 km-daily), large spatial extent (national) and long temporal coverage (2000–2019). We increased accuracy by supplementing sparse $PM_{2.5}$ observations with imputed data and by ensembling the predictions of three base learners. We confirmed that Gaussian Markov random fields predict daily PM concentration better than linear mixed models and provide the first evidence that they may also outperform random forests. Our dataset of daily 1 km $PM_{2.5}$ and PM_{10} is available to health and ecosystems researchers in France and may inform policy makers on air quality issues.

CRediT authorship contribution statement

Ian Hough: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Visualization. **Ron Sarafian:** Methodology, Software, Writing – review & editing. **Alexandra Shtein:** Methodology, Software, Writing – review & editing. **Bin Zhou:** Methodology, Software, Writing – review & editing. **Johanna Lepeule:** Resources, Writing – review & editing, Supervision, Funding acquisition. **Itai Kloog:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition.

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.

Acknowledgements

Funding: This work was supported by the French National Centre for Scientific Research (ANR PRC 2018–2020), Fondation de France (no. 00081169), and Israeli Ministry of Science and Technology (PRC 2018–2020). I.H. is supported by Univ. Grenoble Alpes (ANR-15-IDEX-02) and Ben Gurion University of the Negev. Most data processing and analyses were performed using the GRICAD 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 PM monitor data, INERIS for providing a spatial emissions inventory, and NASA, ECMWF, Copernicus, IGN, and OpenTransportMap for their open datasets.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2021.118693.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F., 2018. Spatial modeling with R-INLA: a review. Wiley Interdiscip. Rev. Comput. Stat. 10, e1443. https://doi.org/10.1002/wics.1443.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 48. https://doi.org/10.18637/jss.v067.i01.

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

Beloconi, A., Kamarianakis, Y., Chrysoulakis, N., 2016. Estimating urban PM10 and PM2.5 concentrations, based on synergistic MERIS/AATSR aerosol observations, land cover and morphology data. Remote Sens. Environ. 172, 148–164. https://doi.org/10.1016/j.rse.2015.10.017.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z.M., 2016. Mlr: machine learning in R. J. Mach. Learn. Res. 17, 1–5.

Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., Lang, M., 2017. In: mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions. https://doi.org/10.13140/RG.2.2.11865.31849 arXiv.

Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv. Stat. Anal. 97, 109–131. https://doi.org/10.1007/s10182-012-0196-3.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A.H., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Vermeulen, R., Brunekreef, B., Hoek, G., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ. Int. 130 https://doi.org/10.1016/j.envint.2019.104934.

Chudnovsky, A.A., Koutrakis, P., Kloog, I., Melly, S.J., Nordio, F., Lyapustin, A., Wang, Y., Schwartz, J., 2014. Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. Atmos. Environ. 89, 189–198. https://doi.org/10.1016/j.atmosenv.2014.02.019.

Chudnovsky, A.A., Lee, H.J., Kostinski, A., Kotlov, T., Koutrakis, P., 2012. Prediction of daily fine particulate matter concentrations using aerosol optical depth retrievals from the Geostationary Operational Environmental Satellite (GOES). J. Air Waste Manag. Assoc. 62, 1022–1031. https://doi.org/10.1080/10962247.2012.695321.

de Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N., Kloog, I., 2018. Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. Environ. Pollut. 233, 1147–1154. https://doi.org/10.1016/j.envpol.2017.10.025.

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., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909. https://doi.org/10.1016/j.envint.2019.104909.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50, 4712–4721. https://doi.org/10.1021/acs.est.5b06121.

Didan, K., Barreto Munoz, A., Solano, R., Huete, A., 2015. MODIS Vegetation Index User's Guide (MOD13 Series).

Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Am. Stat. Assoc. 114, 445–452. https://doi.org/10.1080/01621459.2017.1415907.

Guo, Y., Tang, Q., Gong, D.Y., Zhang, Z., 2017. Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. Remote Sens. Environ. 198, 140–149. https://doi.org/10.1016/j.rse.2017.06.001.

He, Q., Huang, B., 2018. Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling. Remote Sens. Environ. 206, 72–83. https://doi.org/10.1016/j.rse.2017.12.018.

He, Q., Zhang, M., Song, Y., Huang, B., 2021. Spatiotemporal assessment of PM2.5 concentrations and exposure in China from 2013 to 2017 using satellite-derived data. J. Clean. Prod. 286, 124965. https://doi.org/10.1016/j.jclepro.2020.124965.

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.

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 https://doi.org/10.1016/j.envres.2020.109244.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating PM2.5 concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 1–29.

Hu, X., Waller, L.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Sarnat, J.A., Liu, Y., 2013. Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression. Environ. Res. 121, 1–10. https://doi.org/10.1016/j.envres.2012.11.003.

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., Liu, Y., 2014. Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sens. Environ. 140, 220–232. https://doi.org/10.1016/j.rse.2013.08.032.

Inness, A., Ades, M., Agustí-Panareda, A., Barr, J., Benedictow, A., Blechschmidt, A.M., Jose Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.H., Razinger, M., Remy, S., Schulz, M., Suttie, M., 2019. The CAMS reanalysis of atmospheric composition. Atmos. Chem. Phys. 19, 3515–3556. https://doi.org/10.5194/acp-19-3515-2019.

Insee, 2020. In: Estimation de la population au 1er janvier 2020 [WWW Document]. URL. https://www.insee.fr/fr/statistiques/1893198.

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. Atmos. Environ. 239, 117649. https://doi.org/10.1016/j.atmosenv.2020.117649.

Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Moody, E., Wang, Y., Lyapustin, A., Kloog, I., 2015. Using high-resolution satellite aerosol optical depth to estimate daily PM2.5 geographical distribution in Mexico city. Environ. Sci. Technol. 49, 8576–8584. https://doi.org/10.1021/acs.est.5b00859.

Kloog, I., Chudnovsky, A.A., Just, A.C., Nordio, F., Koutrakis, P., Coull, B.A., Lyapustin, A., Wang, Y., Schwartz, J., 2014. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. Atmos. Environ. 95, 581–590. https://doi.org/10.1016/j.atmosenv.2014.07.014.

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.

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B.A., Wang, Y., Just, A.C., Schwartz, J., Broday, D.M., 2015. Estimating daily PM2.5 and PM10 across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. Atmos. Environ. 122, 409–416. https://doi.org/10.1016/j.atmosenv.2015.10.004.

Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. Atmos. Chem. Phys. 11, 7991–8002. https://doi.org/10.5194/acp-11-7991-2011.

Lee, M., Kloog, I., Chudnovsky, A.A., Lyapustin, A., Wang, Y., Melly, S., Coull, B.A., Koutrakis, P., Schwartz, J., 2016. Spatiotemporal prediction of fine particulate matter using high-resolution satellite images in the Southeastern US 2003-2011. J. Expo. Sci. Environ. Epidemiol. 26, 377–384. https://doi.org/10.1038/jes.2015.41.

Liang, F., Xiao, Q., Wang, Y., Lyapustin, A., Li, G., Gu, D., Pan, X., Liu, Y., 2018. MAIAC-based long-term spatiotemporal trends of PM2.5 in Beijing, China. Sci. Total Environ. 616–617, 1589–1598. https://doi.org/10.1016/j.scitotenv.2017.10.155.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Liu, Q., Wu, R., Zhang, W., Li, W., Wang, S., 2020. The varying driving forces of PM2.5 concentrations in Chinese cities: insights from a geographically and temporally weighted regression model. Environ. Int. 145, 106168. https://doi.org/10.1016/j.envint.2020.106168.

Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS Collection 6 MAIAC algorithm. Atmos. Meas. Tech. 11, 5741–5765. https://doi.org/10.5194/amt-11-5741-2018.

Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM2.5 in China using satellite remote sensing. Environ. Sci. Technol. 48, 7436–7444.

Meng, X., Liu, C., Zhang, L., Wang, W., Stowell, J., Kan, H., Liu, Y., 2021. Estimating PM2.5 concentrations in Northeastern China with full spatiotemporal coverage, 2005–2016. Remote Sens. Environ. 253 (March 2020), 112203. https://doi.org/10.1016/j.rse.2020.112203.

Murray, C.J.L., et al., 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396, 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2.

Murray, N.L., Holmes, H.A., Liu, Y., Chang, H.H., 2019. A Bayesian ensemble approach to combine PM2.5 estimates from statistical models using satellite imagery and numerical model simulation. Environ. Res. 178, 108601. https://doi.org/10.1016/j.envres.2019.108601.

Nordio, F., Kloog, I., Coull, B.A., Chudnovsky, A.A., Grillo, P., Bertazzi, P.A., Baccarelli, A.A., Schwartz, J., 2013. Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy. Italy. Atmos. Environ. 74, 227–236. https://doi.org/10.1016/j.atmosenv.2013.03.043.

Park, Y., Kwon, B., Heo, J., Hu, X., Liu, Y., Moon, T., 2020. Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. Environ. Pollut. 256, 113395. https://doi.org/10.1016/j.envpol.2019.113395.

Pu, Q., Yoo, E.H., 2021. Ground PM2.5 prediction using imputed MAIAC AOD with uncertainty quantification. Environ. Pollut. 274, 116574. https://doi.org/10.1016/j.envpol.2021.116574.

R Core Team, 2020. R: A Language and Environment for Statistical Computing.

Randles, C.A., da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., Flynn, C.J., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: system description and data assimilation evaluation. J. Clim. 30, 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1.

Riviere, E., Bernard, J., Hulin, A., Virga, J., Dugay, F., Charles, M.A., Cheminat, M., Cortinovis, J., Ducroz, F., Laborie, A., Malherbe, L., Piga, D., Real, E., Robic, P.Y., Zaros, C., Seyve, E., Lepeule, J., 2019. Air pollution modeling and exposure assessment during pregnancy in the French Longitudinal Study of Children (ELFE). Atmos. Environ. 205, 103–114. https://doi.org/10.1016/j.atmosenv.2019.02.032.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Sarafian, R., Kloog, I., Just, A.C., Rosenblatt, J.D., 2019. Gaussian markov random fields versus linear mixed models for satellite-based PM2.5 assessment: evidence from the northeastern USA. Atmos. Environ. 205, 30–35. https://doi.org/10.1016/j.atmosenv.2019.02.025.

Sarafian, R., Kloog, I., Sarafian, E., Hough, I., Rosenblatt, J.D., 2020. A domain adaptation approach for performance estimation of spatial predictions. IEEE Trans. Geosci. Rem. Sens. 1–9. https://doi.org/10.1109/tgrs.2020.3012575.

Schneider, R., Vicedo-Cabrera, A.M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., Gasparrini, A., 2020. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across great britain. Rem. Sens. 12, 1–19. https://doi.org/10.3390/rs12223803.

Shtein, A., Karnieli, A., Katra, I., Raz, R., Levy, I., Lyapustin, A., Dorman, M., Broday, D. M., Kloog, I., 2018. Estimating daily and intra-daily PM10 and PM2.5 in Israel using a spatio-temporal hybrid modeling approach. Atmos. Environ. 191, 142–152. https://doi.org/10.1016/j.atmosenv.2018.08.002.

Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., Just, A.C., Stafoggia, M., 2019. 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.

Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: a principled, practical approach to constructing priors. Stat. Sci. 32, 1–28. https://doi.org/10.1214/16-STS576.

Song, W., Jia, H., Huang, J., Zhang, Y., 2014. A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. Remote Sens. Environ. 154, 1–7. https://doi.org/10.1016/j.rse.2014.08.008.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., Donato, F. De, Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J., 2019. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environ. Int. 124, 170–179. https://doi.org/10.1016/j.envint.2019.01.016.

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., Kloog, I., Davoli, M., Michelozzi, P., Bellander, T., 2020. A random forest approach to estimate daily particulate matter, nitrogen dioxide, and ozone at fine spatial resolution in Sweden. Atmosphere 11. https://doi.org/10.3390/atmos11030239.

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., De'Donato, F.K., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., Kloog, I., 2017. Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. Environ. Int. 99, 234–244. https://doi.org/10.1016/j.envint.2016.11.024.

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinf. 8 https://doi.org/10.1186/1471-2105-8-25.

Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., Winker, D.M., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 50, 3762–3772. https://doi.org/10.1021/acs.est.5b05833.

WHO, 2006. Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide (Geneva).

Wood, S.N., 2017. In: Generalized Additive Models: an Introduction with R, second ed. CRC Press.

Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Software 77. https://doi.org/10.18637/jss.v077.i01.

Xiao, Q., Geng, G, Liang, F., Wang, X., Lv, Z., Lei, Y., Huang, X., Zhang, Q., Liu, Y., He, K., 2020. Changes in spatial patterns of PM2.5 pollution in China 2000–2018: Impact of clean air policies. Environ. Int. 141, 105776. https://doi.org/10.1016/j.envint.2020.105776.

Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens. Environ. 199, 437–446. https://doi.org/10.1016/j.rse.2017.07.023.

Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y., 2015. Daily estimation of ground-level PM2.5 concentrations over beijing using 3 km resolution MODIS AOD. Environ. Sci. Technol. 49, 12280–12288. https://doi.org/10.1021/acs.est.5b01413.

Yan, X., Zang, Z., Luo, N., Jiang, Y., Li, Z., 2020. New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data. Environ. Int. 144, 106060. https://doi.org/10.1016/j.envint.2020.106060.

Zhai, B., Chen, J., 2018. Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China. Sci. Total Environ. 635, 644–658. https://doi.org/10.1016/j.scitotenv.2018.04.040.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012a. Real-time air quality forecasting, part I: history, techniques, and current status. Atmos. Environ. 60, 632–655. https://doi.org/10.1016/j.atmosenv.2012.06.031.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012b. Real-time air quality forecasting, Part II: state of the science, current research needs, and future prospects. Atmos. Environ. 60, 656–676. https://doi.org/10.1016/j.atmosenv.2012.02.041.

Zhang, Z., Wang, J., Hart, J.E., Laden, F., Zhao, C., Li, T., Zheng, P., Li, D., Ye, Z., Chen, K., 2018. National scale spatiotemporal land-use regression model for PM2.5, PM10 and NO2 concentration in China. Atmos. Environ. 192, 48–54. https://doi.org/10.1016/j.atmosenv.2018.08.046.

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmos. Environ. 124, 232–242. https://doi.org/10.1016/j.atmosenv.2015.06.046.