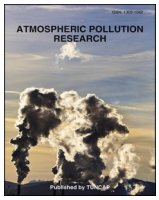
Atmospheric Pollution Research 14 (2023) 101763

Contents lists available at ScienceDirect

Atmospheric Pollution Research

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



Predicting fine-scale daily NO_2 over Mexico city using an ensemble modeling approach

Mike Z. He [a,*], Maayan Yitshak-Sade [a], Allan C. Just [a], Iván Gutiérrez-Avila [a], Michael Dorman [b], Kees de Hoogh [c,d], Bas Mijling [e], Robert O. Wright [a], Itai Kloog [a,b]

[a] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[b] Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel
[c] Swiss Tropical and Public Health Institute, Basel, Switzerland
[d] University of Basel, Basel, Switzerland
[e] Royal Netherlands Meteorological Institute, De Bilt, Netherlands

ARTICLE INFO

Keywords:
Air pollution
NO_2
Random forest
XGBoost
Ensemble modeling

ABSTRACT

In recent years, there has been growing interest in developing air pollution prediction models to reduce exposure measurement error in epidemiologic studies. However, efforts for localized, fine-scale prediction models have been predominantly focused in the United States and Europe. Furthermore, the availability of new satellite instruments such as the TROPOsopheric Monitoring Instrument (TROPOMI) provides novel opportunities for modeling efforts. We estimated daily ground-level nitrogen dioxide (NO_2) concentrations in the Mexico City Metropolitan Area at 1-km^2 grids from 2005 to 2019 using a four-stage approach. In stage 1 (imputation stage), we imputed missing satellite NO_2 column measurements from the Ozone Monitoring Instrument (OMI) and TROPOMI using the random forest (RF) approach. In stage 2 (calibration stage), we calibrated the association of column NO_2 to ground-level NO_2 using ground monitors and meteorological features using RF and extreme gradient boosting (XGBoost) models. In stage 3 (prediction stage), we predicted the stage 2 model over each 1-km^2 grid in our study area, then ensembled the results using a generalized additive model (GAM). In stage 4 (residual stage), we used XGBoost to model the local component at the 200-m^2 scale. The cross-validated R^2 of the RF and XGBoost models in stage 2 were 0.75 and 0.86 respectively, and 0.87 for the ensembled GAM. Cross-validated root-mean-squared error (RMSE) of the GAM was 3.95 μg/m^3. Using novel approaches and newly available remote sensing data, our multi-stage model presented high cross-validated fits and reconstructs fine-scale NO_2 estimates for further epidemiologic studies in Mexico City.

1. Introduction

Mexico City is the most populous city in North America, with a population of over 22 million in its metro area. Air pollution has been a persistent problem in Mexico City, in part due to the city's high elevation, high levels of urbanization, and rapid industrial growth (Bravo-Alvarez and Torres-Jardón, 2002). In 1992, the United Nations declared Mexico City as the most polluted city on Earth (Air Quality Life Index, 2019), prompting additional mitigation efforts that the government had begun back in 1990 with the Comprehensive Program Against Air Pollution (PICCA). Since then, the Mexican government introduced a series of comprehensive policies as part of the ProAire program, which aimed to address pollution from several fronts, including reducing energy consumption, improving public transportation infrastructure, and promoting environmental education (Centre for Public Impact, 2016). Currently in its fourth phase, ProAire has been successful in improving air quality in the Greater Mexico City area. However, because the city sits over 2000 m above sea level and is surrounded by mountains on three sides, atmospheric oxygen levels are naturally lower, which results in incomplete fuel combustion and increased concentrations of various primary and secondary pollutants, including fine particulate matter ($PM_{2.5}$), ozone, and nitrogen dioxide (NO_2).

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.
* Corresponding author. Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1057, New York, NY, USA, 10029.
E-mail address: mike.he@mssm.edu (M.Z. He).

https://doi.org/10.1016/j.apr.2023.101763
Received 31 October 2022; Received in revised form 14 April 2023; Accepted 15 April 2023
Available online 17 April 2023
1309-1042/© 2023 Turkish National Committee for Air Pollution Research and Control. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

NO_2 is an important traffic-related pollutant and has numerous roles in the formation of secondary pollutants in the environment. In the presence of sunlight and volatile organic compounds, NO_2 is a precursor chemical to ozone. NO_2 also plays a major role in the formation of haze and acid rain (U.S. Environmental Protection Agency (EPA), 2016). Because fossil fuel combustion directly contributes to the production of NO_2, NO_2 is often considered a surrogate of traffic-related air pollution (TRAP). The negative effects of NO_2 on human health are also well established. NO_2 can result in many harmful effects to the respiratory system, including reduced lung function and increased inflammation of airways (United States Environmental Protection Agency, 2015). Numerous epidemiologic studies have found associations between NO_2 and increased mortality, hospital admissions, and other subclinical outcomes (Burnett et al., 2004; He et al., 2020; Mills et al., 2015; Ward-Caviness et al., 2016).

Epidemiologic studies of air pollution and health have traditionally relied on the use of data from ground monitoring stations for the purposes of exposure assessment (Bell et al., 2008; Dockery et al., 1993; Peng et al., 2005). Because monitoring stations are somewhat sparse relative to the populations of interest, doing so introduces two potential problems. First, epidemiologic studies relying solely on monitoring stations for exposure are limited to studying populations with a specified distance to the monitors. Since monitors are generally placed in densely populated urban areas, this often results in rural populations being excluded from epidemiologic studies. Second, populations around a certain monitor would typically receive the same exposure assignment, which introduces potential error due to exposure misclassification.

In order to address the above concerns, there has been an increase in the use of air pollution prediction models in recent years. Current modeling efforts often utilize remote sensing data and/or employ machine learning-based techniques that result in high levels of predictive accuracy (Bi et al., 2018; Di et al., 2019; Just et al., 2020; Shtein et al., 2019; Stafoggia et al., 2019; Van Donkelaar et al., 2016). Although existing modeling efforts have predominantly focused on $PM_{2.5}$, there are also many prediction models for other pollutants, including NO_2 (De Hoogh et al., 2019; Di et al., 2020; Gilbert et al., 2005; Hoek et al., 2008; Lee and Koutrakis, 2014; Zhan et al., 2018). Early attempts for NO_2 modeling employed land-use regression models, which are typically more suitable for the modeling of long-term averages (Lee and Koutrakis, 2014). With an increased demand for short-term (e.g., daily) air pollution predictions for use in health studies, efforts have shifted towards the use of satellite-based models, since spatiotemporal modeling allows for short-term predictions. The statistical modeling approach for satellite-based models have also grown increasingly sophisticated over the years, from transitional regression models and linear mixed effects models (De Hoogh et al., 2019) towards machine learning-based methods, including the use of the random forest (RF) algorithm (De Hoogh et al., 2019; Zhan et al., 2018) or alternate algorithms, such as neural networks (Di et al., 2020). Relative to the United States and Europe, there is a critical need for fine-scale prediction models built in low- and middle-income countries, including for Mexico City.

The purpose of this study is to construct a fine-scale daily NO_2 model over Mexico City from 2005 to 2019. We build on the methodology introduced in de Hoogh et al. (De Hoogh et al., 2019), which utilizes NO_2 columns from OMI, and introduce additional novel modeling components, including the use of data from the new TROPOspheric Monitoring Instrument (TROPOMI) NO_2 product and ensemble models for higher predictive accuracy.

2. Data and methods

2.1. Study area

The study area includes Mexico City, previously known as the Federal District (Distrito Federal). It is the capital and largest city of Mexico. It is also Mexico's most densely populated city, spanning an area of 1494.3 km^2 (577.0 $miles^2$) with a population of over 9 million people split across 16 boroughs (INEGI, n.d.). Mexico City sits over 2000 m above sea level and is surrounded by mountains and volcanos nearby on three sides, some reach elevations of over 5000 m.

2.2. Monitoring data

Daily NO_2 monitoring data for the study period (January 1, 2005 to December 30, 2019) were downloaded from Mexico City's automatic air quality monitoring network (Red Automática de Monitoreo Atmosférico; RAMA), which contains 42 monitoring stations across the Mexico City metropolitan area (CDMX, 2022) . Details about the principles of operation and types of equipment used for the monitoring system can be found in monitoring site audit reports (EPA Systems, 2018).

2.3. Satellite data

Two satellite-based NO_2 products were used in this study: OMI and TROPOMI. Launched in 2004, OMI is an instrument on the Aura satellite and provides NO_2 column data at a $0.25^\circ \times 0.25^\circ$ resolution. For the purposes of this study, the variable "ColumnAmountNO2-TropCloudScreened" was extracted from the level-3 *OMNO2d* product, which provided NO_2 column measurements for the entire study period.

Launched in 2017, TROPOMI is an instrument on the Copernicus Sentinel-5 Precursor satellite, and provides NO_2 column measurements at finer spatiotemporal resolutions. Both OMI and TROPOMI fly in a sun-synchronous orbit, with daily overpass times around local solar noon. Processed level-3 TROPOMI products were obtained from the Royal Netherlands Meteorological Institute, which provided 2-km^2 measurements of NO_2 columns from May 2018 to December 2019 (van Geffen et al., 2022).

2.4. Modeled tropospheric column NO_2

We also obtained daily estimates of modeled tropospheric column NO_2 from the Copernicus Atmospheric Monitoring Service's (CAMS) global reanalysis, which provided estimates at a $0.75^\circ \times 0.75^\circ$ resolution. Because satellites cannot accurately measure NO_2 columns at times due to cloud coverage, the daily CAMS estimates, together with other variables, were used to impute missing satellite NO_2 column measurements whenever applicable.

2.5. Spatial and temporal predictors

A number of additional spatial and temporal predictors are also included in the modeling process, including longitude, latitude, 10 m U wind component (*wind_u*), 10 m V wind component (*wind_v*), and 2 m temperature (*temp_2m*) from the global reanalysis ERA5-Land hourly data ($0.1^\circ \times 0.1^\circ$); daily boundary layer height (*blh*), and total cloud cover (*cloud*) from the global reanalysis ERA5 hourly data on single levels ($0.25^\circ \times 0.25^\circ$); sum of length of major roads (*road_length*), road density (*road_density*), density of road intersections (*intersection_density*), and distance to closest major road (*closest_road*) derived from OpenStreetMap; and elevation (*ele*) data from the Shuttle Radar Topography Mission. Our list of selected predictors is initially based on the lists of predictors from similar NO_2 prediction models published in the literature. From there, we tested in the imputation and calibration stages which predictors meaningfully contributed to improved model performance (e.g., higher R^2, reduced RMSE, etc.), which were then kept for the final models. A detailed list of all spatial and temporal predictors and their data sources can be found in the Supplemental Material (Table S1).

2.6. Statistical methods

We estimated daily ground-level nitrogen dioxide (NO_2) concentrations using a four-stage modeling process. Prior to modeling, all

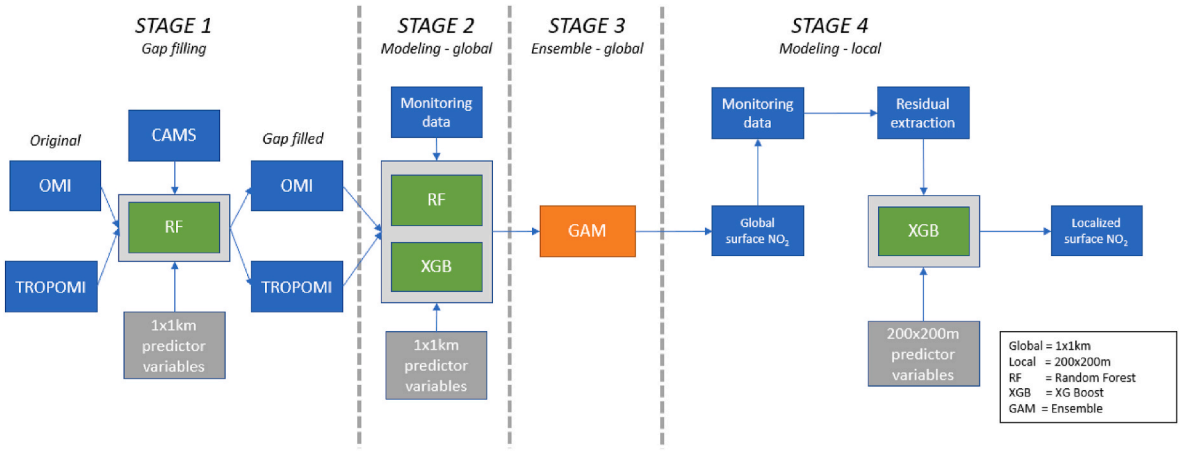


Fig. 1. Flow diagram of modeling approach.

geospatial data were resampled from their initial resolutions into standardized 1-km^2 grids using bilinear interpolation, which takes into account the centroid value based on the four closest cells to calculate each new grid cell.

In the first stage (imputation stage), we imputed missing OMI data using the CAMS ensemble NO_2 estimates along with other spatial and temporal variables (longitude, latitude, *blh*, *temp_2m*) using an RF model. This stage of the model performed well, with a 10-fold cross-validated R^2 of ~0.90. The filled-in OMI data along with additional spatial and temporal predictors are then used as inputs in the next stage.

In the second stage (calibration stage), we link the gap-filled OMI data to ground-level NO_2 from monitoring sites using RF and extreme gradient boosting (XGBoost), two machine-learning based algorithms. Both models included the same set of predictors, which includes gap-filled OMI data, *wind_u*, *wind_v*, *temp_2m*, *blh*, *cloud*, *elevation*, *road_length*, and Julian date. For the RF model, the following hyperparameters were considered: number of variables randomly sampled at each split (*mtry*), number of trees to grow (*num.trees*), and minimum size of terminal nodes (*min.node.size*). For the XGBoost model, the following hyperparameters were considered: maximum tree depths (*max_depth*), learning rate (*eta*), number of covariates sampled in each tree (*colsample_bytree*), and minimum leaf weight (*min_child_weight*).

In the third stage (prediction stage), we employ a generalized additive model (GAM) to geographically weight the cross-validated predictions from the two base learners (RF and XGBoost) in a highly flexible manner, which is then used to predict ground-level NO_2 from monitoring sites. Because the base learners each have its own advantages and disadvantages, geographically weighting the predictions (i.e., ensemble modeling) allows us to maximize the strengths from the individual models to yield the highest quality predictions. The model is based on the following relationship:

$$GAM(NO_2)_{ij} = s(X, Y, by = pred_{RF})_{ij} + s(X, Y, by = pred_{XGBoost})_{ij} + \varepsilon_{ij}$$

Where NO_{2ij} is the NO_2 concentration at monitor i on day j, X and Y are the geographic coordinates of monitor i, and $pred_{RF}$ and $pred_{XGBoost}$ are the cross-validated predictions from the RF and XGBoost models at monitor i on day j, and ε_{ij} is the error of the model. The GAM model is then used to predict ground-level NO_2 values across the entire study area over the study period. As a sensitivity analysis, we also run separate models from May 2018 to December 2020 predicting ground-level NO_2 by replacing OMI with TROPOMI NO_2 column data from the above stages to compare model performance.

In the fourth stage (residual stage), we defined another XGBoost model to explain the residuals between the modeled NO_2 in the prediction stage and other prediction variables at the 200-m^2 scale. In addition to spatial and temporal variables included in earlier stages (*wind_u*, *wind_v*, *temp_2m*, *blh*, *cloud*, *ele*), we considered a number of additional road metrics (*road_density*, *intersection_density*, and *closest_road*). Fig. 1 shows a flow diagram that summarizes the entire stepwise modeling approach.

Model performance from each stage was evaluated based on 10-fold cross-validation. We split the monitors into training and testing groups, and performance evaluation was completed only in the testing groups. Specifically, the monitoring database was divided into 10 random groups, and the cross-validation was completed in two separate substages.

(1) In the first substage, we create out-of-sample predictions for the two models in the calibration stage (RF and XGBoost). At each iteration, the two models were trained on 80% of the monitors to estimate NO_2 values for a separate test dataset that consisted of another 10% of monitors. A remaining 10% of monitors is kept as a validation dataset for evaluating the performance of the GAM model.
(2) At each iteration of this substage, the GAM model is trained using out-of-sample predictions of NO_2 from the first substage, which are then used to predict on the validation dataset that was kept aside during the first substage. This process ensures that the GAM model is trained on unbiased estimations (i.e., no data were used for both model fitting and predicting) and predicts on a dataset that was not used to train the other learners. The other learners are also trained using this same training dataset to predict the validation set. This process is reiterated 10 times until there are NO_2 estimates for all monitors. These estimates were then compared to actual measurements from the monitors.

Using the above procedure, we compute the following performance metrics.

- Coefficient of determination (R^2): the percent of explained variance between the observed and predicted NO_2 values
- 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
- Intercept: the y-intercept of the simple linear regression between observed and predicted NO_2
- Slope: the coefficient of the simple linear regression between observed and predicted NO_2

All statistical analyses were performed using the R Statistical Software, version 4.0.3 (Foundation for Statistical Computing, Vienna, Austria). The *ranger* package was used to fit all RF models (Wright and Ziegler, 2017), the *xgboost* package was used to fit all XGBoost models

Table 1
Cross-validated model performance across all stages.

Measure	RF (Stage 1 – Imputation)	RF (Stage 2 – Calibration)	XGBoost (Stage 2 – Calibration)	GAM (Stage 3 – Ensemble)	XGBoost (Stage 4 – Residual)
R^2	0.90	0.75	0.86	0.87	0.66
RMSE	0.00	5.43	3.98	3.95	1.67
Intercept	0.00	−1.62	−1.25	−0.28	−0.02
Slope	1.11	1.06	1.05	1.01	2.34

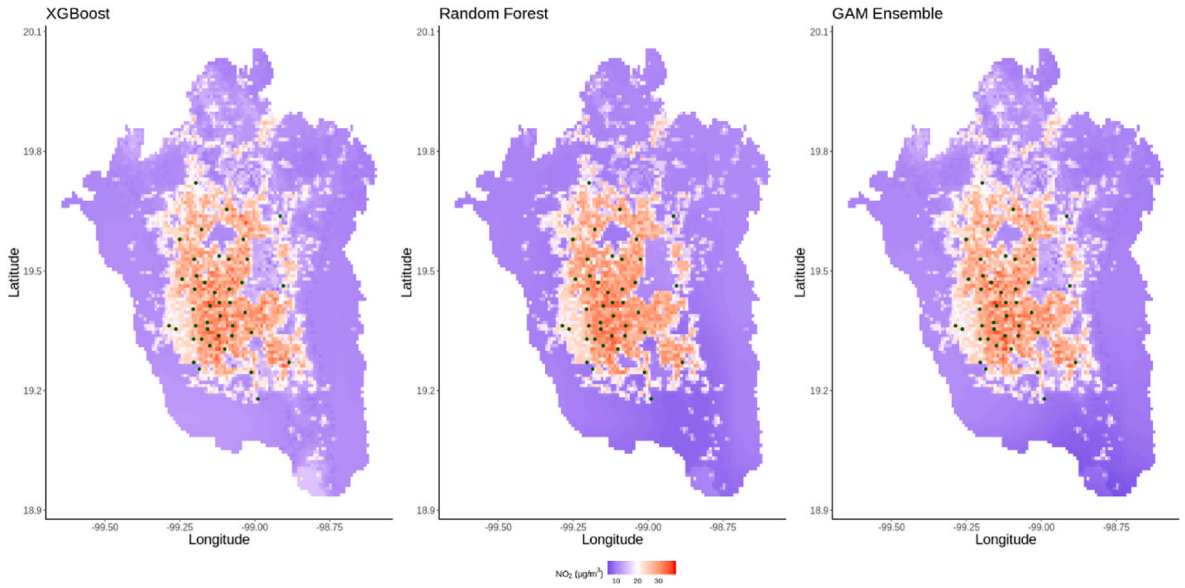


Fig. 2. Mean NO_2 estimates for all three models from 2005 to 2019: locations of monitoring stations are marked on each map.

(Chen and Guestrin, 2016), and the *caret* package was used to tune hyperparameters in the XGBoost models (Kuhn, 2008).

3. Results

A total of 42 unique NO_2 monitors operated across the study area during our study period. The mean daily measured NO_2 in our study area for the entire study duration is 27.1 μg/m^3. Measured NO_2 concentrations in the area decreased over the 15-year period, with a mean concentration of 32.3 μg/m^3 in 2005 to 21.5 μg/m^3 in 2019. A time series of ground-level NO_2 and detailed descriptive statistics of all NO_2 monitoring stations can be found in the Supplemental Material (Fig. S1; Table S2).

Performance metrics across all stages are shown in Table 1. On average, the RF, XGBoost, and GAM models in Stages 2 and 3 explained 75, 86, and 87% of variations in measured NO_2 concentrations. In our evaluation of variable importance, we found that *wind_u*, *road_length*, and satellite NO_2 were the top three contributors to model performance. While the RF and XGBoost models already present high cross-validated fits across the entire study period, the GAM further reduces the RMSE and slope of the model. The GAM is especially efficient at reducing the bias of the model, to a slope of 1.01 (down from 1.06 in the RF and 1.05 in XGBoost). For the residual stage, the XGBoost model explained on average 66% of variation in the residual NO_2 concentrations.

Fig. 2 shows maps of mean concentrations of NO_2 for the three different models across the entire study period, and Fig. 3 shows annual mean NO_2 concentrations of the GAM from 2005 to 2019. Predicted residuals had a mean of 0.05 and were normally distributed and centered around 0. Fig. 4 shows a histogram of predicted mean residuals (based on the residual stage model) across the entire study period. Additional performance metrics and spatial distributions of predictions can be found in the Supplemental Material (Table S3; Fig. S2; Fig. S3).

Performance metrics of the sensitivity analyses using TROPOMI data are shown in Table 2. To allow for a fair comparison, we reran models from our main analysis (using OMI) for the truncated time period for which TROPOMI data were available. Generally speaking, model performance using OMI and TROPOMI yielded similar results.

4. Discussion

This is the first high spatially and temporally resolved ensemble-based NO_2 exposure model built over Mexico City. Specifically, we developed models to estimate daily NO_2 concentrations at the 1-km^2 resolution from 2005 to 2019, and further estimated model residuals using a localized model at the 200-m^2 resolution. In our four-stage modeling process, we used an RF model to impute missing OMI NO_2 column data based on CAMS, established a relationship between ground-level NO_2 and OMI plus additional spatial and temporal predictors using RF and XGBoost models, applied the model to predict daily NO_2 across the entire study area using a GAM model, and finally estimated residuals using an XGBoost model, which can be used to estimate local NO_2 concentrations at the 200-m^2 scale. The model showed excellent performance, explaining on average up to 87% of variation in measured NO_2 concentrations and on average 66% of variation in residual NO_2 concentrations.

Spatially, we find the highest levels of NO_2 to be located in the center of the city, with levels decreasing substantially towards the outskirts of the city limits. This is an expected finding, as NO_2 is a traffic tracer and is highly dependent on traffic and road density. Temporally, outside of the expected seasonal trends within a given year, there is a clear decreasing trend of NO_2 over time. This can be seen in Fig. 3 as well, where mean predicted NO_2 levels for the entire study area decreased by 4.6 μg/m^3 from 2005 to 2019 (from 17.3 μg/m^3 to 12.7 μg/m^3).

To the best of our knowledge, this is also one of the first large-scale NO_2 prediction model to be utilizing TROPOMI data. Relative to OMI, the finer resolution of TROPOMI offered a promising alternative that

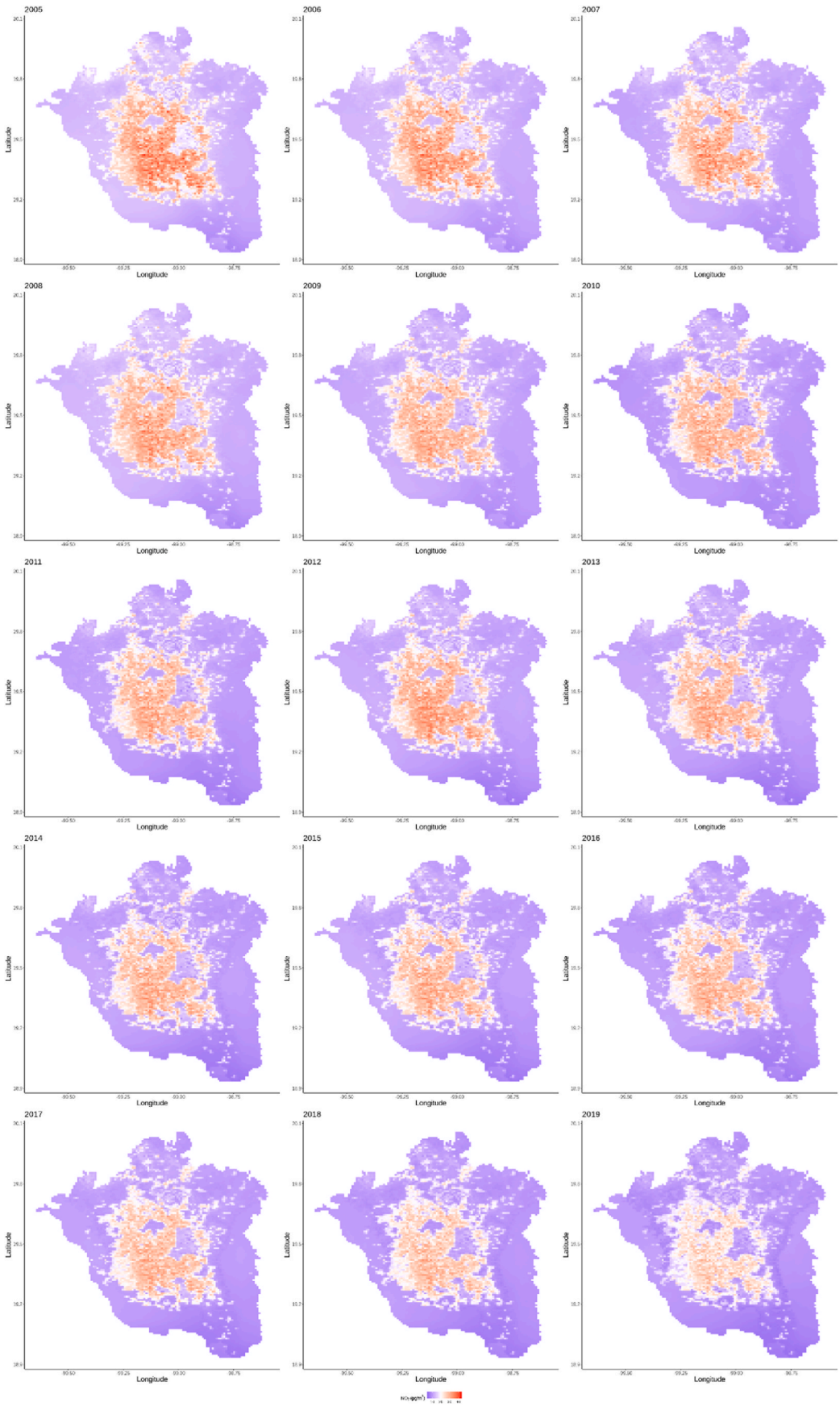


Fig. 3. Annual mean NO_2 estimates in 2005 and 2019 based on GAM.

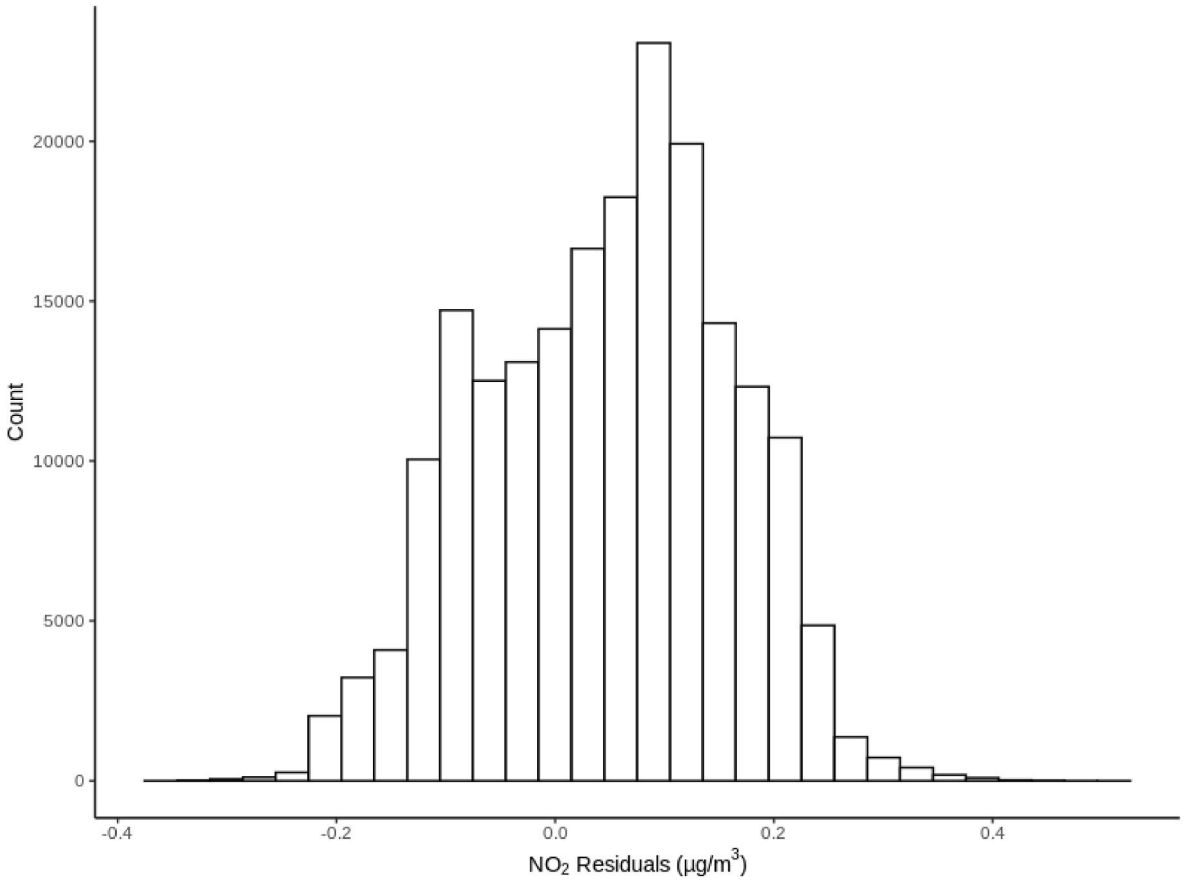


Fig. 4. Histogram of mean predicted NO_2 residuals based on XGBoost.

Table 2
Cross-validated model performance comparing OMI versus TROPOMI, May 2018 to December 2020.

Measure	RF_O[a]	RF_T[a]	XGBoost_O	XGBoost_T	GAM_O	GAM_T
R^2	0.79	0.78	0.88	0.88	0.88	0.88
RMSE	4.04	4.13	2.98	3.02	2.98	3.00
Intercept	−1.83	−1.43	−1.09	−0.97	−0.87	−0.77
Slope	1.08	1.06	1.05	1.04	1.04	1.03

[a] _O indicates OMI, _T indicates TROPOMI.

allows for the enhancement of existing NO_2 modeling capabilities. However, in our sensitivity analysis comparing models using OMI and TROPOMI for the same time period, early analysis of TROPOMI data seems to be on par with OMI data in terms of performance metrics, even though we expected a slight improvement. This may be due to the fact that we are modeling over a relatively small region for a short time period (~1.5 years). The benefits of TROPOMI data for NO_2 model predictions may be more pronounced as more data become available. Furthermore, the TROPOMI data that we used in our study comes from an early-stage algorithm. As with any early product, the algorithm and model are not fully optimized, and we expect the models to capture even more detailed spatiotemporal variations as we improve our understanding of the physical and atmospheric mechanisms that contribute to the model.

There are relatively few studies that have modeled daily NO_2 concentrations using satellite NO_2 products over a large spatial area. De Hoogh et al. built a model predicting fine-scale (1-km^2) daily NO_2 over Switzerland from 2005 to 2016 primarily using mixed effects and RF methods utilizing data from OMI. The models yielded robust predictions explaining ~58% of variations in measured NO_2 concentrations and ~73% of overall residuals (De Hoogh et al., 2019). Zhan et al. also utilized OMI data to construct a daily NO_2 model over China from 2013 to 2016. Their 0.1° × 0.1° model also showed good performance, with cross-validated R^2's of 0.62 and 0.73 for daily and spatial predictions respectively (Di et al., 2020). Di et al. built a 1-km^2 daily NO_2 model using multiple machine learning algorithms across the contiguous United States from 2000 to 2016. The model had a large number of predictor variables, including meteorological data, land-cover variables, and data from chemical transport models, and achieved a cross-validated R^2 of 0.79 (Zhan et al., 2018). More recently, Kim et al. predicted hourly NO_2 in Switzerland and northern Italy from 2018 to 2020 using XGBoost and found that TROPOMI observations were the most important predictors in the model, with the model capturing up to 59% of hourly NO_2 variations (Kim et al., 2021). Compared to existing previous studies, our study achieved a high level of predictive accuracy with cross-validated R^2's of up to 0.87, which is excellent compared to the current state of the art.

There are a few limitations of our study that should be considered. First, because our model relies primarily on NO_2 column data to generate predictions, we are only able to generate NO_2 predictions for which OMI data are available (2005 and on). This limitation in temporal coverage limits the use of our NO_2 predictions for health studies at earlier time points. Second, even though we are generating cross-validated and calibrated NO_2 predictions, the satellite NO_2 data we use come from the entire atmospheric column, and we are missing detailed vertical profiles from OMI and TROPOMI. Ideally, it would be better if we can use stratified NO_2 column data (e.g., column measurements every 10 or 20 m) to obtain even more accurate ground-level NO_2 predictions. Third, aside from the most central parts of the city, Mexico City has a serious lack of monitoring stations, which inhibits our ability to extend the model to the Greater Mexico City region and beyond.

Finally, similar to what we have seen with past modeling efforts for $PM_{2.5}$ and temperature, the transferability of our prediction models is not straightforward: generating predictions for other regions will require separate calibration processes that take into account the anthropogenic, climatic, and atmospheric conditions of the regions.

Our study also has a number of strengths. First, to the best of our knowledge, this is the first fine-scale, ensemble-based NO_2 exposure model built over Mexico City, which provides a valuable data source for those who are interested in conducting health studies of NO_2 in this densely populated metropolitan region. Second, our model demonstrated excellent performance, with R^2's of up to 0.87 in our main prediction model and R^2's of up to 0.66 in our residual model. Lastly, our model takes advantage of newly available TROPOMI NO_2 column data, which we anticipate will provide significant improvements to air pollution prediction models as more data becomes available.

Credit author statement

Mike Z. He: Methodology, Software, Validation, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Visualization, **Maayan Yitshak-Sade:** Writing – Review & Editing, Supervision, **Allan C. Just:** Resources, Writing – Review & Editing, **Iván Gutiérrez-Avila:** Data Curation, Writing – Review & Editing, **Michael Dorman:** Software, Resources, Writing – Review & Editing, **Kees de Hoogh:** Methodology, Writing – Review & Editing, **Bas Mijling:** Resources, Writing – Review & Editing, **Robert O. Wright:** Resources, Writing – Review & Editing, Supervision, Funding acquisition, **Itai Kloog:** Methodology, Software, Resources, Writing – Review & Editing, Supervision.

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

Research reported in this publication was supported by the National Institute of Environmental Health Sciences (NIEHS) under award number P30ES023515, and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) under award number T32HD049311.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apr.2023.101763.

References

Air Quality Life Index, 2019. Mexico City: ProAire (1990) [WWW Document]. URL. https://aqli.epic.uchicago.edu/policy-impacts/mexico-city-proaire-1990/. (Accessed 8 April 2022). accessed.

Bell, M.L., Ebisu, K., Peng, R.D., Walker, J., Samet, J.M., Zeger, S.L., Dominici, F., 2008. Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999-2005. Am. J. Epidemiol. 168, 1301–1310. https://doi.org/10.1093/aje/kwn252.

Bi, J., Belle, J.H., Wang, Y., Lyapustin, A.I., Wildani, A., Liu, Y., 2018. Impacts of snow and cloud covers on satellite-derived PM2.5 levels. Remote Sens. Environ. 221, 665–674. https://doi.org/10.1016/j.rse.2018.12.002.

Bravo-Alvarez, H., Torres-Jardón, R., 2002. Air Pollution Levels and Trends in the Mexico City Metropolitan Area, pp. 121–159. https://doi.org/10.1007/978-0-387-22520-3_6.

Burnett, R.T., Stieb, D., Brook, J.R., Cakmak, S., Dales, R., Raizenne, M., Vincent, R., Dann, T., 2004. Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. Arch Env. Heal. 59, 228–236.

CDMX, 2022. Dirección de Monitoreo Atmosférico [WWW Document]. URL. http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27. (Accessed 8 April 2022). accessed.

Centre for Public Impact, 2016. Mexico City's ProAire programme [WWW Document]. URL. https://www.centreforpublicimpact.org/case-study/mexico-citys-proaire-program. (Accessed 8 April 2022). accessed.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13–17, 785–794. https://doi.org/10.1145/2939672.2939785. -Augu.

De Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E.A., Strassmann, A., Puhan, M., Roösli, 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.

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.

Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., Speizer, F.E., 1993. An association between air pollution and mortality in six U.S. Cities. N. Engl. J. Med. 329, 1753–1759. https://doi.org/10.1056/NEJM199312093292401.

Epa Systems, L.L.C., 2018. Sistema de Monitoreo Atmosférico de la Ciudad de México Submitted to : Dirección General de Gestión de la Calidad del Aire Sistema de Monitoreo Atmosférico de la Ciudad de México Air Quality Monitoring Site Audit Report.

Gilbert, N.L., Goldberg, M.S., Beckerman, B., Brook, J.R., Jerrett, M., 2005. Assessing spatial variability of ambient nitrogen dioxide in montréal, Canada, with a land-use regression model. J. Air Waste Manag. Assoc. 55, 1059–1063. https://doi.org/10.1080/10473289.2005.10464708.

He, M.Z., Kinney, P.L., Li, T., Chen, C., Sun, Q., Ban, J., Wang, J., Liu, S., Goldsmith, J., Kioumourtzoglou, M.A., 2020. Short- and intermediate-term exposure to NO2 and mortality: a multi-county analysis in China. Environ. Pollut. 261, 114165 https://doi.org/10.1016/j.envpol.2020.114165.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2008.05.057.

INEGI, n.d. México en Cifras [WWW Document]. URL https://www.inegi.org.mx/app/areasgeograficas/default.aspx (accessed 8.April.2022).

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.

Kim, M., Brunner, D., Kuhlmann, G., 2021. Importance of satellite observations for high-resolution mapping of near-surface NO2 by machine learning. Remote Sens. Environ. 264, 112573 https://doi.org/10.1016/j.rse.2021.112573.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software 28, 1–26. https://doi.org/10.18637/jss.v028.i05.

Lee, H.J., Koutrakis, P., 2014. Daily ambient NO2 concentration predictions using satellite ozone monitoring instrument NO2 data and land use regression. Environ. Sci. Technol. 48, 2305–2311. https://doi.org/10.1021/es404845f.

Mills, I.C., Atkinson, R.W., Kang, S., Walton, H., Anderson, H.R., 2015. Quantitative systematic review of the associations between short-term exposure to nitrogen dioxide and mortality and hospital admissions. BMJ Open 5. https://doi.org/10.1136/bmjopen-2014-006946.

Peng, R.D., Dominici, F., Pastor-Barriuso, R., Zeger, S.L., Samet, J.M., 2005. Seasonal analyses of air pollution and mortality in 100 US cities. Am. J. Epidemiol. 161, 585–594. https://doi.org/10.1093/aje/kwi075.

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. https://doi.org/10.1021/acs.est.9b04279.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., 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.

U.S. Environmental Protection Agency (EPA), 2016. Basic Information about NO2 | Nitrogen Dioxide (NO2) Pollution [WWW Document]. URL. https://www.epa.gov/no2-pollution/basic-information-about-no2#What. (Accessed 6 January 2020). accessed.

United States Environmental Protection Agency, 2015. Integrated Science Assessment for Oxides of Nitrogen Health Criteria (Second External Review Draft), pp. 1–82.

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.

van Geffen, J.H.G., Eskes, H.K.F.B., Veefkind, J., 2022. TROPOMI ATBD of the Total and Tropospheric NO 2 Data Products.

Ward-Caviness, C.K., Nwanaji-Enwerem, J.C., Wolf, K., Wahl, S., Colicino, E., Trevisi, L., Kloog, I., Just, A.C., Vokonas, P., Cyrys, J., Gieger, C., Schwartz, J., Baccarelli, A.A., Schneider, A., Peters, A., 2016. Long-term exposure to air pollution is associated

with biological aging. Oncotarget 7, 74510–74525. https://doi.org/10.18632/oncotarget.12903.
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.
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.