Importance of satellite observations for high-resolution mapping of near-surface NO₂ by machine learning

Minsu Kim^{*}, Dominik Brunner, Gerrit Kuhlmann

Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland

ARTICLE INFO

Edited by: Menghua Wang

Keywords:

Nitrogen dioxides (NO₂)
TROPOspheric Monitoring Instrument (TROPOMI)
Sentinel-5 precursor satellite
High resolution map of NO₂
Machine learning
Extreme gradient boosting (XGBoost)
Shapley additive exPlanations (SHAP)
COVID-19 pandemic lockdown

ABSTRACT

Nitrogen dioxide (NO₂) is an important air pollutant with negative health effects and a precursor of ozone and particulate matter responsible for photo-chemical smog and wintertime air pollution. To evaluate human exposure to NO₂ for public health assessment, maps of near-surface NO₂ concentrations at a high resolution of 100 m are desirable. In this study, we report hourly maps of gridded near-surface NO₂ concentrations that are produced using an extreme gradient-boosted tree ensemble for an Alpine domain (Switzerland and northern Italy) spanning two years, from June 2018 to May 2020. To estimate the NO₂ distribution at ground level, we used satellite observations of NO₂ vertical column density, land use data, meteorological fields and topographical information to train models with *in situ* NO₂ ground measurements. The best model with this approach captured up to 59% of hourly NO₂ variation for 40 test stations in the domain with a mean absolute error of 7.69 μg/m³, performing especially well for urban regions with dense sampling. We present the first hourly maps of NO₂ concentrations that reveal previously unresolved spatio-temporal variations. Local interpretations of the machine learning model demonstrate that TROPOMI NO₂ satellite observations make a strong contribution to the information content of the near-surface NO₂ maps besides their relatively coarse resolution (3.5 × 5.5 km²) and the fact that they are only available once a day under cloud-free conditions. The COVID-19 pandemic lockdown presents a case study that offers new insights into the importance of satellite data that can partially re-mediate statistical models' unsusceptibility to unusual events (like changes due to political intervention) with regard to model training.

1. Introduction

Nitrogen oxides (NO_x = NO + NO₂) have a range of impacts on air quality either directly or as precursors of ozone (O₃) and particulate matter (PM) (Faustini et al., 2014; Beelen et al., 2014). The European Union (EU) and the World Health Organization (WHO) set the limits of ambient concentration of NO₂ at an annual mean of 40 μg/m³. This limit is frequently exceeded in many European countries and worldwide resulting in health impacts and damage to ecosystems (Brunekreef and Holgate, 2002; Bouwman et al., 2002; EEA, 2020a). Although the annual means of NO₂ in European countries are decreasing (EEA, 2020a), studies have reported that there is a lack of evidence for a safe annual exposure threshold regarding the association between air pollution and mortality even at cities with lower annual NO₂ concentrations (Achalwist et al., 2019; Khomenko et al., 2021). Besides, meta-analysis found positive relations between short-term exposure of NO₂ (1 h to 24 h) and mortality from all-cause and cause-specific diseases and with

cardiovascular and respiratory hospital admissions in different age groups (Mills et al., 2015) directing to the necessity of high resolution NO₂ maps that can capture the high spatial and temporal variability of NO₂ concentrations (Hanigan et al., 2017).

By using city-scale dispersion models, a European-wide study suggested that population weighted concentration maps with a resolution of 1 km, but preferably of 100 m, are required to avoid systematic population-biases in densely populated cities (Maiheu et al., 2017). Furthermore, spatial and temporal dynamics of personal exposure assessment (e.g. trajectory based approaches, individual mobility etc.) is only possible with sufficient spatial and temporal resolution (Jerrett et al., 2005; Dias and Tchepel, 2018). Several studies have generated maps of near-surface NO₂ concentration by combining *in situ* ground measurements with land-use regression models (Johnson et al., 2010; Wang et al., 2012; Mueller et al., 2015; Larkin et al., 2017). Others incorporate satellite observations and spatial kriging to produce maps at global, continental, and country scales (Xu et al., 2019; Chen et al.,

^{*} Corresponding author.

E-mail address: mins.kim@empa.ch (M. Kim).

<https://doi.org/10.1016/j.rse.2021.112573>

Received 11 February 2021; Received in revised form 21 June 2021; Accepted 23 June 2021

Available online 7 July 2021

0034-4257/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2019; Zhan et al., 2018; Young et al., 2016; Bechle et al., 2015; de Hoogh et al., 2019). The temporal resolution of these produced maps ranged from daily to yearly depending on their spatial coverage. These previous studies have shown the feasibility of bridging the spatial and temporal gaps between *in situ* measurements of near-surface NO₂ concentration

(temporally dense, spatially sparse) and satellite observations from space (temporally sparse, spatially dense). However, no previous studies have reported hourly maps of near-surface NO₂ concentrations at 100 m spatial resolution at regional scale.

Here, we employed tree-based ensemble models using the extreme

Fig. 1. Model performance for predicting hourly NO₂ concentration at 100 m resolution. (a) Map of the model domain showing the distribution of *in situ* stations used for training (blue circles) and for testing (red squares) and the topographic elevation. (b) Observation-prediction plot of near surface NO₂ concentration for hourly data from 40 test stations during the study period. The red colour scale represents the point density and the dashed line is the 1:1 relationship. Station-wise summary statistics are provided with (c,f) R^2 , (d,g) Spearman's rank correlation coefficient ρ_s , and (e,h) mean absolute error (MAE). (c-e) Time series of the original hourly data (H, blue) are aggregated daily (D, orange), weekly (W, green), and monthly (M, red). (f-h) Station wise data are aggregated for global human settlement modes; urban centers (UC, muted blue), urban clusters (Uc, muted orange), rural (R, muted green), and natural (N, muted red). Each box plot indicates the interquartile ranges and median. The black whiskers mark the 5th and 95th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

gradient boosting (XGBoost) algorithm (Chen and Guestrin, 2016) for predicting high-resolution hourly NO₂ maps for Switzerland and northern Italy ("Alpine domain"). Tree-based ensemble models offer an efficient approach to high resolution mapping, which is computationally intensive (e.g., about 20 million point predictions required in this study for the entire domain at a given hour). As covariates in the model, we selected NO₂ satellite observations, land use data (population density, road network, traffic volume, etc.), meteorological data (wind speed, boundary layer height, temperature, etc.) and topographic data (digital elevation map). The model was trained with *in situ* NO₂ observations and the trained model was used to analyze the contributions of satellite observations for high-resolution mapping. Finally, the predicted maps of near-surface NO₂ concentrations are underpinned by a case study of changes in NO₂ levels during the pandemic lockdowns in 2020.

2. Methods

2.1. Study region

In this work, we focused on an alpine domain that includes the Alps, the Swiss plateau and the Po Valley, representing terrain with partially complex topography. Due to the variety of landscape features, ranging from heavily industrialized and densely populated areas to rural alpine valleys, this study region is ideal to investigate interactions between natural (e.g. topographic and meteorological variables) and anthropogenic factors (e.g. land use and traffic data) that determine the NO₂ distribution. The study region is located within 42–48 °N and 6–12 °E (see Fig. 1 (a)).

2.2. Data preprocessing

Our model includes various predictors to capture the spatio-temporal variability of near-surface NO₂. Spatial predictors are land use, emission source proxies (population density, road length density, average traffic volume, industrial sources), and topographic information. These spatial maps are harmonized to the target grid (100 m × 100 m) that is the same as the land use grid. As spatio-temporal predictors, satellite observations of NO₂ and meteorological data (10 m wind speed, boundary layer heights, 2 m temperature, total precipitation, solar radiation) are also harmonized hourly at 100 m scale like other spatial predictors. As additional temporal predictors, hour of day, day of year, and weekdays are also included. A summary graphic of the model is shown in SI Fig. S1.

2.2.1. Ground level NO₂ observations

Hourly measurements of NO₂ concentration were collected from AirBase (Air Quality e-Reporting data repositories), which is maintained by the European Environmental Agency (EEA) (EEA, 2020b). Additional data for Switzerland were obtained from Switzerland's National Air Pollution Monitoring Network (NABEL) including non-AirBase monitoring sites managed by the Swiss cantons. In total, NO₂ time series data from 380 monitoring sites (315 Airbase stations and 65 Swiss cantonal stations) were compiled for this study. The locations of monitoring sites are shown in Fig. 1 (a) (see a list of stations used in this work in Supplementary material 1).

2.2.2. Satellite observations

Tropospheric NO₂ vertical column densities (VCDs in molecules/cm²) were taken from the TROPOMI NO₂ standard product (Version 01.03.02) from June 2018 to May 2020 (van Geffen et al., 2019). The Level-2 data were preprocessed by filtering quality assurance (f_{QA}) values accompanied with the data (Eskes et al., 2019). In this work, NO₂ VCDs with f_{QA} values <0.5 were filtered to remove erroneous data and problematic retrievals (Eskes et al., 2019). We did not use the stricter f_{QA} value (e.g. 0.75) recommended by the data product, which would also filter retrievals over clouds or snow/ice cover mainly because using stricter thresholding results in considerable loss of data (>40%) under

certain meteorological conditions (SI Fig. S2). The filtered satellite observations were projected to a longitude-latitude grid with 0.01° resolution (about 1 km) assuming constant values within the satellite pixels (Kuhlmann et al., 2014). After re-gridding, we gap-filled missing or filtered satellite observations by using normalized convolution with a Gaussian kernel for the spatial interpolation (Knutsson and Westin, 1993) which is ideal for gap-filling of small regions. The uncertainty of the inserted values were also derived by the algorithm as the cumulative distance from the certain values before the normalized convolution. The daily satellite observations and the uncertainty measures were linearly interpolated across days to obtain hourly maps and then included as covariates in the model. This simpler linear interpolation was chosen to fill temporal gaps without introducing additional information and to reduce a step-wise behaviour across hours.

2.2.3. Geographic covariates

This study includes six spatial predictors: (1) digital elevation map, (2) land use data, (3) population density, and (4) road length density. Specific measures that are directly related to sources of NO₂ were also considered such as (5) average traffic volume and (6) industrial emissions. The European Digital Elevation Model (EU-DEM) version 1.1 and the CORINE Land Cover (CLC 2018) from the Copernicus Land Monitoring Service were used at spatial resolutions of 25 m and 100 m, respectively (EU, EEA, 2016, 2018). Topographic (EU-DEM) data were downsampled to 100 m by averaging to match the resolution of categorical land use data. For population density, the global human settlement layer at 250 m resolution was re-gridded with the constant value method to match the resolution of land use data (Freire et al., 2016). The same dataset also provides a classification of human settlements based on degree of urbanization as urban centers (UC), urban clusters (Uc), rural (R), and regions without inhabitants (N, denoted as natural) (Dijkstra and Poelman, 2014), which we additionally used to classify the NO₂ monitoring stations. This classification was also used to aggregate results for regional comparisons. Road-length-weighted average of traffic volume and road-length density were calculated at 100 m resolution from the Open Transport Map (OTM), which provides publicly available and harmonized transport data with averaged traffic volume per road (Veckman et al., 2017). We also included NO_x emissions of point sources from the TNO/MACC-3 inventory, which provides the emission amount of various categories of industrial emission, such as power plants and other industries (Kuenen et al., 2014). Point source emissions were combined, harmonized, and rasterized at the same resolution as land use data (100 m) to be included in our model. Adding point sources is important as they are not sufficiently represented by other proxies such as land use data. For additional information on data sources see SI Table S1 and Table S2.

2.2.4. Spatial context of geographic covariates

Local NO₂ concentrations (in a given grid cell at 100 m resolution) depend not only on the local properties but also on the neighborhood, since NO₂ can be transported from neighboring emission sources. Geographic covariates directly related to NO₂ emissions like population density, road length density, traffic volume, and industrial emission were, therefore, decomposed into magnitude-related and distance-related data by using a decomposition scheme with Gaussian convolution kernels (see SI Text S1 and Fig. S3 for details).

A wavelet transform was used to decompose topographic information (digital elevation map) to multiple covariates (see SI Text S2 and Fig. S4 for details). This approach allows us to include surrounding topography of measurement stations at various length scales ranging from 100 m to 100 km in place of incorporating only their altitude. In this way we can account for the effect of terrain such as the trapping of air pollution in valleys and opposite effects on mountain ridges.

2.2.5. Meteorological data

While the geographic covariates presented above are static in this

study, meteorological data are included as spatial and temporal predictors. Hourly estimates of meteorological fields, including 2 m temperature, total precipitation, 10 m wind speed, solar radiation, and boundary layer height were obtained from the ERA5 reanalysis provided by the European Center for Medium-Range Weather Forecast (ECMWF). Specifically, we used publicly available ERA5 hourly data on single levels because it is available globally with high temporal resolution, despite its relatively coarse horizontal resolution of $0.25^\circ \times 0.25^\circ$ (approximately 25 km). We bilinearly interpolated this data to the center points of the 100 m grid to harmonize with other data (see details in SI Table S1 and Table S2).

2.2.6. Temporal covariates

In reality, emissions from traffic or industrial activity are highly variable in time. However, the publicly available spatial predictors used here were only available as static data. To compensate for this absence of emission dynamics, we included hour of day, weekdays, and day of year as temporal predictors. Weekdays are included in three categories, Monday to Friday (working days), Saturday, and Sunday.

2.3. Model selection and evaluation

2.3.1. Data-driven model (machine learning algorithm)

Using the preprocessed data as input, a decision tree regression model was trained with *in situ* NO₂ observations to predict near-surface NO₂ concentrations. Decision tree regression models recursively break down a dataset into subsets of multiple covariates in a binary manner and the final subset predicts output on the basis of conditional probability. Subsets can be sampled independently for each tree to make predictions and the predictive output can be aggregated by averaging outputs from multiple trees as an ensemble, which is called random forest (Breiman, 2001). Such tree-based regression models are non-parametric and have advantage over classical statistical models as they do not rely on any prior assumptions such as normality, independence, or homoscedasticity of the data. They are particularly suitable for handling large datasets with strong non-linearities and complex dependencies. Gradient boosting of regression trees, which optimizes an arbitrary differentiable loss function (e.g. minimizing errors) using a general gradient descent method for sequential learning, has been applied successfully to various problems (Friedman, 2001). In this work, we used the extreme gradient-boosting model (hereafter, XGBoost), which is a scalable end-to-end tree boosting system optimized for high performance computing with reduced training and prediction time (Chen and Guestrin, 2016). We chose this algorithm for computational efficiency to build near surface NO₂ concentration maps at hourly and 100 m resolution. Unlike previous studies that used XGBoost together with other machine learning algorithms in ensembles (Xiao et al., 2018; Li, 2019), we use solely an XGBoosted tree ensemble to maintain interpretability of the outcome. The hyper-parameters of the XGBoost model were selected by an automated algorithm (Le et al., 2020; Akiba et al., 2019) (SI Table S3). In this work, hyper-parameter tuning was only performed for a trial to guide the best performance but we did not pursue intense parameter tuning to avoid the risk of tuning the parameters to the specific training data.

2.3.2. Cross-validation and model evaluation

We used stratified sampling for cross-validation, which is referred to 'clustered cross-validation' to select test stations from clustered groups (Wang et al., 2012; Young et al., 2016). This approach was chosen to avoid over-fitting to urban sites which are dominating the measurement network and to provide similar weight to regions that are under-represented in the network (Babyak, 2004). To cluster AirBase stations into representative groups, we used dimensionality reduction by uniform manifold approximation (UMAP) (McInnes et al., 2018) and the hierarchical density based clustering method (HDBSCAN) (McInnes and Healy, 2017) on the reduced dimensional space. For this clustering

procedure, all geographic covariates were used together with annual means of meteorological data of each station (see Supplementary material 2 for a summary of values used for clustering). As a result we aim to represent each cluster with the combined characteristics of geographical and meteorological factors that are expected to influence near-surface NO₂ levels. After clustering, we randomly assigned 10% of stations in each cluster to a test set (a holdout set) for model evaluation. This clustering/categorization method has a merit of being solely determined by data and being observation-independent without subjectivity. However, clustering and selection of a test set includes stochasticity, hence we trained multiple models to evaluate our approach. We performed 10 clustering approaches and randomly select 5 sets of test sets from each clusters resulting 50 models with 50 test sets (see Supplementary material 3 for the clustered stations). For performance evaluation, we used classical R^2 , Spearman's rank correlation coefficient (ρ_s), and mean absolute error (MAE).

2.3.3. Interpretation of models using SHAP

We analyzed the best trained model by using the Shapley Additive exPlanations (SHAP) method (Lundberg and Lee, 2017; Lundberg et al., 2018; Rodríguez-Pérez and Bajorath, 2020). This method takes a game-theoretical approach based on the Shapley value (Shapley, 1953), which estimates the impact of each predictor for determining output in terms of magnitude and direction. The impact of a covariate is calculated as the difference in the model's predictions with and without the covariate in every possible permutations of all covariates. The impact is large if the differences are higher in model predictions when the covariate was omitted in the model. The SHAP values are calculated per incident (per case), thus the magnitude of SHAP values indicates the local importance of the predictor and the direction indicates reduction or accretion of the output relative to the global average as a base value. This method is beneficial to access the local interpretation of a complex model under various conditions (Lundberg et al., 2020). By interpreting the model with SHAP, we show the individual contributions of predictors with a particular focus on the evaluation of the contributions of the TROPOMI observations.

3. Results and discussions

3.1. Model performance

For June 2018 to May 2020, NO₂ station data and all covariates were compiled to build the data-driven model. After inserting missing predictor values, a total of 5,214,561 hourly observations from 380 stations were used in this study. Notably, about 5% of TROPOMI values were gap-filled in space and about 95.8% were linearly interpolated in time between daily observations. Stations were clustered into several groups, then 10% of each group were randomly selected as test stations (resulting in total 40 stations) to assess the model performance. The remaining stations were used to train tree ensembles using XGboost. Owing to the stochasticity in the clustering algorithm and the selection of the test set, we built 50 different tree ensemble models. In general, the trained models were not over-fitted, which can be concluded from the fact that increasing the number of stations used for training resulted in further performance gains (See SI Fig. S5). We found, however, that the selection of training/test stations has a strong impact on model performance even despite the 10-fold cross validation while training. This indicates that the clustering of stations did not completely eliminate sampling bias. About 72% of the stations in the network are located in urban environments according to the global human settlement classification (137 and 139 stations for urban centers and urban clusters, respectively) while there are only 15 stations (about 4%) in natural environments. Such a sampling bias is unavoidable due to the design of air quality monitoring networks, which primarily target highly polluted and densely populated areas where higher sampling is necessary owing to the high spatial variability in NO₂ concentrations. As a result, any

data-driven approach using this monitoring network is inherently tailored to the prediction of near-surface NO_2 in rather polluted regions where the majority of population resides. This can be seen as a disadvantage for spatially representative prediction but as an advantage for exposure assessment studies. Nevertheless, the clustered cross-validation allowed us to improve performance for under-represented regions of the measurement network.

We selected the best performing model in terms of goodness of prediction for a test set. The best model using the sets of training/test stations (Fig. 1 (a)) marked $R^2 = 0.59$, $\rho_s = 0.78$ and $\text{MAE} = 7.69 \mu\text{g}/\text{m}^3$ for the hourly data at the 40 test stations (Fig. 1 (b)). For observations from all stations including those used for training, the best performing model achieves $R^2 = 0.84$, $\rho_s = 0.92$ and $\text{MAE} = 4.74 \mu\text{g}/\text{m}^3$. Station-wise summary statistics improve when temporally aggregating the hourly time series of test stations to daily, weekly, and monthly values (Fig. 1 (c-e)). We found that some test stations exhibit very low performance (Fig. 1 (f-h)) which also could be attributed to the sampling bias towards urban regions with high concentrations caused by the design of this monitoring network. However, these problematic stations are located in regions with generally low NO_2 and with limited variation. Thus, the model prediction for these stations is poor in terms of R^2 values, but good enough in terms of MAE, because NO_2 concentrations at these stations are low. Some other stations with bad statistics were located at the boundary of our domain that were underestimated in terms of emission proxies.

3.2. High resolution map of near-surface NO_2

We used the model with the best performance based on the full dataset presented in Fig. 1 for analysis and for extrapolation to maps of near-surface NO_2 concentration at 100 m and hourly resolutions (a subset of produced maps during March 2019 is provided in Supplementary Video 1). Fig. 2 (a-b) shows a comparison of monthly mean tropospheric NO_2 vertical column density from the TROPOMI satellite instrument with a map of near-surface NO_2 concentration predicted for the Alpine domain at high resolution for March 2019. Enlarged urban regions of Zurich and Milan reveal that the model predicts fine-structures in near-surface NO_2 concentrations that cannot be captured by the coarser TROPOMI observations (Fig. 2 (c)). Comparisons of hourly time series (Fig. 2 (d)) illustrate that the model captures diurnal and day-to-day variability very well, although it tends to underestimate high peak (as well as very low) concentrations owing to the regression approach used in this model, which is optimized for predicting the mean behaviour. In general, while TROPOMI can only see ‘hotspots’ of NO_2 at the scale of entire cities, the near-surface concentration maps show much more detailed structures with high spatial and temporal variability along roads and valleys. The annual cycles of NO_2 distribution in this Alpine domain were presented in SI Fig. S6 with monthly averaged maps from February 2019 to May 2020. Our maps provide the first predicted spatial distribution of hourly near-surface NO_2 deduced from these observation networks. The result produced in this work is freely available at a data repository (Kim et al., 2021).

Fig. 2. (a) Average tropospheric NO_2 vertical column density observed by the satellite (TROPOMI) during March 2019. (b) Predicted map of average near-surface NO_2 concentrations at 100 m resolution during March 2019. (c) The magnified maps at city scales are given for Zurich and Milan indicated by the black squares in (a) and (b). Circles indicate monthly averaged values from in situ observations during March 2019. The colour scales are the same as used in (a) and (b) for TROPOMI VCD of NO_2 and model prediction of near-surface NO_2 , respectively. (d) Time series of near-surface NO_2 concentrations at test stations in Zurich (Airbase EoI station code, CH0058A) and Milan (IT0477A) during March 2019. Orange bars indicate the TROPOMI NO_2 observations at overpass time after spatial gap-filling. The locations of the stations are marked as black arrows in (c). Model predictions (red) are compared with hourly observations (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Daily cycles of near-surface NO₂

The model predictions exhibit typical daily cycles with peaks of NO₂ during morning and evening rush hours. Fig. 3 shows predicted maps of near-surface NO₂ at midnight (12 AM), morning rush hour (7 AM), and at midday (12 PM) and their differences to the monthly mean (shown in Fig. 2 (b)) during March 2019. We stress that diurnal patterns cannot be captured by the satellite observations (Fig. 2 (a)) since TROPOMI has a local overpass time during midday (13:30 local time) only, but it is present in the *in situ* observations used for training the model. Temporally varying covariates like hour of day and meteorological parameters allow the model to capture this variability.

The temporal difference to the monthly mean exhibits high spatial variability depending on the time of the day. We analyzed how these differences depend on location with mean daily cycles for different regions as defined by the global human settlement layer: urban centers, urban clusters, rural, and regions without inhabitants (denoted as natural). As expected, the averaged level of near-surface NO₂ depends on the degree of urbanization. Urban centers show the highest monthly

mean concentrations (about 25.3 $\mu\text{g}/\text{m}^3$) and the largest deviations from the mean during rush hours (up to 20 $\mu\text{g}/\text{m}^3$) (Fig. 3). Rural/natural regions remain at much lower levels and the diurnal variability is reduced. The diurnal cycle with morning and evening peaks is compliant with our general understanding that traffic is the dominant source of NO₂ especially for urban regions. The level of NO₂ during midday is lower than during night time, which suggests that other factors in addition to traffic volume, such as the depth of the boundary layer, play an important role as well.

3.4. Contribution of covariates to high resolution mapping of NO₂

Although air pollutant observations from satellites have been used in previous studies aiming at high-resolution mapping of NO₂, assessing the contribution of satellites to the final product was resorted to simple comparisons of model performance using summary statistics (Vienneau et al., 2013; Bechle et al., 2015; Young et al., 2016; Xu et al., 2019). These studies showed an increase in R^2 values by about 0.02–0.15 when including satellite-based NO₂ observations as a covariate for high-

Fig. 3. Distributions of near-surface NO₂ at (a) midnight 12 AM (UTC + 01:00), (b) 7 AM (UTC + 01:00) during morning rush hour, and (c) midday 12 PM (UTC + 01:00). Difference to the monthly average for (d) 12 AM (UTC + 01:00), (e) 7 AM (UTC + 01:00) and (f) 12 PM (UTC + 01:00). (g) Monthly mean daily cycles of NO₂ concentrations spatially averaged depending on degree of urbanization. Four classes are shown: urban centers (red), urban clusters (orange), rural (cyan), and regions without human settlement, marked as natural (navy). Error bars indicate 1 standard deviation from spatial averaging. The dashed lines indicate the monthly mean for each region, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resolution maps of monthly or annual mean NO₂ concentrations. Our study also agrees with this result showing the decrease in R^2 by about 0.04 when TROPOMI input was removed from our model (SI Fig. S5). In this work, we took a further approach to quantitatively assess satellite contributions by using the SHAP method in place of comparing models with different number of covariates. SHAP values represent the localized linear relation between covariates and model output as a local impact ('force') of the covariate for determining the prediction relative to the base value. In other words, a SHAP value in our model quantifies the contribution of a covariate to a predicted NO₂ value at a given time and location.

In Fig. 4, we present the SHAP values in terms of average, local, and relative contributions of each covariate. The average contribution (across all stations for two years of observations) was quantified in terms of mean absolute SHAP values (considering only magnitude). TROPOMI

NO₂ observations show the highest average contribution despite the fact that they are only available once a day (around midday) at most and linearly interpolated across observations. We note that the covariate of satellite uncertainty had a negligible impact in the model, indicating that the gap-filled and linear-interpolated values were admissible for prediction. With regard to the average impacts, the TROPOMI input was followed by hour of day, traffic volume weight, boundary layer height and day of year. This indicates that these five covariates played decisive roles in the prediction of the high-resolution hourly NO₂ maps. We further note that, in terms of total information gain of the tree model as an alternative measure of global importance, these covariates also ranked among the top five with slightly different order (see SI Fig. S7). TROPOMI NO₂ observations ranked the highest also in this metric representing almost 20% of relative information gain to determine the final output.

Fig. 4. Contribution of each covariate to near-surface NO₂ concentrations quantified with the Shapley Additive explanations (SHAP) method. (a) Average impact as quantified by the mean of absolute SHAP values. The 15 highest ranked covariates are listed in order from top to bottom. (b) Distribution of SHAP values ordered in the same way as the average impact. The colors ranging from blue to red represent the normalized values of each covariate from low to high, respectively. (c) Relative impact of each covariate grouped in categories of satellite input (red), traffic/industry data (gray), population/road length data (purple), land use data (light gray), meteorological data (blue) topographic data (green), and time information (orange). Specific values of all covariates are given in percentage together with legends. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The local impacts summarized in Fig. 4 (b) show how the values of each covariate influence near-surface NO_2 levels for each observation/prediction. For instance, generally positive SHAP values for high values of TROPOMI observation, traffic volume, and population density indicate their positive effects on near-surface NO_2 . In contrast, negative SHAP values associated with high values of elevation, boundary layer heights, and 10 m wind speed indicate a negative effect of these parameters on near-surface NO_2 . However, the wide ranges of SHAP values, spanning from negative to positive, point to their dynamic impact on local predictions and possible interactions with other covariates.

Fig. 4 (c) summarizes the normalized average impact of each covariate grouped by category: satellite observations, traffic/industrial data, population/road length data, land use, meteorological, topographic, and time. These relative contributions show a balanced input of information to the model and demonstrate the necessity of each category for high-resolution mapping of NO_2 . The overall contribution of static covariates representing the spatial distribution of emission sources

(population density, traffic, industrial emissions) is rather small (about 25%), which can be explained by the fact that hourly NO_2 concentrations are also determined by temporal variability. Indeed, the time category has a comparable contribution of 26.3% and complements the static data of emission sources with a representation of the temporal dynamics of these sources, especially traffic.

3.5. Dynamic roles of covariates to determine spatio-temporal distribution of near-surface NO_2

The power of the SHAP method lies in the possibility to investigate the temporal and spatial dynamics of the contribution of the covariates. As an example, we analyzed these dynamics for the area of Milan during March 2019. We found that roles of the covariates vary considerably at different spatial and temporal scales (Fig. 5). As shown in panel (a), both TROPOMI and boundary layer height make a strong contribution to the temporal evolution of NO_2 over the domain, with TROPOMI largely modulating the day-to-day variability and boundary layer heights the

Fig. 5. Dynamic contribution of covariates to the prediction of near-surface NO_2 in terms of temporal and spatial SHAP values for the area of Milan for March 2019. (a) Hourly SHAP values averaged across all 15 measurement stations in Milan. Two covariates with high impact, TROPOMI observation (red) and boundary layer height (blue) are given in the upper panel together with observed (cyan) and predicted (magenta) near-surface NO_2 in the lower panel. (b) Mean weekly patterns and (c) daily patterns of SHAP values derived from the hourly values presented in (a). The four most important covariates, TROPOMI observations (red), hour (orange), boundary layer height (blue), and weekday (black) are plotted. Lines indicate median values and shaded colors indicate interquartile ranges (IQR) across the 15 stations. (d) Spatial patterns of SHAP values in Milan at 7 AM (UTC + 1) and 12 PM (UTC + 1) averaged for March 2019. From top to bottom, TROPOMI observations, hour of day, and traffic volume weight, and boundary layer height are presented with the same colour scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

diurnal pattern. Boundary layer height is well known to affect the diurnal variability of NO_2 through its influence on vertical mixing (van Stratum et al., 2012). Further insights into the role of individual covariates in controlling the weekly and daily patterns are presented in panels (b) and (c). The weekday factor takes account of the lower emissions during weekends, especially on Sundays. The diurnal

variability is not only controlled by the boundary layer height but also by the 'hour' covariate, which strongly resembles the evolution of traffic emissions with a morning and an evening peak and lower activity during night (see Supplement material 4). The statistical model thus seems to be able to disentangle these two factors, which are in opposite phase and partly counterbalance each other. Despite being available only once a

Fig. 6. Changes in near-surface NO_2 concentrations due to COVID-19 pandemic lockdown events. (a, b) Difference in time-averaged NO_2 distributions in spring between 2019 and 2020 at 1 PM (TROPOMI overpass time) and 7 PM (evening rush hour). Here, we averaged from 9th of March to 18th of May corresponding to the period of the spring lockdown in northern Italy. (c) Six panels compare changes in average daily cycles of NO_2 concentrations between 2019 (dashed black lines) and 2020 (solid red lines) during March, April, and May for urban centers (top) and rural regions (bottom). Lines indicate median values and shaded colors indicate interquartile ranges. (d) Monthly averaged difference in near-surface NO_2 at city scale for Zurich and Milan (for a domain of $10 \text{ km} \times 10 \text{ km}$) during March, April, and May. Here, the colour ranges are the same as in panels (a) and (b) on the left. (e) Probability density functions (PDF) of measured and predicted near surface NO_2 in 2019 and 2020. (f) Changes in PDFs of SHAP values of selected covariates between 2019 (red) and 2020 (black). From PDFs, expected SHAP values ($E[X]_{\text{year}}$) are also given by years together with the yearly difference in its impact ($\Delta E[X]$ in blue texts). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

day at most, TROPOMI observations sustain relatively high SHAP values throughout the day and throughout the whole month, shifting the mean prediction from the global average. The spatial distributions of SHAP in Fig. 5 (d) demonstrate the role of static emission proxies in terms of mapping at high spatial resolution. The detailed structure of SHAP for traffic volume weight indicate that the input of emission source information was crucial to provide the fine details of the NO₂ distribution.

3.6. Case study: effect of pandemic lockdowns on near-surface NO₂

The study period in this work includes lockdowns related to the COVID-19 pandemic implemented by Italian and Swiss authorities. These unusual lockdown events had widespread impacts on atmospheric compositions, notably NO₂, owing to the great changes in human activities (Bauwens et al., 2020; Prunet et al., 2020; Grange et al., 2020). To illustrate the effect of these lockdowns, we analyzed the changes of NO₂ between the pre-pandemic period (spring, 2019) and during/after the lockdown (spring, 2020). The official duration of the spring lockdown in Italy was from 9th of March to 18th of May 2020 with the most strict regulations lasting from 9th of March - 14th April. The lockdown period in Switzerland was shorter, from 16th of March to 11th of May 2020. Fig. 6 (a) and (b) show the difference of time averaged near-surface NO₂ at 1 PM (UTC + 01:00) (around TROPOMI overpass time) and 7 PM (UTC + 01:00) (evening rush hour) between the two years, averaged from 9th of March to 18th of May. The reductions in NO₂ were considerable for big cities in the Po Valley, especially for Milan, exceeding 20 µg/m³ at some locations during evening (7 PM). The monthly averaged daily cycles for different regions during the lockdown period (Fig. 6(c)) show that, especially during March, the reduction occurred mainly in the morning/evening rush hours similar to observations in the United States (Tanzer-Gruener et al., 2020). During April and May, the reduction seems to persist through out the average day, this can be attributed to more stricter lockdown during this time in both countries, Switzerland and Italy. However, this interpretation should be taken carefully due to the differences in meteorology between 2019 and 2020 (Goldberg et al., 2020). The reductions were smallest during midday (TROPOMI overpass time) suggesting that the TROPOMI satellite observations cannot capture the full magnitude of the changes occurring mainly during the rush hours. By combining with the spatially sparse ground observations of near-surface NO₂, the hourly maps of our model, on the other hand, provide information on the temporal and spatial change of NO₂ concentrations during the lockdown. For rural regions, the difference of monthly averaged near-surface NO₂ concentrations was not as evident as for urban centers although there was a small decrease during evening hours in March and throughout the day during April.

Difference maps of near-surface NO₂ at city scale (~10 km) are displayed in Fig. 6 (d) for Zurich and Milan as examples. Compared to the mega city Milan (population 1.3 M), the reduction of NO₂ in Zurich (population 0.4 M) was less pronounced although the changes were still evident along the main roads within the city, especially during April. The differences in March were quite small since the Swiss lockdown was only introduced in the second half of the month.

Although the changes in human activities must be the main cause of this reduction, effects of meteorological variability need to be considered as well when comparing NO₂ concentrations between 2019 and 2020 (Goldberg et al., 2020). The SHAP method does not only allow us to investigate the spatio-temporal evolution of NO₂ during the lockdowns but also to separate meteorological effects from changes in emissions (Fig. 6 (e, f)). Relative changes in probability distribution functions of SHAP values between 2019 and 2020 (9th of March - 18th of May) show that the predicted reduction was mainly driven by day of year, and by traffic volume weight with the leading contribution for reduction by TROPOMI observations (Fig. 6(f)). This indicates that the model can successfully respond to unusual events if the events are included in the training. The model detected this unusual event from the

in situ observations and directed the output towards lower NO₂ by adjusting the impact of these covariates, especially 'day of year', which captured the abrupt lockdowns together with TROPOMI observations. In our model, the impact of boundary layer height was similar between 2019 and 2020. Although the changes were rather small compared to other variables, 2 m temperature also exhibited its role in reducing near surface NO₂ owing to higher temperatures in spring 2020 compared to 2019. This could possibly be caused by reduced emissions from residential heating due to the decrease in heating degree days (see SI Fig. S8), which can be used as a proxy for estimating emissions from residential heating (Zhang and Zhou, 2019).

3.7. Discussion and conclusions

By combining multiple covariates in various categories (demographic, topographic, meteorological, and satellite observations of NO₂), we demonstrated a machine learning approach based on the XGboost method to generate near-surface NO₂ concentration maps spatially at 100 m and temporally at hourly resolution for a domain covering northern Italy and Switzerland. Utilizing hourly NO₂ concentration measurements (June 2018–May 2020) from 380 operational air quality stations yielded NO₂ distribution maps that capture the general trends of daily, weekly and seasonal cycles in urban, suburban and rural areas across scales ranging from city districts to countries at unprecedented resolution (see Supplementary Video 1, Fig. S6). The model performance was reasonable for selected test stations including different human settlement modes, although there was an unavoidable bias towards urban regions owing to the design of monitoring network (Fig. 1). The case study of the COVID-19 pandemic lockdowns illustrates the possible implication of such models for the assessment of political interventions with respect to air pollutant distributions and dynamics.

The choice of tree models using the XGboost method was based on considerations of training and prediction efficiency; a computational necessity for high resolution mapping. Especially, the availability of XGboost on graphical processing units (GPUs) proved to be a great advantage for training compared to other approaches. The time to train an ensemble of 100 trees with 5 million observations was ~2 min with one GPU versus ~15 min with one CPU with 12 cores on the Cray XC50 (Intel Xeon E5-2690 v3 @ 2.60GHz (12 cores, 64GB RAM) and NVIDIA Tesla P100 16GB) at the Piz Daint high performance computer at Swiss National Supercomputing Centre (CSCS). The time to predict 20 million grid points on a single map was ~5 s with one GPU and ~14 s with one CPU. We note that, aside from using the XGBoost model itself, additional computational time is also required for data pre-processing (remapping, gap-filling, interpolation). In this study, we parallelized the map productions and it takes approximately 45 s for reading pre-processed input, reindexing, concatenating, and finally writing an output file of a NO₂ map using 36 CPU cores on the Piz Daint Cray XC40 system (Two Intel Xeon E5-2695 v4 @ 2.10GHz (2 × 18 cores)). Tree models, however, exhibit limitations especially regarding the ability of producing spatially continuous maps without step-wise changes. Using deep tree-depths and linear interpolation of spatial input data with resolutions much coarser than the output grid, we were able to resolve such discontinuities to a large extent. Nevertheless, some artifacts remained, which were most pronounced at locations with a complex distribution of emission sources. Other approaches have been proposed that are less affected by such issues, for instance kriging in residuals etc. (Young et al., 2016; Zhan et al., 2018; Chen et al., 2019). In this work, however, we did not take such an approach as the generation of hourly high-resolution maps would have been too costly. Furthermore, by keeping the model as a simple tree ensemble it was possible to explore the tree structure, analyze information gain and weights, and to attain local interpretability using SHAP. This enables a physical interpretation of the results, which would not be possible (or not as easy) with other methods.

We introduced two schemes that enable a prediction of NO₂ concentrations not only in terms of local conditions but also in terms of

spatial context. The first scheme is a decomposition of the spatial distributions of NO₂ emission proxies (population, traffic, etc.) into magnitude and distance (SI Text S1 and Fig. S3), which allows accounting for the impact of neighboring sources depending on their distance and magnitude. The second scheme is a wavelet transform of the elevation map providing spatial context in terms of surrounding topography (SI Text S2 and Fig. S4). Differentiating between the impact of topography on NO₂ concentration at stations located in valley floors versus mountain ridges was especially important in this study owing to the complex terrain of the Alpine domain. The decomposition of topography proved its importance in terms of SHAP values (Fig. 4). In the model, these topographic features provide critical information by interacting with meteorological conditions, especially boundary layer height, in determining the near-surface NO₂ values (see SI Fig. S9). We note that the impact of these topographic data will be reduced in areas with less-complex terrain since the model training eventually drops covariates with lower discrimination information. This ‘feature engineering’ improved the model performance in terms of overall R^2 , yet the produced maps of NO₂ show visible differences if we omit these two schemes of spatial context (see SI Fig. S10).

The choice of a suitable f_{QA} to filter TROPOMI observations needs to trade the impact of reduced data quality versus the impact of reduced data quantity. Filtering TROPOMI observations with a stricter f_{QA} of 0.75 causes a considerable imputation of TROPOMI data for training (> 40%) and gap-filling for cloud-covered scenes becomes increasingly uncertain with normalized convolutions (see SI Fig. S11). The uncertainties associated with a large fraction of gap-filled data had a negative impact on model performance: The model using TROPOMI observations with $f_{QA} > 0.75$ showed worse performance than the model including lower quality data ($0.50 < f_{QA} < 0.75$) (R^2 was reduced to 0.55 from 0.59). This implies that the TROPOMI data of reduced quality with $0.50 < f_{QA} < 0.75$, which mostly correspond to retrievals over fractional cloud covers, were still informative to the model. The model appears to be able to learn and compensate the enhanced errors in the lower quality TROPOMI data from the uncertainty measure of gap-filling and the meteorological data, especially from solar radiation as a proxy of cloud cover (see SI Fig. S12). In future studies it should be investigated whether additional information on cloud fraction and cloud top pressures from TROPOMI products or meteorological analyses could further improve the model predictions through a better compensation of errors in lower quality TROPOMI data.

The SHAP method and the total information gain of the tree model are two complementary ways of assessing the importance of each covariate, and both point to the pivotal role of TROPOMI NO₂ observations for the prediction of near-surface NO₂ concentrations. This is consistent with previously reported observations of a strong correlation between satellite-based NO₂ columns and near surface NO₂ concentrations (Zhou et al., 2009; Ialongo et al., 2019; Zheng et al., 2019). The role of TROPOMI observations was mainly in defining the concentration levels at large scale (placed at the root of the tree) while fine structures were driven by average traffic volume (Fig. 2) as demonstrated by the spatial distribution of SHAP values in Fig. 5 and Fig. S13 in SI. The SHAP analysis also showed that the temporal dynamics of the impact of emission sources such as traffic, which are provided as static average distributions, can be captured by additional time covariates such as hour of day (see Supplementary material 1) and weekday but also by time-dependent meteorological factors. Most notably, the statistical model was able to separate the influence of (traffic) emissions on the diurnal cycle of NO₂ concentrations from the influence of the diurnal evolution of boundary layer heights. Hourly traffic/industrial activity data could potentially improve the representation of temporal variability in the model, which would likely reduce the relative importance of the satellite observations. Such data are only gradually becoming available (Liu et al., 2020) and may be incorporated in the future.

SHAP analysis also informed a limitation of our model regarding the covariate, day of year. Initially, this covariate was included to capture

anthropogenic activity, such as annual holidays, and it was an effective covariate to detect the sudden changes caused by political interventions like pandemic lockdowns. However, we note that, owing to our training period including only 2 years, the importance of ‘day of year’ could have been overrated. Furthermore, the SHAP analysis showed the day of year covariate diminishes the importance of annual patterns in meteorological data. When the day of year is not included as a covariate, the importance of meteorological data increased together with TROPOMI observations, indicating that omitting this covariate can be an option for future applications for better interpretability with the price of model performance (see SI Fig. S14).

In this study, we focused on the Alpine domain but the approach can be readily extended to any other region in Europe thanks to the publicly available high quality data (see SI Table S1). For global applications, several challenges with respect to data availability remain, especially regarding in situ observations. To achieve a good model performance, a sufficient number of measurement stations covering different regions and air pollution levels is needed for model training, but the network does not have to be as dense as in our study region. In fact, we found that there is a saturation in model performance with regard to the number of stations in a training set (see SI Fig. S5). The best performing model using a representative selection of only 67 stations for training was able to outperform another model with a random selection of 340 stations. This implies that a representative set of in situ stations is more important than a dense network of stations, which tends to provide increasingly redundant information.

Another challenge for global applications is the lack of detailed information on traffic volume and industrial emission sources. However, omitting these two variables reduced model performance only moderately (R^2 decrease of 0.04) and including the satellite observations as a covariate could successfully compensate for the lack of these variables (see Fig. S5 in SI for comparisons). We note that this compensation by satellite observations was especially crucial when using only a small number of stations for training, even leading to better model performances compared to models using only ground information without the view from space. Therefore, using globally available data sets, such as population density, road length density, land use data, and topography, would still enable a downscaling of TROPOMI NO₂ observations to high-resolution surface NO₂ maps of adequate quality. As mentioned previously, a set of well-represented (even in a small number) stations for training would be an advantage aiming at high quality.

The COVID-19 lockdowns proved the critical role of political interventions and more stringent regulations for reducing the concentrations of critical air pollutants like NO₂ in the air we breathe. Detailed maps of NO₂ and of its temporal dynamics as presented in this study are critically needed as input for epidemiological studies and to assess public (even personal) exposure not only in relation to emission sources but also to meteorological factors. Regarding the exposure-mitigation strategies, an adequate guideline for the annual threshold of NO₂ levels can be revisited considering this spatio-temporal variability as shown in different regions and seasons with the help from satellites.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2021.112573>.

Author contributors

Minsu Kim: Methodology, Investigation, Software, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing. Dominik Brunner: Conceptualization, Investigation, Funding acquisition, Supervision, Writing- Reviewing and Editing. Gerrit Kuhlmann: Conceptualization, Investigation, Data curation, Funding acquisition, Supervision, Project administration, Writing- Reviewing and Editing.

Declaration of Competing Interest

None.

Acknowledgement

The authors thank the NABEL network operated by BAFU and Empa for providing additional data from Swiss air quality stations. We also acknowledge the free use of tropospheric NO₂ column data from the TROPOMI sensor from www.temis.nl. This project was funded by an Empa internal project and the model training/prediction was carried out using CSCS (Swiss National Supercomputing Centre) resources.

References

- Achakulwisut, P., Brauer, M., Hystad, P., Anenberg, S.C., 2019. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO₂ pollution: estimates from global datasets. *Lancet Planet Health* 3, e166–e178.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY, USA, pp. 2623–2631.
- Babyak, M.A., 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* 11.
- Bauwens, M., Compennolle, S., Stavrakou, T., Müller, J.F., van Gent, J., Eskes, H., Levelt, P.F., van der A, R., Veefkind, J.P., Vlietinck, J., Yu, H., Zehner, C., 2020. Impact of coronavirus outbreak on NO₂ pollution assessed using TROPOMI and OMI observations. *Geophys. Res. Lett.* 47, e2020GL087978.
- Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National spatiotemporal exposure surface for NO₂: monthly scaling of a satellite-derived land-use regression, 2000–2010. *Environ. Sci. Technol.* 49, 12297–12305.
- Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W.W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A.S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., Faire, U. D., Pedersen, N.L., Ostenson, C.G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., de Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Gironi, S., Krogh, V., Tsai, M.Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulos, A., Brunekreef, B., Hoek, G., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 383, 785–795.
- Bouwman, A.F., Van Vuuren, D.P., Derwent, R.G., Posch, M., 2002. A global analysis of acidification and eutrophication of terrestrial ecosystems. *Water Air Soil Pollut.* 141, 349–382.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. *Lancet* 360, 1233–1242.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. ACM Press, San Francisco, California, USA, pp. 785–794.
- Chen, Z.Y., Zhang, R., Zhang, T.H., Ou, C.Q., Guo, Y., 2019. A kriging-calibrated machine learning method for estimating daily ground-level NO₂ in mainland China. *Sci. Total Environ.* 690, 556–564.
- de Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E.A., Strassmann, A., Puhon, M., Röösi, 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.
- Dias, D., Tchepel, O., 2018. Spatial and temporal dynamics in air pollution exposure assessment. *Int. J. Environ. Res. Public Health* 15, 558.
- Dijkstra, L., Poelman, H., 2014. A harmonised definition of cities and rural areas: the new degree of urbanisation. *Eur. Comm. Urb. Reg. Pol.* 28.
- EEA, 2020a. Air Quality in Europe - 2020 Report : European Environment Agency. Publication 9/2020. European Environment Agency (EEA).
- EEA, 2020b. AirBase – The European Air Quality Database (Version 8) - European Environment Agency. <https://www.eea.europa.eu/data-and-maps/data/airepo-rting-8>.
- Eskes, H., van Geffen, J., Boersma, F., Eichmann, K.U., Apituley, A., Pedernana, M., Sneep, M., Veefkind, J.P., Loyola, D., 2019. Sentinel-5 Precursor/TROPOMI Level 2 Product User Manual Nitrogen dioxide.
- EU, EEA, 2016. Copernicus Land Monitoring Service, European Digital Elevation Model (EU-DEM v 1.1). <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/>.
- EU, EEA, 2018. Copernicus Land Monitoring Service, CORINE Land Cover (CLC 2018). <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>.
- Faustini, A., Rapp, R., Forastiere, F., 2014. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. *Eur. Respir. J.* 44, 744–753.
- Freire, S., Dossay-Whitfield, E., MacManus, K., Mills, J., Pesaresi, M., 2016. Development of new open and free multi-temporal global population grids at 250 m resolution. In: *AGILE. AGILE*, Helsinki, p. 6.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- Goldberg, D.L., Anenberg, S.C., Griffin, D., McLinden, C.A., Lu, Z., Streets, D.G., 2020. Disentangling the impact of the COVID-19 lockdowns on urban NO₂ from natural variability. *Geophys. Res. Lett.* 47, e2020GL089269.
- Grange, S.K., Lee, J.D., Drysdale, W.S., Lewis, A.C., Hueglin, C., Emmenegger, L., Carslaw, D.C., 2020. COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban areas. *Atmos. Chem. Phys. Discuss.* 1–25.
- Hanigan, I.C., Williamson, G.J., Knibbs, L.D., Horsley, J., Rolfe, M.L., Cope, M., Barnett, A.G., Cowie, C.T., Heyworth, J.S., Serre, M.L., Jalaludin, B., Morgan, G.G., 2017. Blending multiple nitrogen dioxide data sources for neighborhood estimates of long-term exposure for health research. *Environ. Sci. Technol.* 51, 12473–12480.
- Ialongo, I., Virta, H., Eskes, H., Hovila, J., Douras, J., 2019. Comparison of TROPOMI/Sentinel 5 Precursor NO₂ Observations with Ground-Based Measurements in Helsinki. Preprint. *Gases/Remote Sensing/Validation and Intercomparisons*.
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *J. Expos. Sci. Environ. Epidemiol.* 15, 185–204.
- Johnson, M., Isakov, V., Touma, J.S., Mukerjee, S., Özkaynak, H., 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* 44, 3660–3668.
- Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., 2021. Premature mortality due to air pollution in European cities: a health impact assessment. *Lancet Planet Health* 5 (3), e121–e134. [https://doi.org/10.1016/S2542-5196\(20\)30272-2](https://doi.org/10.1016/S2542-5196(20)30272-2).
- Kim, Minsu, Gerrit, Kuhlmann, Dominik, Brunner, 2021. Dataset for: Importance of satellite observations for high-resolution mapping of near-surface NO₂ by machine learning. Zenodo. <http://doi.org/10.5281/zenodo.5036643>, 2021.
- Knutsson, H., Westin, C.F., 1993. Normalized and differential convolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Comput. Soc. Press, New York, NY, USA, pp. 515–523.
- Kuenen, J.J.P., Visschedijk, A.J.H., Jozwicka, M., Denier van der Gon, H.A.C., 2014. TNO-MACC II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. *Atmos. Chem. Phys.* 14, 10963–10976.
- Kuhlmann, G., Hartl, A., Cheung, H.M., Lam, Y.F., Wenig, M.O., 2014. A novel gridding algorithm to create regional trace gas maps from satellite observations. *Atmos. Measur. Tech.* 7, 451–467.
- Larkin, A., Geddes, J.A., Martin, R.V., Xiao, Q., Liu, Y., Marshall, J.D., Brauer, M., Hystad, P., 2017. Global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* 51, 6957–6964.
- Le, T.T., Fu, W., Moore, J.H., 2020. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. 36 (1) *Bioinformatics*, pp. 250–256.
- Li, L., 2019. Geographically weighted machine learning and downscaling for high-resolution spatiotemporal estimations of wind speed. *Remote Sens.* 11, 1378.
- Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S.J., Feng, S., Zheng, B., Cui, D., Dou, X., Zhu, B., Guo, R., Ke, P., Sun, T., Lu, C., He, P., Wang, Y., Yue, X., Wang, Y., Lei, Y., Zhou, H., Cai, Z., Wu, Y., Guo, R., Han, T., Xue, J., Boucher, O., Boucher, E., Chevallier, F., Tanaka, K., Wei, Y., Zhong, H., Kang, C., Zhang, N., Chen, B., Xi, F., Liu, M., Bréon, F. M., Lu, Y., Zhang, Q., Guan, D., Gong, P., Kammen, D.M., He, K., Schellnhuber, H.J., 2020. Near-real-time monitoring of global CO₂ emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* 11, 5172.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA, pp. 4768–4777.
- Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.W., Newman, S.F., Kim, J., Lee, S.I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* 2, 749–760.
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67.
- Maiheu, B., Lefebvre, W., Walton, H., Dajnak, D., Janssen, S., Blyth, L., Beevers, S., 2017. Improved Methodologies for NO₂ Exposure Assessment in the EU. Technical Report 2017/RMA/R/1250.
- McInnes, L., Healy, J., 2017. Accelerated hierarchical density based clustering. In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 33–42.
- McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. [arXiv:1802.03426 \[cs, stat\]](https://arxiv.org/abs/1802.03426) 1802.03426.
- 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, e006946.
- Mueller, M.D., Wagner, M., Barmapadimos, I., Hueglin, C., 2015. Two-week NO₂ maps for the City of Zurich, Switzerland, derived by statistical modelling utilizing data from a routine passive diffusion sampler network. *Atmos. Environ.* 106, 1–10.
- Prunet, P., Lezeaux, O., Camy-Peyret, C., Thevenon, H., 2020. Analysis of the NO₂ tropospheric product from S5P TROPOMI for monitoring pollution at city scale. *City Environ. Interact.* 8, 100051.
- Rodríguez-Pérez, R., Bajorath, J., 2020. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J. Comput. Aided Mol. Des.* 34, 1013–1026.
- Shapley, L.S., 1953. 17. Princeton University Press, A Value for n-Person Games.
- Tanzer-Gruener, R., Li, J., Eilenberg, S.R., Robinson, A.L., Presto, A.A., 2020. Impacts of modifiable factors on ambient air pollution: a case study of COVID-19 shutdowns. *Environ. Sci. Technol. Lett.* 7 (8), 554–559.
- van Geffen, J., Eskes, H., Boersma, K., Maasakkers, J., Veefkind, J., 2019. TROPOMI ATBD of the Total and Tropospheric NO₂ Data Products, Report S5P-KNMI-L2-0005-RP. Technical Report. KNMI. De Bilt, The Netherlands.
- van Stratum, B.J.H., Vilà-Guerau de Arellano, J., Ouwersloot, H.G., van den Dries, K., van Laar, T.W., Martinez, M., Lelieveld, J., Diesch, J.M., Drewnick, F., Fischer, H., Hosaynali Beygi, Z., Harder, H., Regelin, E., Sinha, V., Adame, J.A., Sörgel, M.,

- Sander, R., Bozem, H., Song, W., Williams, J., Yassaa, N., 2012. Case study of the diurnal variability of chemically active species with respect to boundary layer dynamics during DOMINO. *Atmos. Chem. Phys.* 12, 5329–5341.
- Veeckman, C., Jedlička, K., De Paepe, D., Kozhukh, D., Kafka, S., Colpaert, P., Čerba, O., 2017. Geodata interoperability and harmonization in transport: a case study of open transport net. *Open Geospatial Data, Software and Standards* 2, 3.
- Vienneau, D., de Hoogh, K., Bechle, M.J., Beelen, R., van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite- and ground-based measurements of NO₂ and PM₁₀. *Environ. Sci. Technol.* 47, 13555–13564.
- Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO₂. *Environ. Sci. Technol.* 46, 4481–4489.
- Xiao, Q., Chang, H.H., Geng, G., Liu, Y., 2018. An ensemble machine-learning model to predict historical PM_{2.5} concentrations in China from satellite data. *Environ. Sci. Technol.* 52, 13260–13269.
- Xu, H., Bechle, M.J., Wang, M., Szpiro, A.A., Vedal, S., Bai, Y., Marshall, J.D., 2019. National PM_{2.5} and NO₂ exposure models for China based on land use regression, satellite measurements, and universal kriging. *Sci. Total Environ.* 655, 423–433.
- Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO₂ and model validation in a national prediction model based on universal kriging and land-use regression. *Sci. Technol.* 50, 3686–3694.
- 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.
- Zhang, N., Zhou, Y., 2019. Climate-driven changes in CO₂ emissions associated with residential heating and cooling demand by end-century in China. *Environ. Res. Lett.* 14, 084043.
- Zheng, Z., Yang, Z., Wu, Z., Marinello, F., 2019. Spatial variation of NO₂ and its impact factors in China: an application of sentinel-5P products. *Remote Sens.* 11, 1939.
- Zhou, Y., Brunner, D., Boersma, K.F., Dirksen, R., Wang, P., 2009. An improved tropospheric NO₂ retrieval for OMI observations in the vicinity of mountainous terrain. *Atmos. Measur. Tech.* 2, 401–416.