Atmos. Chem. Phys., 23, 10267–10285, 2023
https://doi.org/10.5194/acp-23-10267-2023

Atmospheric Chemistry and Physics
Open Access
EGU

Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning

Vigneshkumar Balamurugan[1], Jia Chen[1], Adrian Wenzel[1], and Frank N. Keutsch[2,3]

[1]Environmental Sensing and Modeling, Technical University of Munich (TUM), Munich, Germany
[2]School of Engineering and Applied Science, Harvard University, Cambridge, MA, USA
[3]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

Correspondence: Vigneshkumar Balamurugan (vigneshkumar.balamurugan@tum.de) and Jia Chen (jia.chen@tum.de)

Received: 13 March 2023 – Discussion started: 6 April 2023
Revised: 28 July 2023 – Accepted: 31 July 2023 – Published: 14 September 2023

Abstract. Machine learning (ML) models are becoming a meaningful tool for modeling air pollutant concentrations. ML models are capable of learning and modeling complex nonlinear interactions between variables, and they require less computational effort than chemical transport models (CTMs). In this study, we used gradient-boosted tree (GBT) and multi-layer perceptron (MLP; neural network) algorithms to model near-surface nitrogen dioxide (NO_2) and ozone (O_3) concentrations over Germany at 0.1° spatial resolution and daily intervals.

We trained the ML models using TROPOspheric Monitoring Instrument (TROPOMI) satellite column measurements combined with information on emission sources, air pollutant precursors, and meteorology as feature variables. We found that the trained GBT model for NO_2 and O_3 explained a major portion of the observed concentrations ($R^2 = 0.68$–0.88 and RMSE $= 4.77$–$8.67\,\mu g\,m^{-3}$; $R^2 = 0.74$–0.92 and RMSE $= 8.53$–$13.2\,\mu g\,m^{-3}$, respectively). The trained MLP model performed worse than the trained GBT model for both NO_2 and O_3 ($R^2 = 0.46$–0.82 and $R^2 = 0.42$–0.9, respectively).

Our NO_2 GBT model outperforms the CAMS model, a data-assimilated CTM but slightly underperforms for O_3. However, our NO_2 and O_3 ML models require less computational effort than CTM. Therefore, we can analyze people's exposure to near-surface NO_2 and O_3 with significantly less effort. During the study period (30 April 2018 and 1 July 2021), it was found that around 36 % of people lived in locations where the World Health Organization (WHO) NO_2 limit was exceeded for more than 25 % of the days during the study period, while 90 % of the population resided in areas where the WHO O_3 limit was surpassed for over 25 % of the study days. Although metropolitan areas had high NO_2 concentrations, rural areas, particularly in southern Germany, had high O_3 concentrations.

Furthermore, our ML models can be used to evaluate the effectiveness of mitigation policies. Near-surface NO_2 and O_3 concentration changes during the 2020 COVID-19 lockdown period over Germany were indeed reproduced by the GBT model, with meteorology-normalized near-surface NO_2 having significantly decreased (by 23 ± 5.3 %) and meteorology-normalized near-surface O_3 having slightly increased (by 1 ± 4.6 %) over 10 major German metropolitan areas when compared to 2019. Finally, our O_3 GBT model is highly transferable to neighboring countries and locations where no measurements are available ($R^2 = 0.87$–0.94), whereas our NO_2 GBT model is moderately transferable ($R^2 = 0.32$–0.64).

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

Air pollution is a major threat to human health and impacts ecosystems (Bell et al., 2011; Lelieveld et al., 2015; Zhang et al., 2019; Xie et al., 2019). Based on the source of the pollution, air pollutants are classified as being primary (directly emitted from anthropogenic or natural sources) or secondary (formed through complex atmospheric chemical reactions). Near-surface nitrogen oxide ($NO_x = NO+NO_2$) is a primary air pollutant emitted largely by fossil-fuel-consuming sectors such as vehicles, industries, and power plants, but there are also natural sources such as lightning, soil emissions, and biomass burning. Near-surface ozone (O_3) is a secondary air pollutant produced solely by the photolysis of NO_2 (nitrogen dioxide) in the presence of sunlight (Crutzen, 1988; Council, 1992).

Tropospheric NO_x and O_3 are chemically strongly coupled via complex atmospheric chemical reactions (Jacob, 1999). The majority of NO_x, from primary sources such as fossil fuel combustion, is emitted in the form of nitric oxide (NO), which rapidly converts to NO_2 by reacting with O_3. In turn, O_3 and NO are generated again by the photolysis of NO_2, thus forming a null cycle. Therefore, the amount of sunlight present and the total concentration of NO_x determine ozone production via this NO_x null cycle. In addition, the oxidation of volatile organic compounds (VOCs) can alter the NO/NO_2 ratio. The presence of the hydroxyl radical (OH) initiates the VOC oxidation process, followed by the formation of hydro- and organic peroxy radicals. These radicals convert NO to NO_2 and form additional O_3, while also converting HO_2 back to OH, thus forming a catalytic cycle known as the HO_x catalytic cycle. However, ozone production is nonlinear in relation to its precursors (NO_x and VOC) due to termination reactions that occur within the catalytic cycle (Lin et al., 1988; Nussbaumer and Cohen, 2020; Pusede and Cohen, 2012; Pusede et al., 2014). To that end, the response of ozone production is categorized into three regimes, namely NO_x-saturated (high NO_x with low VOC); NO_x-limited (low NO_x with high VOC); and transitional (Sillman et al., 1990; Sillman, 1999). In the NO_x-saturated regime (typically urban areas), ozone production is inversely proportional to NO_x concentration, whereas ozone production is directly proportional to VOC concentration. However, in NO_x-limited regimes (typically rural areas), ozone production is directly proportional to NO_x concentration, whereas VOCs have little effect on ozone production. This complex ozone production vs. precursor emission response is also evident in real-time observations, such as urban weekend ozone levels being higher than weekday levels (Sicard et al., 2020) and high-ozone levels during public holidays and national shutdowns (e.g., the COVID-19 lockdown), due to low NO_x emissions (Balamurugan et al., 2021, 2022b).

Chemical transport models (CTMs) are commonly used to study air pollution and its drivers (Hu et al., 2016; Lou et al., 2015), but these models are dependent on emissions as represented in emission inventories (Pisoni et al., 2018). Emission inventories are typically developed using the bottom-up method, based on data such as economic activity, fuel consumption, and traffic density (McDuffie et al., 2020; Osses et al., 2022). However, bottom-up emission inventories can be highly uncertain due to inaccuracies in the data used in the bottom-up method, especially from unaccounted-for sources (Chen et al., 2020; Crippa et al., 2019; Forstmaier et al., 2023; Trombetti et al., 2018). Because of the significant computational effort and storage space requirements, CTMs often perform at a coarse spatial resolution, making it unable to solve fine transport and chemical mechanisms, particularly over complex topography (Singh et al., 2021). Machine learning (ML) models have been shown to be an effective complement to these computationally expensive CTMs (Vlasenko et al., 2021). The performance of machine learning models for modeling air pollutants is promising (Balamurugan et al., 2022a; Cheng et al., 2022; Lee et al., 2020; Li et al., 2023; Liang et al., 2020; Liu et al., 2022; Zaini et al., 2022; Zhao et al., 2023). Meteorological variables such as solar radiation and temperature have been shown to be important parameters in near-surface ozone modeling using machine learning (Diao et al., 2021; Hu et al., 2021). Meteorological conditions influence the concentration of O_3 both directly and indirectly. Solar UV radiation is responsible for the photolysis of O_3 precursors (NO_2 and VOCs). Temperature directly influences the photochemical reaction rate. Furthermore, meteorology influences biogenic and fuel-leak-related VOC emissions (exponentially proportional to temperature), which account for a significant portion of total VOC emissions (Guenther et al., 1993). In addition to meteorology, when the emission source information is included, ML models predict near-surface NO_2 very well (Ghahremanloo et al., 2021; De Hoogh et al., 2019).

In situ air quality measurements are sparse and concentrated primarily in urban areas. Recent advancements in satellite remote sensing allow us to analyze urban and non-urban air quality with adequate spatiotemporal coverage; however, they typically only measure the total or tropospheric column of specific air quality species, making it difficult to interpret people's exposure to the near-surface air pollutant concentration. Therefore, in this study, we trained two ML models for near-surface NO_2 and O_3 concentrations over Germany using available information on proxies for near-surface air pollutants (satellite column measurements) and emission sources, precursors of air pollutants, and meteorology. Many recent studies, similar to ours, have attempted to model near-surface NO_2 and O_3 concentrations at national or regional scales (De Hoogh et al., 2019; Kang et al., 2021; Kim et al., 2021; Li et al., 2020; Zhu et al., 2022); there are, however, very few attempts over Germany. To the best of the authors' knowledge, only one study (Chan et al., 2021) used the TROPOspheric Monitoring Instrument (TROPOMI) satellite NO_2 tropospheric column measurements and other auxiliary information (e.g., meteorology) to

model near-surface NO_2 concentrations over Germany using a multi-layer perceptron (MLP) model. Furthermore, previous studies have focused on a single pollutant (e.g., NO_2), whereas in this study, we model and analyze the spatiotemporal variations in both NO_2 and O_3, which are chemically strongly coupled. In terms of anthropogenic emissions, we also evaluate the ML model performance of NO_2 and O_3 during the 2020 COVID-19 lockdown period, which serves as a natural experiment period with significantly lower primary anthropogenic emissions (Gensheimer et al., 2021).

2 Study region, data sets, model, and method

All data sets used in this study, and their spatial and temporal resolutions, are summarized in Table 1.

2.1 Study region and near-surface NO_2 and O_3 measurements

We focused on the spatial domain of 5–15° E and 47–55.5° N, particularly over Germany. Near-surface NO_2 and O_3 data from measurement stations across Germany were used in this study. However, not all measuring stations collect data on both pollutants; there are fewer stations measuring O_3 than those measuring NO_2. There were also temporal gaps in the measurement data. Therefore, we only considered stations that had more than 80 % data coverage during the study period. In the end, we considered 321 stations for modeling NO_2 and 256 stations for modeling O_3. The selected measurement stations are located throughout the entire country and are situated in high-traffic, industrial, and background locations (Fig. 1 and Table A1).

2.2 Predictor variables of ML model

Predictor variables or input features for the ML models include satellite column measurements of air pollutants and meteorology and auxiliary data containing information on the area of interest.

2.2.1 Satellite column measurements

Tropospheric column NO_2, total column O_3, and tropospheric column formaldehyde (HCHO) data are used, which are level-2 retrieval products from TROPOMI, which is aboard the Sentinel-5P satellite. Sentinel-5P overpasses the study area between 13:00 and 14:00 LST (local standard time). The spatial resolution of TROPOMI data is 7 km × 3.5 km (increased to 5.5 km × 3.5 km after 6 August 2019). We applied the data quality filtering described in the product manual to each data product (S5P, 2022b, for NO_2; S5P, 2022c, for O_3; S5P, 2022a, for HCHO). Tropospheric column NO_2 is used in the NO_2 ML model because it can be considered to be a proxy for near-surface NO_2. Since NO_2 is the precursor for O_3, we also included the tropospheric column NO_2 in the O_3 ML model. Because HCHO is an intermediate gas product of VOC oxidation, it can be used as a proxy for VOC oxidation (Jin et al., 2017). Therefore, we included tropospheric column HCHO in the O_3 model. We also considered the TROPOMI FNR (ratio of TROPOMI HCHO and TROPOMI NO_2) in the O_3 ML model, which in previous studies has been shown to be a useful indicator of ozone production regime (Jin et al., 2020; Wang et al., 2021). We included total column O_3 in the O_3 ML model by considering total column O_3 as a proxy for near-surface O_3.

2.2.2 Vegetation index

Normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) data were obtained from MODIS (Moderate Resolution Imaging Spectroradiometer) measurements aboard the Terra and Aqua satellites. We used the MOD13A2 (16 d; 1 km) Vegetation Indices (VI) data set, which contains NDVI and EVI data at 1 km spatial resolution and 16 d temporal resolution. To generate daily intervals, the NDVI and EVI data were linearly interpolated. We considered these vegetation indexes in the O_3 ML model because vegetation contributes a considerable number of VOCs. We also considered these vegetation indexes in the NO_2 ML model as supplementary information to check whether changes in vegetation cover have any implications for changes in the NO_2 concentration.

2.2.3 Meteorology

Meteorology has both direct and indirect effects (e.g., dispersion and photochemical reactions) on pollutant concentrations. Meteorological variables such as temperature (T), relative humidity (RH), wind speed (WS), and wind direction (WD) were obtained from the ERA5 reanalysis product. These variables were derived from the lowest model level (1000 hPa) of the ERA5 hourly data on pressure levels data set. Downward UV solar radiation at the surface (DUV), boundary layer height (BLH), surface pressure (SP), and temperature of the air at 2 m above the surface ($T_{2\,m}$) were derived from the ERA5 hourly data on single levels data set. These meteorological data have a spatial resolution of 0.25° and a temporal resolution of 1 h. In both the NO_2 and O_3 ML models, we took all meteorology variables into account.

2.2.4 Proxy for NO_x emission source

Because vehicle (transport sector) emissions are a significant source of NO_x emissions, considering a proxy for vehicle emissions is crucial. Therefore, we used road density as a proxy for the source of NO_x emissions. We are aware that traffic volume or density would be the ideal proxy, but data on the traffic volume or density on a national or regional scale are not available. The road density (RD) data were obtained

Table 1. Data sets and related information used in this study.

Data source	Data (purpose)	Temporal resolution	Spatial resolution
Governmental in situ measurements	Near-surface NO_2 and O_3 (ground truth data)	1 h	–
TROPOMI satellite measurements	Tropospheric column NO_2, total column O_3 and total column HCHO (input features)	Daily	7 km × 3.5 km (5.5 km × 3.5 km, after 6 August 2019)
ERA5 (ECMWF reanalysis)	Temperature, relative humidity, wind speed, wind direction, downwind UV solar radiation at surface, boundary layer, height surface pressure, and temperature of air at 2 m above the surface (input features)	1 h	0.25° × 0.25°
U.S. Geological Survey	Surface elevation (input features)	–	1 km × 1 km
GRIP global roads database	Road density (input features)	–	8 km × 8 km
CAMS European air quality forecasts	Near-surface NO_2 and O_3 (for validation)	1 h	0.1° × 0.1°
GEOS-Chem chemical transport model	Near-surface NO_2 and O_3 (for disentangling meteorology impacts)	1 h	0.5° × 0.625°

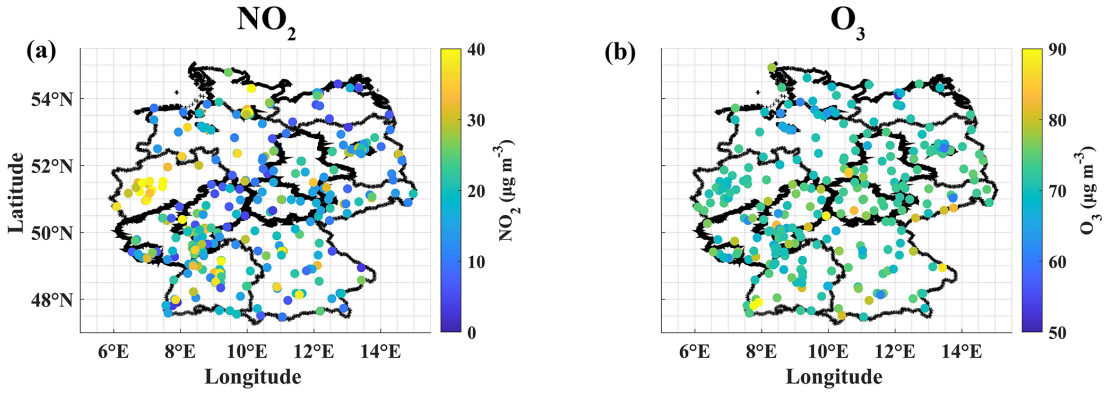

Figure 1. Locations of near-surface NO_2 **(a)** and O_3 **(b)** measurement stations considered in this study. The color bar depicts the mean of near-surface NO_2 and O_3 for each measurement station during the study period.

from the Global Roads Inventory Project (GRIP) database, with a spatial resolution of 8 km.

2.2.5 Additional features

Additional supplementary data, such as surface elevation (E), were obtained from the U.S. Geological Survey (USGS), with a spatial resolution of 1 km. Surface elevation was taken into account because it influences the tropospheric or total column value of measurements. We also considered the DOW (day of the week) and season (season of the year) information in both the NO_2 and O_3 models, since both NO_2 and O_3 have distinct weekly and seasonal cycles. Because NO_2 is an important precursor to O_3, in addition to TROPOMI NO_2, we also included near-surface NO_2 modeled from the NO_2 ML model as a feature variable in the O_3 ML model.

2.3 Study period and data pre-processing

The study period was chosen to be between 30 April 2018 and 1 July 2021, which corresponds to the availability of TROPOMI data retrievals with the same processing version. Despite the fact that satellites pass over the study area between 13:00 and 14:00 LST, we found that the satellite data represent the daily mean of air pollutants well. Therefore, we

considered the daily 24 h mean for near-surface NO_2 and the daily maximum 8 h mean (i.e., the mean of the eight highest hourly values during a day) for near-surface O_3 as our variables of interest (dependent variables to model), as these are commonly used metrics in air quality research (Hoffmann et al., 2021).

Because each data set has a different spatiotemporal resolution, we resampled all of the data to the same spatial ($0.1° \times 0.1°$) and temporal (daily) resolution. The $0.1°$ ($\approx$ 10 km) resolution was chosen because it corresponds to the resolution of the main features, such as road density (spatial resolution of 8 km), TROPOMI satellite measurements (spatial resolution of 7 km $\times$ 3.5 km), and concurrent high-resolution ($0.1°$) air quality forecasts from CAMS (Copernicus Atmosphere Monitoring Service). We computed the daily 24 h mean for near-surface NO_2 and the daily maximum 8 h mean for near-surface O_3 for each in situ measurement station and then calculated the mean of all stations that fell within the $0.1°$ grid. The mean of surface elevation, NDVI, EVI, TROPOMI (NO_2, HCHO, and O_3), and road density for each day were then calculated for the corresponding $0.1°$ grids. The surface elevation and road density were assumed to be constant during the study period. The ERA5 meteorology product was resampled to $0.1°$ resolution using the nearest-neighbor method, and the 24 h mean was computed.

2.4 Machine learning model and evaluation strategies

We primarily used the gradient-boosted tree (GBT) machine learning algorithm, XGBoost (Chen and Guestrin, 2016), to model near-surface NO_2 and O_3 concentrations. The GBT algorithm is a gradient-boosted decision-tree-based algorithm that is expected to outperform deep-neural-network-based algorithms for structured data (Lundberg et al., 2020). Furthermore, tree-based models are more interpretable and require less time to train than deep neural network algorithms. However, for comparison, we also used the multi-layer perceptron (MLP; neural network) algorithm (Gardner and Dorling, 1998). The GBT and MLP algorithms were implemented using scikit-learn, a Python module (https://scikit-learn.org/stable/, last access: 10 March 2023). When training the MLP model, we normalized the discrete feature variables between 0 and 1. The corresponding predictor variables and data flow for the NO_2 and O_3 ML model is shown in Fig. 2.

To evaluate the ML model, we used the R^2 (coefficient of determination) and RMSE (root mean square error) metrics. We split the available data into training (70 % of the data) and testing (the remaining 30 %). The training data set was used to iteratively vary the hyperparameters (combinations) and select the best set of hyperparameters using a 5-fold CV (cross-validation). The hyperparameters used in this study are shown in Tables A2 and A3. We also evaluated the ML model using three different 5-fold CV testing strategies (random 5-fold CV, time-leave-out 5-fold CV, and location-leave-out 5-fold CV) with 100 % of the data (Meyer et al., 2018). In the random 5-fold CV testing strategy, the data were randomly split into five parts, four of which were used for training and one for testing. This procedure was repeated until all five parts had been used as test. The mean (and standard deviation) of R^2 and RMSE from the 5-fold CV were then computed. In the time-leave-out 5-fold CV testing strategy, the 5-fold CV procedure was the same, but the data were split based on time period (by date; i.e., from the start of study period to the end of study period). Similarly, in the location-leave-out 5-fold CV testing strategy, the data were split based on location (by latitude). Figure A1 shows the first 1-fold step in a 5-fold CV for time-leave-out and location-leave-out testing strategies. To interpret the importance of feature variables in the fitted model, we use SHAP (SHapley Additive exPlanations) values. The SHAP method (https://christophm.github.io/interpretable-ml-book/shap.html, last access: 10 March 2023) is the most commonly used method for interpreting ML model output, which calculates the contribution of each feature variable to the final prediction. Thus, higher SHAP values indicate greater feature importance.

2.5 CAMS model data

We obtained near-surface NO_2 and O_3 air quality forecasts from CAMS in order to compare the performance of our ML model to that of the chemical transport model. This data set is based on a data assimilation technique that combines real-time measurements with an ensemble of 11 air quality models to provide air quality data with high spatial resolution ($0.1°$) and 1 h temporal resolution over Europe; however, it is only available for 3 years in the rolling archive. We used data from 17 July 2019 to 31 January 2020. We did not use data after 31 January 2020 due to COVID-19 lockdown restrictions, during which many anthropogenic emission activities were limited, and CAMS had not adjusted the emission inventory for changes in emissions. This is because NO_2 has a short lifetime, so the effect of assimilated observations is minimal, and the CAMS-forecasted NO_2 product mostly reflects emissions prescribed in the inventory (Inness et al., 2015).

2.6 GEOS-Chem model data

In this study, GEOS-Chem (Goddard Earth Observing System with Chemistry; hereafter GC) chemical transport model simulations were used to normalize the meteorology effects when estimating the influence of the COVID-19 lockdown restrictions on air pollutant concentration changes. The GC simulations over the study area were obtained with a spatial resolution of $0.5° \times 0.625°$ and 1 h temporal resolution for the 2020 strict COVID-19 lockdown period (21 March to 31 May) and the same period in 2019. Identical anthropogenic emissions from the 2014 Community Emissions Data System (CEDS) inventory were used for both 2020 and 2019 but with the corresponding meteorology, natural, and

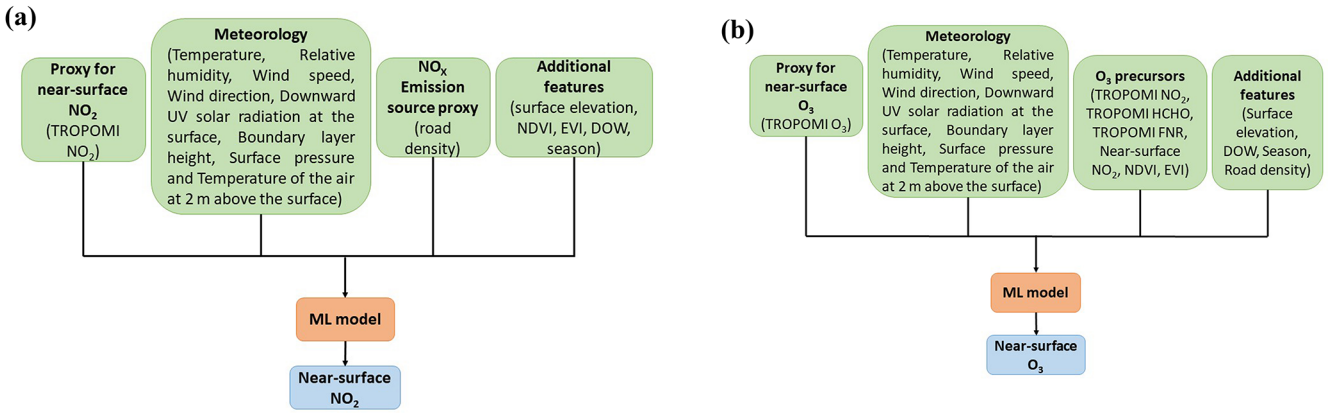

Figure 2. Predictor variables and data flow for the NO_2 **(a)** and O_3 **(b)** ML model.

fire emissions in the respective years. Therefore, the difference in GC-simulated species (X) concentrations between 2020 and 2019 results from changes in meteorology, natural, and fire emissions between 2020 and 2019 (GC $X_{2020–2019}$); here, X refers to either NO_2 or O_3. Then, we subtracted the GC $X_{2020–2019}$ from the observed near-surface $X_{2020–2019}$ to estimate the changes in the concentrations of species X due to changes in anthropogenic emissions in the 2020 lockdown period (refer to the studies of Balamurugan et al., 2021; Qu et al., 2021, for detailed descriptions of the method).

3 Results

3.1 ML model evaluation and feature importance

The trained GBT model with 70 % of the data (78 433) for NO_2 reproduced the observed NO_2 concentration well in the test case (33615), with an R^2 of 0.88 and RMSE of 4.77 $\mu g\,m^{-3}$ (Fig. 3a and Table 2). The random 5-fold CV results were in the same range ($R^2 = 0.89 \pm 0.002$ and RMSE $= 4.65 \pm 0.034\,\mu g\,m^{-3}$). The other two testing strategies (time-leave-out 5-fold CV and location-leave-out 5-fold CV) showed slightly worse agreement (Table 2), indicating that different validation strategies should be performed to interpret the ML model capability. Otherwise, it may result in an overoptimistic view of ML models (Meyer et al., 2018). Furthermore, the worse agreement in the location-leave-out 5-fold CV testing strategy suggests that there is less confidence in modeling the near-surface NO_2 over new locations that the GBT model has not been trained on before. However, these results outperformed the MLP model trained by another study (Chan et al., 2021; $R = 0.8$ and RMSE $= 6.32\,\mu g\,m^{-3}$ obtained for the testing strategy of the random split of 90 % of the data used for training and 10 % of the data used for testing) for near-surface NO_2 over Germany. Feature importance, based on the SHAP values, indicates that road density is the most important feature in the fitted model for NO_2 (Fig. 3c) because traffic is the main source of near-surface NO_x in urban areas. The next most important features were TROPOMI NO_2, boundary layer height, and elevation. Because the majority of NO_x sources are present at the surface, tropospheric column NO_2 data play an important role in explaining near-surface NO_2. Near-surface NO_2 typically has a negative correlation with the boundary layer height, as an increasing BLH disperses more, and vice versa (Balamurugan et al., 2021). Therefore, BLH is one of the most important features. It is unexpected that elevation was an important feature. The cause could be that the surface elevation varies greatly across Germany, influencing the total tropospheric column of NO_2 and thus serving as a link between the tropospheric column of NO_2 and near-surface NO_2. A previous study (Chan et al., 2021) also found that elevation was an important feature in the fitted MLP model for near-surface NO_2 over Germany.

The GBT model trained with 70 % of the data (65 705) for O_3 also had a good representation of the observed O_3 concentrations in the test case (28 160), with an R^2 of 0.92 and RMSE of 8.53 $\mu g\,m^{-3}$ (Fig. 3b). Similar to the NO_2 GBT model findings, time-leave-out 5-fold CV and location-leave-out 5-fold CV testing strategies showed less agreement than the random 5-fold CV testing strategy (Table 2). In comparison to our NO_2 GBT model, our O_3 GBT model demonstrated greater confidence in modeling near-surface O_3 over locations that the model was not trained on. According to SHAP values, the five most important features were DUV, T, RH, BLH, and season, with DUV having the greatest influence (Fig. 3d). Because ozone is formed in the atmosphere from the photolysis of NO_2, DUV plays a significant role in the fitted model that explains near-surface O_3. Temperature is the second most important feature, which is also not surprising, as it drives biogenic VOC emissions (an important precursor to O_3). Previous studies also show similar findings (Diao et al., 2021; Hu et al., 2021). GBT-modeled near-surface NO_2 was the sixth most important feature in the fitted model, according to the SHAP values, and it was also more important than TROPOMI NO_2.

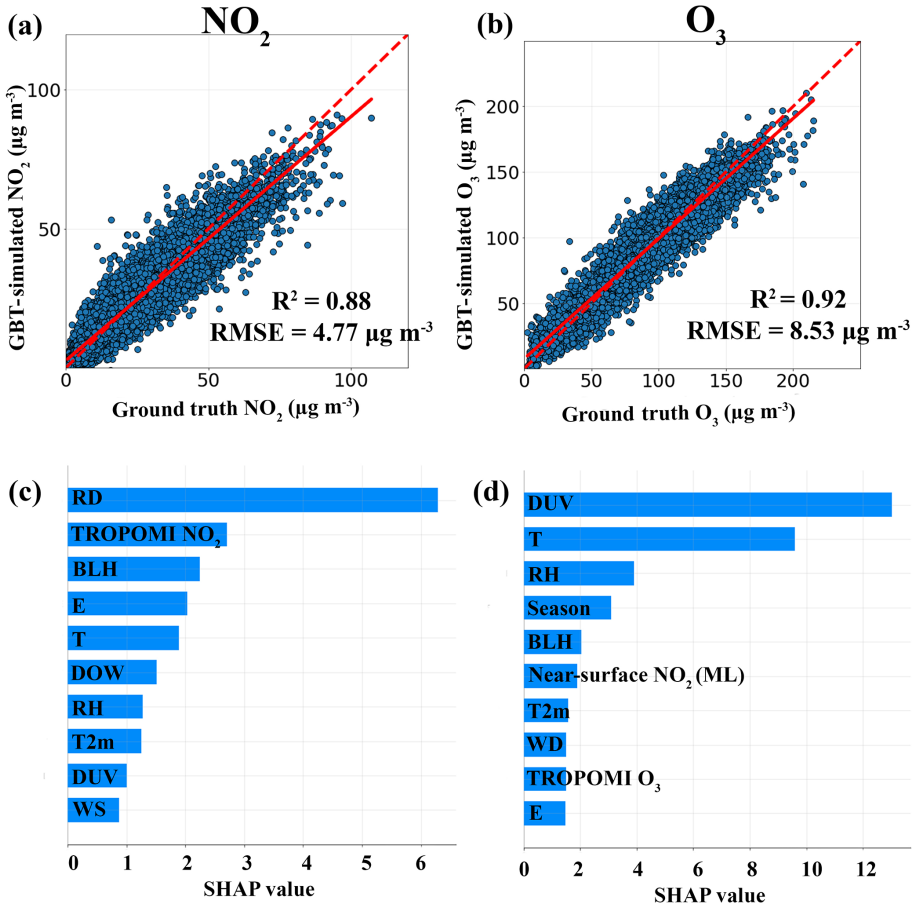

Figure 3. Comparison between the ground truth and GBT-simulated near-surface NO_2 **(a)** and O_3 **(b)**. The feature importance (top 10) is calculated based on SHAP (SHapley Additive exPlanations) values for NO_2 **(c)** and O_3 **(d)** in the GBT model. RD is for road density, BLH is for boundary layer height, E is for surface elevation, T is for temperature, DOW is for day of the week, RH is for relative humidity, $T_{2\,m}$ is for temperature at 2 m height, DUV is for downwind UV radiation, WS is for wind speed, and WD is for wind direction.

Table 2. Evaluation metrics of our GBT model in different testing strategies.

		Random (1-fold)	Random (5-fold)	Time leave out (5-fold)	Location leave out (5-fold)
NO_2 GBT model	R^2	0.88	0.89 ± 0.002	0.74 ± 0.07	0.68 ± 0.12
	RMSE ($\mu g\,m^{-3}$)	4.77	4.65 ± 0.034	6.77 ± 0.7	8.67 ± 1
O_3 GBT model	R^2	0.92	0.92 ± 0.001	0.74 ± 0.09	0.8 ± 0.06
	RMSE ($\mu g\,m^{-3}$)	8.53	9.36 ± 0.068	13.2 ± 1.1	12.45 ± 1.3

Figure A2 shows the results obtained from the MLP model. Both the NO_2 and O_3 MLP models performed worse than the NO_2 and O_3 GBT models, respectively (Table A4 vs. Table 2). In particular, the MLP model findings showed poorer agreement in the time-leave-out 5-fold CV and location-leave-out 5-fold CV testing strategies. This supports previous studies (Heaton, 2020; Lundberg et al., 2020) and shows that the MLP model is unlikely to outperform tree-based models for tabular data. Because the GBT model outperforms the MLP model, we only considered the GBT model results in the following.

It is important to note that deep learning models are data intensive, and their performance and generalization capabilities tend to improve with larger amounts of data. In our study, we utilized the simplest deep learning algorithm known as MLP. However, it is essential to explore the capabilities of

other deep learning algorithms, such as the CNN (convolutional neural network) and LSTM (long short-term memory), in future studies to gain further insight. Additionally, employing multiple ML models through bagging techniques could potentially lead to improved performance, despite the computational expense involved (He et al., 2022).

3.2 GBT model performance compared to CAMS

To evaluate how well our GBT model performs compared to CAMS, we compared the high-resolution near-surface NO_2 and O_3 forecasts from CAMS with observations and GBT-simulated near-surface NO_2 and O_3 with observations for the period between 17 July 2019 and 31 January 2020 (i.e., CAMS comparison period; Fig. 4). Please note this time period was not used for training the GBT model for this comparison. Our NO_2 GBT model reproduced the observed near-surface NO_2 concentrations well during this comparison period, with an R^2 of 0.82 and RMSE of 5.76 $\mu g\,m^{-3}$, while CAMS NO_2 forecasts showed poor representation ($R^2 = 0.37$ and RMSE $= 14.96\,\mu g\,m^{-3}$). However, CAMS O_3 forecasts agreed slightly better with the observed concentrations ($R^2 = 0.93$ and RMSE of 9.2 $\mu g\,m^{-3}$) when compared to our O_3 GBT model ($R^2 = 0.85$ and RMSE $= 13\,\mu g\,m^{-3}$). Our NO_2 GBT model outperforms CAMS due to the fact that the effect of the data assimilation on the CAMS NO_2 forecast product is minimal, with CAMS simulations mostly reflecting the emissions provided in the inventory. Additionally, it should be noted that our GBT model requires less computational effort than the CAMS model.

3.3 Spatiotemporal changes in near-surface NO_2 and O_3 over the study domain

After the discussed model evaluation, we trained the GBT model using 100 % of the data and modeled the near-surface NO_2 and O_3 concentrations over the study domain at 0.1° resolution and daily (24 h mean for NO_2 and 8 h maximum mean for O_3) intervals. The averaged GBT-modeled near-surface NO_2 concentrations over the study domain during the study period are shown in Fig. 5a. The spatial variability in the near-surface NO_2 correlates with Germany's population density, and the main hotspots correspond to Germany's major metropolitan areas (Fig. A3). The study domain's main hotspot is western Germany (North Rhine-Westphalia; a federal state of Germany), which is Germany's industrial heartland. The number of days (%) that exceeded the 2021 World Health Organization (WHO) NO_2 limit (24 h mean $> 25\,\mu g\,m^{-3}$) over major metropolitan areas in Germany was more than 50 %, with western Germany (North Rhine-Westphalia) experiencing the most days of exceedance during the study period (Fig. 7). Around 36 % of people live in locations where more than 25 % of the days exceed the WHO NO_2 limit during the study period (Fig. 8). The GBT-simulated near-surface O_3 showed a distinct spatial variability compared to NO_2, with high O_3 concentrations over southern Germany and low O_3 concentrations over northern Germany (Fig. 6). This could be due to the fact that O_3 is a secondary pollutant that is primarily driven by photochemical reactions influenced by meteorology, DUV, and temperature values, which were the factors with the most influence on photochemical reactions; accordingly, the most important features fitted in the O_3 GBT model were higher in southern Germany than northern Germany (Fig. A4). During the study period, more than 50 % of the study days in southern Germany exceeded the 2021 WHO O_3 limit (maximum 8 h mean $> 100\,\mu g\,m^{-3}$). Nearly 90 % of people live in locations where more than 25 % of the study days exceed the WHO O_3 limit (Fig. 8). Another interesting fact is that southern metropolitan areas and high NO_x regions have fewer days that exceeded the WHO O_3 limit than southern rural regions (Fig. 7). It is a well-known fact that rural regions have higher ozone levels than urban regions (Malashock et al., 2022). It could be because NO is a significant O_3 scavenger in higher NO_x (NO_2 is a proxy for NO_x) regions or because NO is in a NO_x-saturated regime. Furthermore, it is due to the fact that rural regions are the downwind locations of emission plumes and are the primary source of biogenic VOC emissions (Zong et al., 2018).

We also evaluated the model's capability to capture the exceedance events (above the WHO limit) using time-leave-out evaluation strategy. The exceedances of NO_2 and O_3 events simulated by the GBT model were compared with ground truth events in each iteration. This allows us to assess the model's ability to reproduce the exceedance events that have not been used in the training process. In total, 82 % of the WHO NO_2 and O_3 exceedance events in the whole data set (ground truth) were correctly identified as WHO NO_2 and O_3 exceedance events (true positives) in both the NO_2 and O_3 GBT models (Table A5). This indicates that our GBT model might slightly underestimates the exceedance events for both NO_2 and O_3. It could be due to unknown drivers that are not included in the model. However, we also noted that 6.6 % and 7.3 % of the data were incorrectly identified as exceedance events (false positives) by our NO_2 and O_3 GBT models, respectively.

The GBT-simulated near-surface NO_2 showed seasonal variations, as expected, with higher values in the winter season (Fig. 5). This is because of the high residential heating demand and favorable meteorology (e.g., a low boundary layer height) for pollutant accumulation and less NO_2 photolysis due to low solar radiation in the winter. The near-surface NO_2 hotspots were the same in all seasons, as seen in the overall study period average. In contrast, near-surface O_3 showed strong seasonal variations, with high values in the spring and summer due to high solar radiation (Fig. 6). It is worth noting that, as seen in the overall study period average, O_3 values in southern Germany were significantly higher in spring and summer than in northern Germany. Because near-surface O_3 is mainly driven by meteorology (DUV and tem-

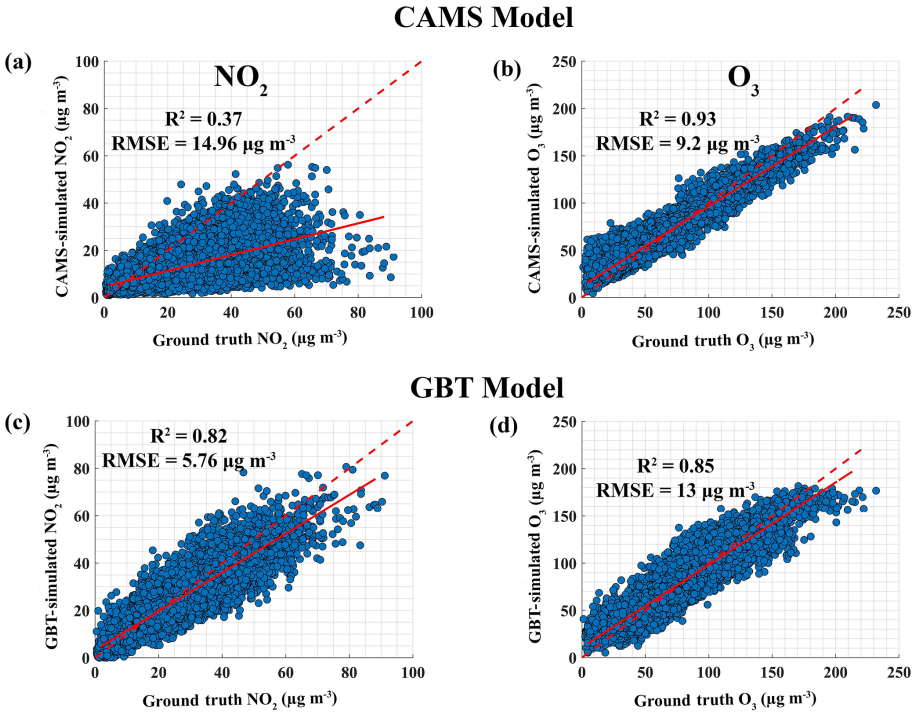

Figure 4. The top panels show the comparison between ground truth near-surface NO_2 and CAMS forecasts of near-surface NO_2 **(a)** and O_3 **(b)** for the period between 17 July 2019 and 31 January 2020. The bottom panels show the comparison between ground truth near-surface NO_2 and GBT-simulated near-surface NO_2 **(c)** and O_3 **(d)** values for the period between 17 July 2019 and 31 January 2020. The dotted line represents a 1 : 1 line, while the solid line represents a linear fit.

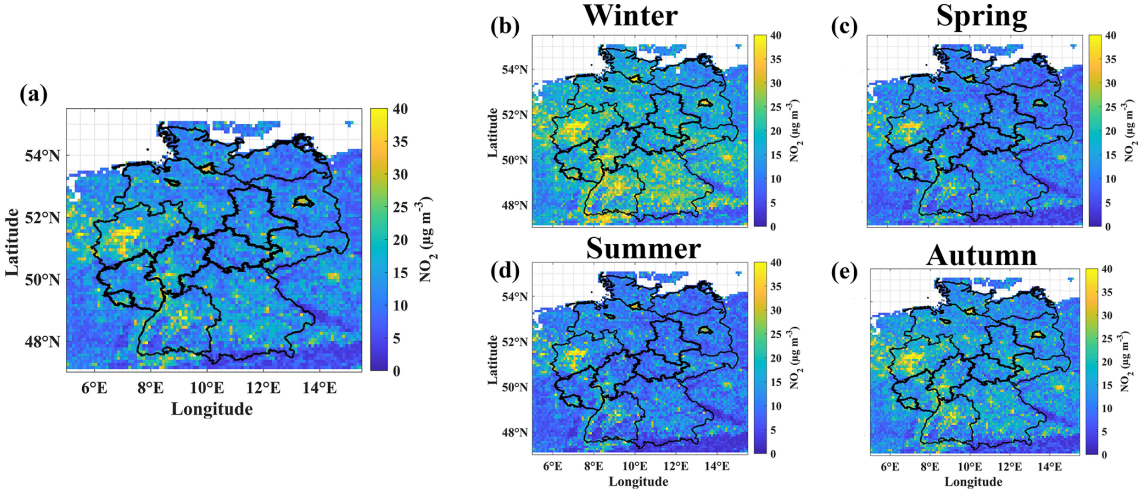

Figure 5. **(a)** Averaged GBT-simulated daily near-surface NO_2 concentrations over the study domain for the study period between 30 April 2018 and 1 July 2021. **(b–e)** Averaged GBT-simulated daily near-surface NO_2 concentrations for each season during the study period. Winter comprises the months of December, January, and February. Spring comprises the months of March, April, and May. Summer comprises the months of June, July, and August. Autumn comprises the months of September, October, and November.

perature, which drive photochemical reactions and precursor emissions), the spatiotemporal variability is attributed to changes in meteorology. We also compared the spatial variability in the GBT-simulated near-surface NO_2 and O_3 to the CAMS forecasts product for the period between 17 July 2019 and 31 January 2020 (Figs. A5 and A6). The spatial variability in the GBT-simulated near-surface NO_2 and O_3 agreed well with the CAMS model. This implies that the ML model can supplement or replace the computationally expensive chemical transport models.

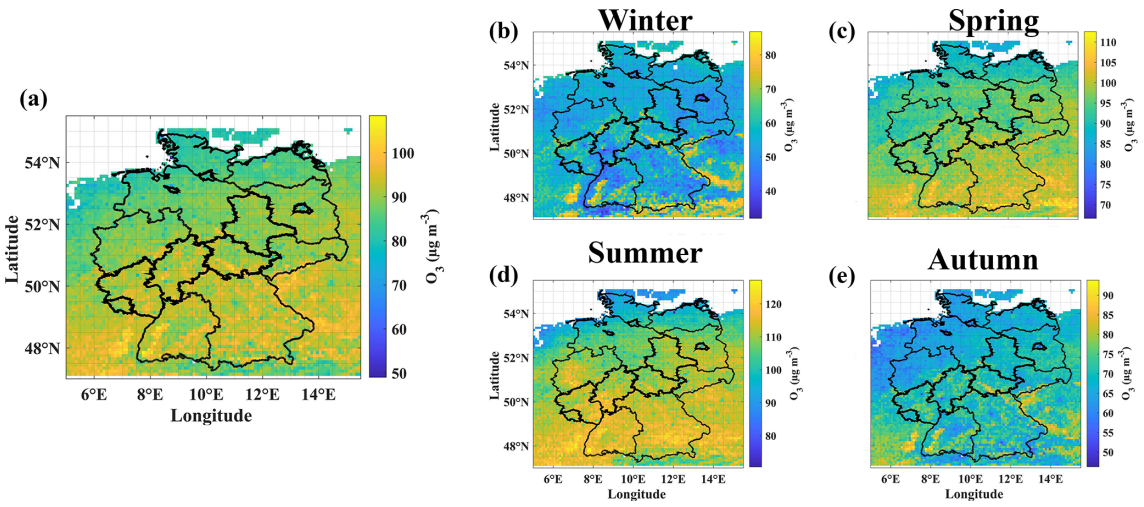

Figure 6. **(a)** Averaged GBT-simulated daily near-surface O_3 concentrations over the study domain for the study period between 30 April 2018 and 1 July 2021. **(b–e)** Averaged GBT-simulated daily near-surface O_3 concentrations for each season during the study period. Winter comprises the months of December, January, and February. Spring comprises the months of March, April, and May. Summer comprises the months of June, July, and August. Autumn comprises the months of September, October, and November.

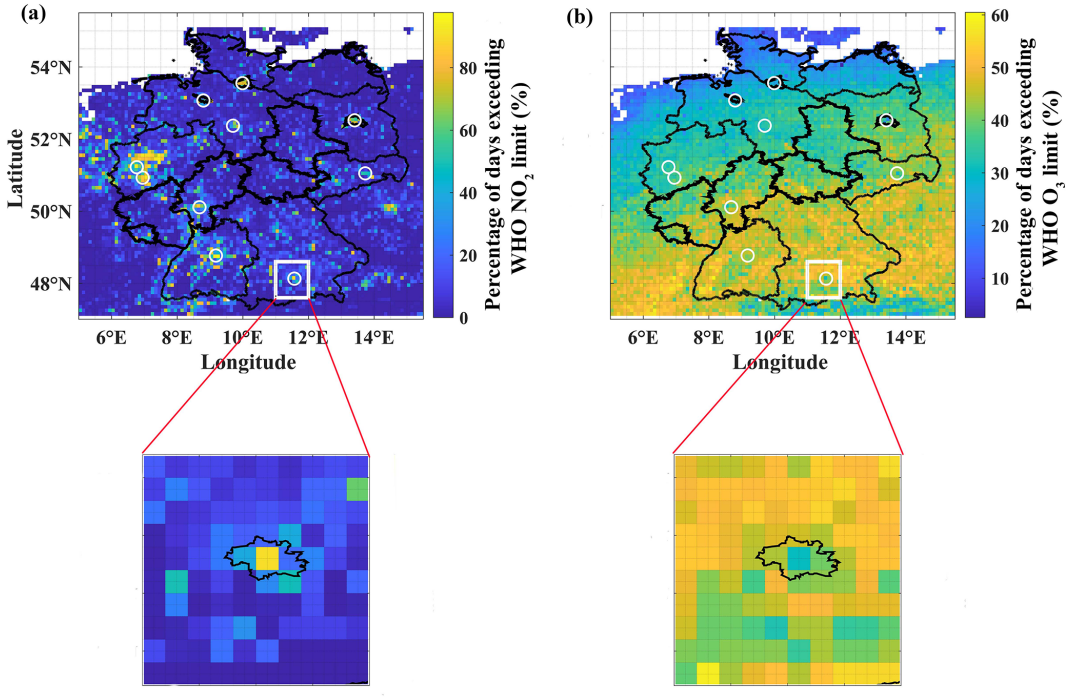

Figure 7. Number of days (%) that exceeded the WHO 24 h mean NO_2 **(a)** and maximum 8 h mean O_3 **(b)** limits over the study domain during the study period based on GBT model simulations. White circles represent major metropolitan areas. The metropolitan area of Munich and its surroundings (rectangular box) are enlarged to illustrate the urban vs. rural gradient. The administrative boundaries of Munich are marked in black in the insets.

3.4 Influence of COVID-19 lockdown restrictions on near-surface NO_2 and O_3 changes

Due to the COVID-19 outbreak, many nations, including Germany, announced a lockdown in the spring of 2020. During that time period, various anthropogenic emission activities were restricted, particularly affecting traffic-related emissions. To estimate the influence of the lockdown restrictions on air pollutant concentration changes, we compared the GBT-simulated 2020 lockdown concentration with the

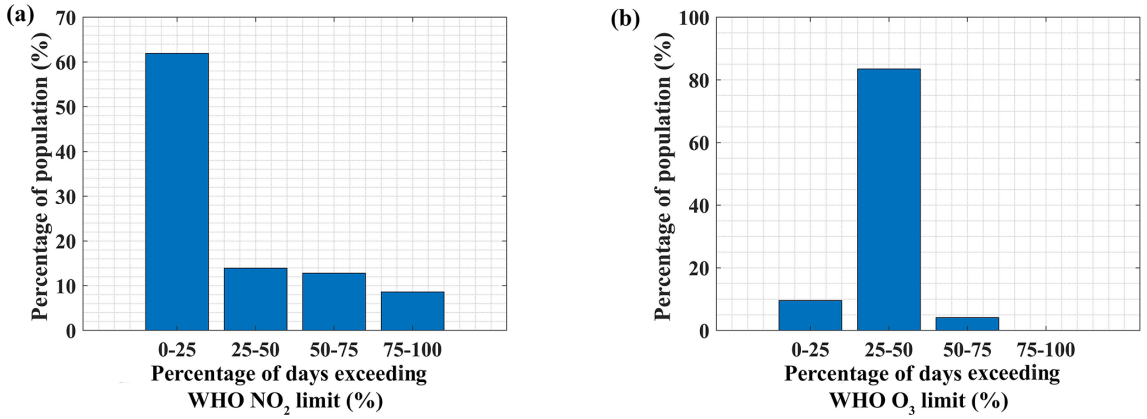

Figure 8. The population distribution in terms of the number of days (%) that exceeded the WHO 24 h mean NO_2 **(a)** and maximum 8 h mean O_3 **(b)** limits over the study domain during the study period based on GBT model simulations.

same period in 2019. The 2020 lockdown period measurements were not used for GBT model training in this comparison. This can also be regarded as the critical performance evaluation of the GBT model.

When comparing different time periods, it is crucial to normalize the meteorology effects when estimating the impact of anthropogenic emission reductions (i.e., lockdown effects) on changes in air pollutant concentrations. Therefore, as described in Sect. 2, we used GC simulations to normalize the meteorology effects from GBT-simulated concentrations. After normalizing the meteorology effects, it is noticeable that high near-surface NO_2 levels decreased primarily over the previously observed hotspots (Fig. 9). The near-surface O_3 increased over western Germany, while decreasing elsewhere, particularly over low NO_x regions. We already observed that western Germany was a NO_x hotspot, possibly due to being a NO_x-saturated regime, so a reduction in NO_x increases ozone. Also, we could see that changes in near-surface O_3 were either negligible or slightly increased over metropolitan areas. The meteorology-normalized mean lockdown near-surface NO_2 decreased by about 23 % (± 5.3 %), while the meteorology-normalized mean lockdown near-surface O_3 increased by 1 % (± 4.6 %), over 10 major metropolitan areas (Berlin, Bremen, Cologne, Dresden, Düsseldorf, Frankfurt, Hamburg, Hanover, Munich, and Stuttgart) when compared to 2019. It increased by about 9 % in the Cologne and Düsseldorf metropolitan areas (located in western Germany) and slightly increased or decreased (between −3 % and +2 %) in other metropolitan areas when compared to 2019. This finding is consistent with other studies that found a decrease in the meteorology-normalized lockdown near-surface NO_2 and the small increase in the lockdown near-surface O_3 over German metropolitan areas when compared to 2019, using in situ measurements (Balamurugan et al., 2021, 2022b). We also evaluated our GBT model's ability to represent different emission scenarios by comparing weekends and weekdays; typically, anthropogenic NO_x emissions on weekends are lower than on weekdays due to reduced vehicle transportation. Our GBT model was also able to distinguish between the weekend and weekday emission scenarios; weekend near-surface NO_2 was lower than weekday near-surface NO_2, and as expected, there were no or only slight changes in weekend near-surface O_3 when compared to weekdays, with slight increases particularly over metropolitan areas (Fig. A7).

3.5 Transferability of our GBT model

Although our study domain also covered parts of other European countries, we trained our GBT model using data from German measurement stations only. Therefore, comparing our trained GBT model simulations with measurements in other countries demonstrates how well our GBT model can model near-surface NO_2 and O_3 concentrations in neighboring parts of the world (similar to the location-leave-out testing strategy). We chose five major cities (Salzburg, Prague, Strasbourg, Liège, and Groningen) in different European countries covered by our study domain and compared their measured NO_2 and O_3 concentrations with GBT-modeled NO_2 and O_3 concentrations (Fig. 10 and Table A6).

Our trained NO_2 GBT model based on German measurement stations explained 32 %–64 % (R^2 ranges between 0.32 and 0.64; RMSE ranges between 9.76 and 13 $\mu g\,m^{-3}$) of the near-surface NO_2 measured in five metropolitan areas located outside of Germany, while the O_3 GBT model simulations agreed well with the observations (R^2 ranges between 0.87 and 0.94; RMSE ranges between 9.55 and 14.32 $\mu g\,m^{-3}$). Since near-surface O_3 is mainly driven by meteorology, the O_3 GBT model trained using German measurement stations explains a large portion of near-surface O_3 in other locations. The worse agreement between the NO_2 GBT model predictions and NO_2 observations in other European countries suggests that information is lacking in the

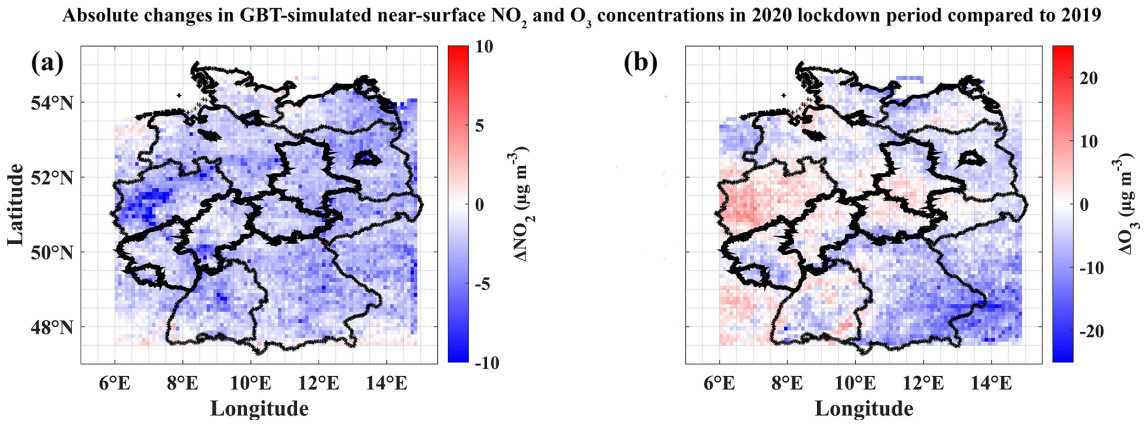

Figure 9. Absolute changes in GBT-simulated near-surface NO_2 and O_3 concentrations in 2020 lockdown period compared to the same period in 2019 after meteorology normalization.

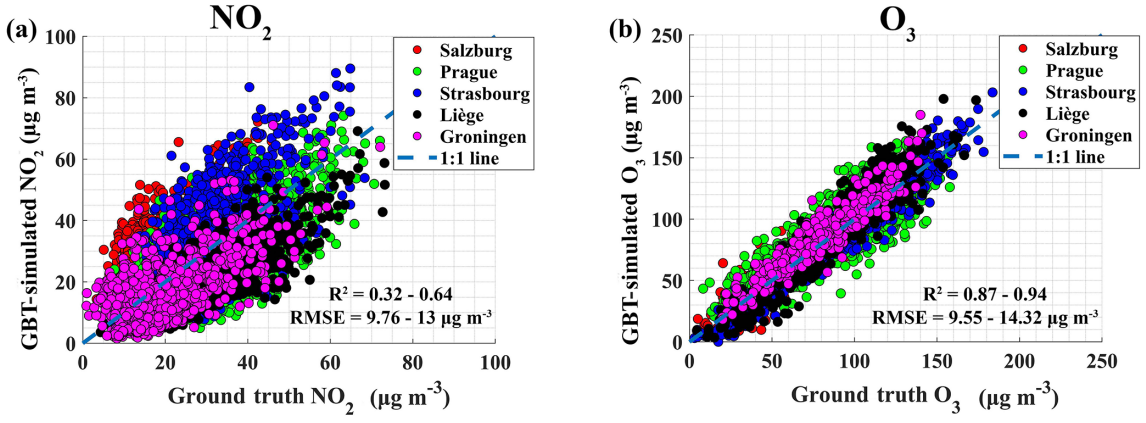

Figure 10. Comparison between ground truth and GBT-simulated near-surface NO_2 **(a)** and O_3 **(b)** for five different European metropolitan areas.

NO_2 GBT model to enable better representations of other locations, similar to location-leave-out 5-fold CV, which also showed poorer agreement for the NO_2 GBT model when modeling new locations (Table 2). Differences in the vehicle fleet composition and emission standards across different countries and locations would have an impact on our NO_2 GBT model predictions when applied to other countries or locations. In future work, other features and proxies besides road density could be considered to represent traffic emission.

4 Conclusion

This study simulated near-surface NO_2 and O_3 concentrations using an ML model over Germany at 0.1° resolution and daily intervals. The ML model was used to link satellite column measurements (proxies for near-surface air pollutants), meteorology, and proxies for emission source information to near-surface NO_2 and O_3 concentrations. The ML models are extremely effective at learning the complex nonlinear relationships between variables. Therefore, in this study, we explored the capabilities of the ML models with respect to the spatiotemporal prediction of air pollutants. In addition, we investigated three aspects of the ML model, namely (1) how well our ML model performs compared to the chemical transport model, (2) how well our ML model can be used to assess the effectiveness of mitigation initiatives, and (3) how well our ML model can be transferred to locations where measurements are unavailable.

The following four different testing strategies were performed to evaluate the ML model's spatiotemporal predictions: (1) random split of data (70 % for training and 30 % for testing); (2) random 5-fold CV; (3) time-leave-out 5-fold CV; and (4) location-leave-out 5-fold CV. The gradient-boosted tree (GBT) model trained for NO_2 explained about 68 %–88 % of the observed NO_2 concentrations in Germany, with RMSE values of 4.77–8.67 $\mu g\,m^{-3}$, whereas the GBT model trained for O_3 performed even better, with an R^2 of 0.74–0.92 and RMSE of 8.53–13.2 $\mu g\,m^{-3}$. The evaluation metrics of the GBT model for different testing strategies differed

significantly. This points out the importance of performing different testing strategies to interpret the true capability of the ML model. The road NO_x emission source proxy (road density) and TROPOMI tropospheric column NO_2 were the most important features in the fitted NO_2 GBT model. For O_3, the most important features were downward UV radiation at the surface and temperature. Since the multi-layer perceptron (MLP) model performed worse than the GBT model, the latter was used in further investigations in our study.

We also showed that our NO_2 GBT model outperforms the CAMS model, while slightly underperforming for near-surface O_3. The CAMS model forecast data set uses real-time observations with an ensemble of 11 air-quality models through data assimilation techniques, which are expected to be more computationally expensive than our GBT model. Therefore, the spatiotemporal variability in the near-surface NO_2 and O_3 concentrations and human exposure at a locations where no measurements are available can be studied with lower computational effort when using our GBT model. Near-surface NO_2 hotspots were found over German metropolitan areas, particularly in western Germany. The near-surface NO_2 hotspot locations did not change with the seasons but had high values in the winter. However, near-surface O_3 showed high seasonal variability, with high values in the spring and summer and no definite hotspots. Overall, southern Germany experiences higher ozone levels than northern Germany due to higher downward UV radiation and temperatures in southern Germany compared to northern Germany. Even though metropolitan areas were the NO_2 hotspots, rural regions, particularly in southern Germany, had higher O_3 concentrations than metropolitan areas. It is because rural areas are dominated by meteorology-driven biogenic VOC emissions and are generally situated downwind of the emission plume. About 36 % of people live in locations where the WHO NO_2 limit exceeds more than 25 % of the days during the study period. Meanwhile, 90 % of people live in areas where the WHO O_3 limit is exceeded for more than 25 % of the study days.

Our study also demonstrated the GBT model's capability to assess the efficacy of mitigation strategies. For example, our GBT model reproduced the observations that, during the 2020 COVID-19 lockdown period, meteorology-normalized near-surface NO_2 was significantly reduced, while meteorology-normalized near-surface O_3 was slightly increased or decreased over metropolitan and industrial areas over Germany when compared to 2019. These findings agreed with those of other studies that used in situ measurements.

Our GBT ML model's transferability is assessed by comparing simulations from our GBT model trained with measurements in Germany to measurements in other European countries. Our NO_2 GBT model showed moderate agreement with observations from other countries (R^2 ranges between 0.32 and 0.64, and RMSE ranges between 9.76 and 13 $\mu g\,m^{-3}$), implying a lack of information in the GBT model when modeling near-surface NO_2 over other countries, which may have different vehicle fleet composition and emissions standards. However, our O_3 GBT model performed well (R^2 ranges between 0.87 and 0.94, and RMSE ranges between 9.55 and 14.32 $\mu g\,m^{-3}$), indicating that our O_3 GBT model can be used to model the O_3 concentrations in other countries, at least in neighboring European countries.

Appendix A

Table A1. Different types of stations (%) considered in this study (based on locations specified by the European Environment Agency).

	Traffic	Industrial	Background
Near-surface NO_2	37.1 %	5.3 %	57.6 %
Near-surface O_3	2.7 %	5.8 %	91.4 %

Table A2. The hyperparameters of the GBT model for each pollutant used in the study.

Hyperparameters	NO_2 model	O_3 model
Max_depth	10	10
Learning_rate	0.3	0.3
reg_lambda	12	4
reg_alpha	18	26
gamma	20	8
min_child_weight	16	8
n_estimators	2500	2500

Table A3. The hyperparameters of the MLP model for each pollutant used in the study.

Hyperparameters	NO_2 model	O_3 model
Hidden_layers (neurons in each layer)	3 (200, 100, 50)	4 (350, 150, 75, 37)
activation	Tanh	Tanh
alpha	0.04	0.1
learning rate	Adaptive	Adaptive
solver	SGD	L-BFGS
Max_iter	2000	1500

L-BFGS is the limited-memory Broyden–Fletcher–Goldfarb–Shanno. SGD is the stochastic gradient descent.

Table A4. Evaluation metrics of our MLP model in different testing strategies.

		Random (70 %/30 %)	Random (5-fold)	Time leave out (5-fold)	Location leave out (5-fold)
NO_2 MLP model	R^2	0.79	0.82 ± 0.006	0.54 ± 0.29	0.46 ± 0.25
	RMSE ($\mu g\,m^{-3}$)	4.77	5.9 ± 0.11	8.6 ± 1.76	13.2 ± 1.07
O_3 MLP model	R^2	0.83	0.9 ± 0.001	0.42 ± 0.37	0.71 ± 0.13
	RMSE ($\mu g\,m^{-3}$)	12.15	9.6 ± 0.027	20.1 ± 7.3	14.9 ± 3.2

Table A5. Comparison between the WHO NO_2 and O_3 exceedance events in the ground truth data set and GBT-simulated WHO NO_2 and O_3 exceedance events using time-leave-out testing strategy.

	Ground truth exceedance	Correct detection as exceedance by NO_2 GBT model (true positives)	Correct detection as exceedance by O_3 GBT model (false positives)
Near-surface NO_2	36 772	30 125	7439
Near-surface O_3	35 860	29 396	6924

Table A6. Metropolitan areas in other European cities considered for the evaluation of GBT model. The evaluation metrics (comparison between GBT simulations and in situ measurements) for NO_2 and O_3 shown in last two columns for each city.

Metropolitan area (country)	Coordinates	R^2 and RMSE ($\mu g\,m^{-3}$) for NO_2	R^2 and RMSE ($\mu g\,m^{-3}$) for O_3
Salzburg (Austria)	47.80° N, 13.05° E	0.32 and 12.52	0.87 and 12.43
Prague (Czech Republic)	50.07° N, 14.43° E	0.43 and 10.05	0.79 and 14.32
Strasbourg (France)	48.57° N, 7.75° E	0.47 and 13	0.94 and 9.55
Liège (Belgium)	50.63° N, 5.56° E	0.64 and 11.9	0.88 and 12.04
Groningen (Netherlands)	53.21° N, 6.56° E	0.34 and 9.76	0.87 and 11.33

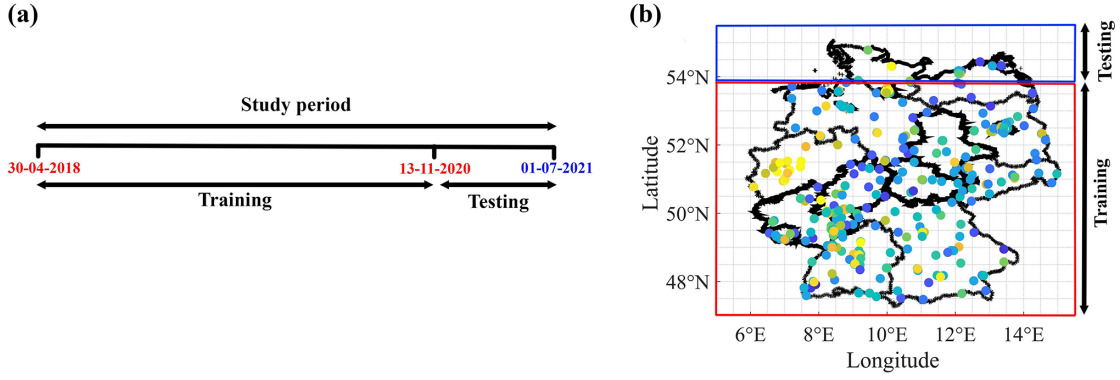

Figure A1. A first 1-fold step in 5-fold CV is illustrated for time-leave-out **(a)** and location-leave-out **(b)** testing strategies. In time-leave-out 5-fold CV, the data were divided into five parts based on the time period (date-wise), with four parts used for training and one part being tested. This process is repeated until each part (a total of five) has been tested. Similarly, in location-leave-out 5-fold CV, the data were divided into five parts based on location (latitude), with four parts used for training and one part being tested. This process is repeated until each part (a total of five) has been tested.

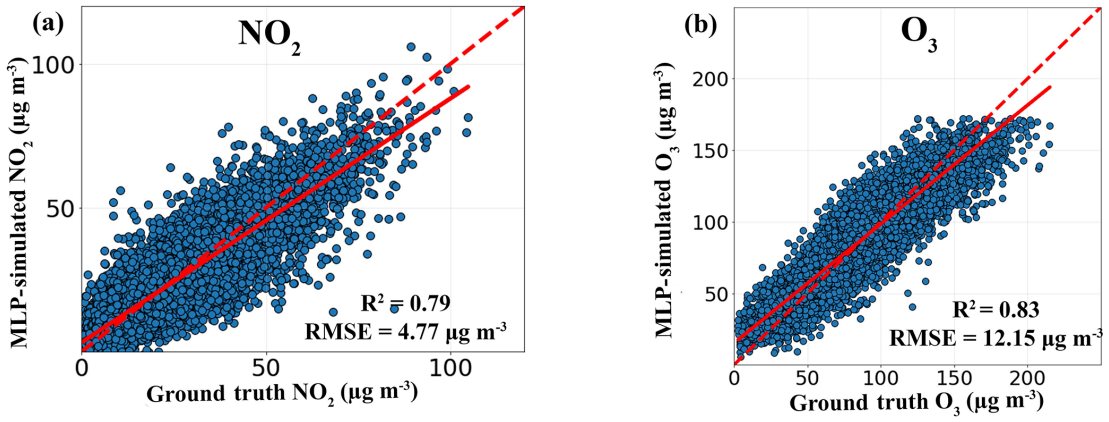

Figure A2. Comparison between ground truth and MLP-simulated near-surface NO_2 **(a)** and O_3 **(b)**. The dotted line represents a 1 : 1 line, while the solid line represents a linear fit.

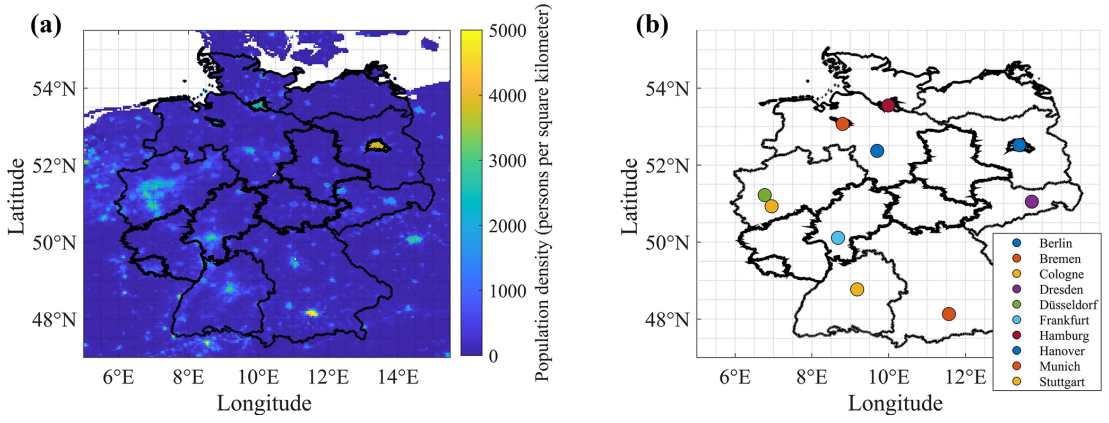

Figure A3. Population density for the year 2020 **(a)** and the locations of major German metropolitan areas **(b)**.

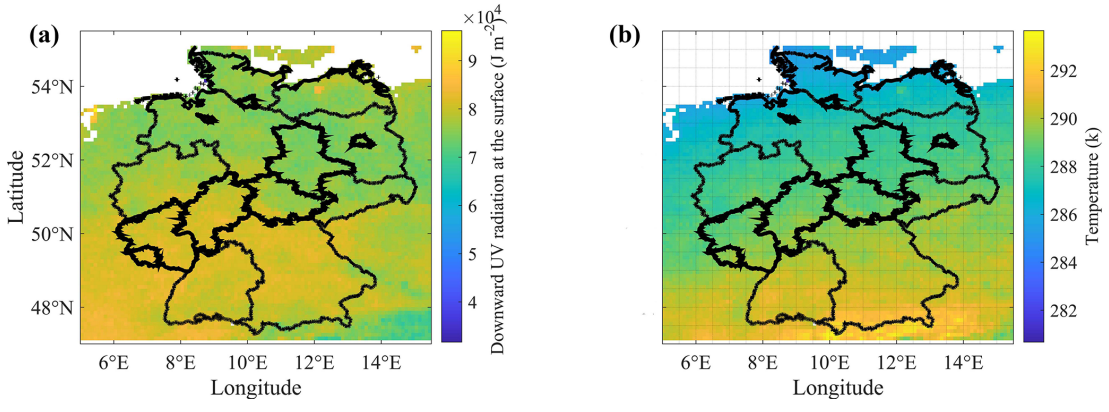

Figure A4. Averaged downward UV radiation at the surface **(a)** and temperature **(b)** over the study domain during the study period.

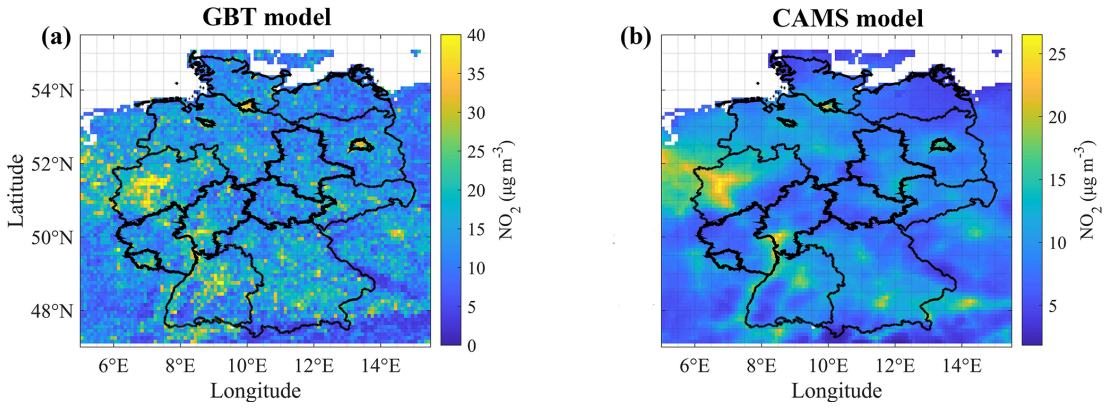

Figure A5. Averaged GBT-simulated near-surface NO_2 concentrations **(a)** and CAMS forecast near-surface NO_2 concentrations **(b)** over the study domain for the period between 17 July 2019 and 31 January 2020.

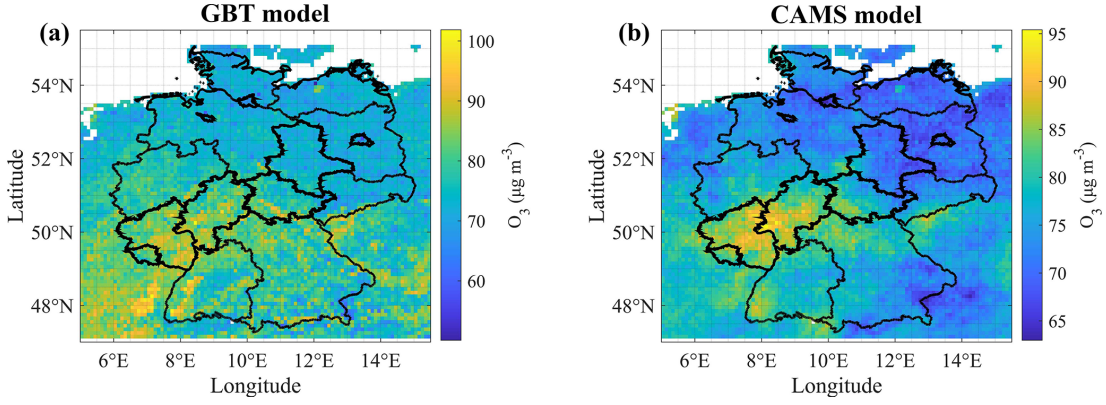

Figure A6. Averaged GBT-simulated near-surface O_3 concentrations **(a)** and CAMS forecast near-surface O_3 concentrations **(b)** over the study domain for the period between 17 July 2019 and 31 January 2020.

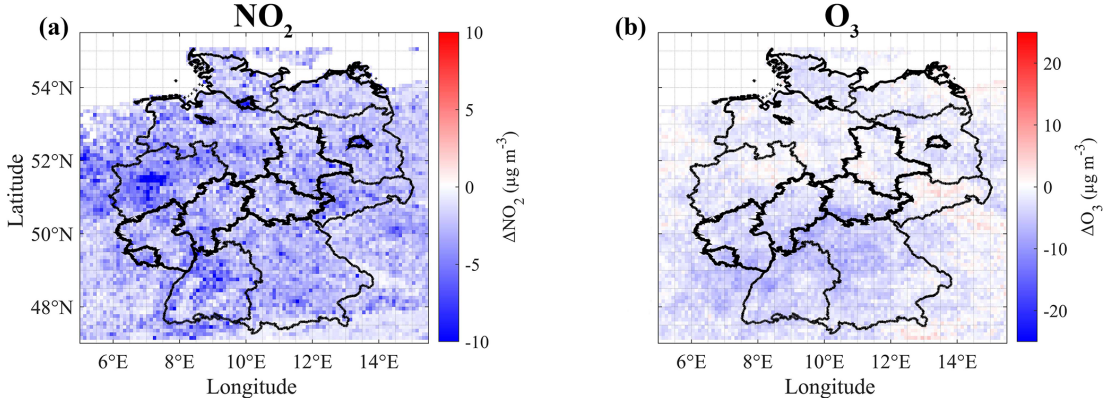

Figure A7. The difference in GBT-simulated near-surface NO_2 **(a)** and O_3 **(b)** concentrations between weekends and weekdays during the study period.

Code and data availability. The code and pre-processed data used to conduct this study are available on Zenodo (https://doi.org/10.5281/zenodo.8330479, Balamurugan et al., 2023).

Author contributions. VB, JC, and FNK conceived the study and designed the concept. VB obtained all of the data, performed the modeling work and analyzed the results. VB developed the machine learning model methodology, with inputs from AW, JC, and FNK. JC and FNK acquired the funding and supervised the work. VB wrote the paper. JC, AW, and FNK reviewed and edited the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Atmospheric Chemistry and Physics*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors thank the European Environment Agency, the Copernicus Services, the Goddard Earth Sciences (GES) Data and Information Services Center (DISC) data archive, and the United States Geological Survey for providing free access to the various data sets used in this study.

Financial support. This research has been funded by the Institute for Advanced Study, Technical University of Munich (grant no. 291763).

This work was supported by the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

Review statement. This paper was edited by Harald Saathoff and reviewed by two anonymous referees.

References

Balamurugan, V., Chen, J., Qu, Z., Bi, X., Gensheimer, J., Shekhar, A., Bhattacharjee, S., and Keutsch, F. N.: Tropospheric NO_2 and O_3 response to COVID-19 lockdown restrictions at the national and urban scales in Germany, J. Geophys. Res.-Atmos., 126, e2021JD035440, https://doi.org/10.1029/2021JD035440, 2021.

Balamurugan, V., Balamurugan, V., and Chen, J.: Importance of ozone precursors information in modelling urban surface ozone variability using machine learning algorithm, Sci. Rep.-UK, 12, 1–8, 2022a.

Balamurugan, V., Chen, J., Qu, Z., Bi, X., and Keutsch, F. N.: Secondary $PM_{2.5}$ decreases significantly less than NO_2 emission reductions during COVID lockdown in Germany, Atmos. Chem. Phys., 22, 7105–7129, https://doi.org/10.5194/acp-22-7105-2022, 2022b.

Balamurgan, V., Chen, J., Wenzel, A., and Keutsch, F. N.: Spatio temporal ML model for NO2 and O3: Initial release, Version V1.0.0, Zenodo [code], https://doi.org/10.5281/zenodo.8330479, 2023.

Bell, J., Power, S. A., Jarraud, N., Agrawal, M., and Davies, C.: The effects of air pollution on urban ecosystems and agriculture, Int. J. Sust. Dev. World, 18, 226–235, 2011.

Chan, K. L., Khorsandi, E., Liu, S., Baier, F., and Valks, P.: Estimation of surface NO_2 concentrations over Germany from TROPOMI satellite observations using a machine learning method, Remote Sens.-Basel, 13, 969, 2021.

Chen, J., Dietrich, F., Maazallahi, H., Forstmaier, A., Winkler, D., Hofmann, M. E. G., Denier van der Gon, H., and Röckmann, T.: Methane emissions from the Munich Oktoberfest, Atmos. Chem. Phys., 20, 3683–3696, https://doi.org/10.5194/acp-20-3683-2020, 2020.

Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, California, San Francisco, USA, 13 August 2016, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Cheng, X., Zhang, W., Wenzel, A., and Chen, J.: Stacked ResNet-LSTM and CORAL model for multi-site air quality prediction, Neural Comput. Appl., 34, 13849–13866, 2022.

Council, N. R.: Rethinking the ozone problem in urban and regional air pollution, The National Academies Press, Washington, DC, https://doi.org/10.17226/1889, 1992.

Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Van Dingenen, R., and Dentener, F.: Contribution and uncertainty of sectorial and regional emissions to regional and global $PM_{2.5}$ health impacts, Atmos. Chem. Phys., 19, 5165–5186, https://doi.org/10.5194/acp-19-5165-2019, 2019.

Crutzen, P. J.: Tropospheric ozone: An overview, Tropospheric ozone: regional and global scale interactions, Springer, 227, 3–32, https://doi.org/10.1007/978-94-009-2913-5_1, 1988.

De Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E. A., Strassmann, A., Puhan, M., Röösli, M., Stafoggia, M., and Kloog, I.: Predicting fine-scale daily NO_2 for 2005–2016 incorporating OMI satellite data across Switzerland, Environ. Sci. Technol., 53, 10279–10287, 2019.

Diao, L., Bi, X., Zhang, W., Liu, B., Wang, X., Li, L., Dai, Q., Zhang, Y., Wu, J., and Feng, Y.: The Characteristics of Heavy Ozone Pollution Episodes and Identification of the Primary Driving Factors Using a Generalized Additive Model (GAM) in an Industrial Megacity of Northern China, Atmosphere-Basel, 12, 1517, 2021.

Forstmaier, A., Chen, J., Dietrich, F., Bettinelli, J., Maazallahi, H., Schneider, C., Winkler, D., Zhao, X., Jones, T., van der Veen, C., Wildmann, N., Makowski, M., Uzun, A., Klappenbach, F., Denier van der Gon, H., Schwietzke, S., and Röckmann, T.: Quantification of methane emissions in Hamburg using a network of FTIR spectrometers and an inverse modeling approach, Atmos. Chem. Phys., 23, 6897–6922, https://doi.org/10.5194/acp-23-6897-2023, 2023.

Gardner, M. W. and Dorling, S.: Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, Atmos. Environ., 32, 2627–2636, 1998.

Gensheimer, J., Chen, J., Turner, A. J., Shekhar, A., Wenzel, A., and Keutsch, F. N.: What Are the Different Measures of Mobility Telling Us About Surface Transportation CO_2 Emissions During the COVID-19 Pandemic?, J. Geophys. Res.-Atmos., 126, e2021JD034664, https://doi.org/10.1029/2021JD034664, 2021.

Ghahremanloo, M., Lops, Y., Choi, Y., and Yeganeh, B.: Deep Learning Estimation of Daily Ground-Level NO_2 Concentrations From Remote Sensing Data, J. Geophys. Res.-Atmos., 126, e2021JD034925, https://doi.org/10.1029/2021JD034925, 2021.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R.: Isoprene and monoterpene emission rate variability: model evaluations and sensitivity analyses, J. Geophys. Res.-Atmos., 98, 12609–12617, 1993.

He, S., Dong, H., Zhang, Z., and Yuan, Y.: An Ensemble Model-Based Estimation of Nitrogen Dioxide in a Southeastern Coastal Region of China, Remote Sens.-Basel, 14, 2807, https://doi.org/10.3390/rs14122807, 2022.

Heaton, J.: Applications of Deep Neural Networks, https://www.heatonresearch.com/book/applications-deep-neural-networks-keras.html (last access: 10 March 2023), 2020.

Hoffmann, B., Boogaard, H., de Nazelle, A., Andersen, Z. J., Abramson, M., Brauer, M., Brunekreef, B., Forastiere, F., Huang, W., Kan, H., Kaufman, J. D., Katsouyanni, K., Krzyzanowski, M., Kuenzli, N., Laden, F., Nieuwenhuijsen, M., Adetoun, M., Powell, P., Rice, M., Roca-Barceló, A., Roscoe, C. J., Soares, A., Straif, K., and Thurston, G.: WHO Air Quality Guidelines 2021 – Aiming for Healthier Air for all: A Joint Statement by Medical, Public Health, Scientific Societies and Patient Representative Organisations, Int. J. Public Health, 6, 1604465, https://doi.org/10.3389/ijph.2021.1604465, 2021.

Hu, C., Kang, P., Jaffe, D. A., Li, C., Zhang, X., Wu, K., and Zhou, M.: Understanding the impact of meteorology on ozone in 334 cities of China, Atmos. Environ., 248, 118221, https://doi.org/10.1016/j.atmosenv.2021.118221, 2021.

Hu, J., Chen, J., Ying, Q., and Zhang, H.: One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system, Atmos. Chem. Phys., 16, 10333–10350, https://doi.org/10.5194/acp-16-10333-2016, 2016.

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

Jacob, D. J.: Introduction to Atmospheric Chemistry, Princeton University Press, ISBN: 9780691001852, 1999.

Jin, X., Fiore, A. M., Murray, L. T., Valin, L. C., Lamsal, L. N., Duncan, B., Folkert Boersma, K., De Smedt, I., Abad, G. G., Chance, K., and Tonnesen, G. S.: Evaluating a space-based indicator of surface ozone-NO_x-VOC sensitivity over midlatitude source regions and application to decadal trends, J. Geophys. Res.-Atmos., 122, 10439–10461, 2017.

Jin, X., Fiore, A., Boersma, K. F., Smedt, I. D., and Valin, L.: Inferring changes in summertime surface Ozone–NO_x-VOC chemistry over US urban areas from two decades of satellite and ground-based observations, Environ. Sci. Technol., 54, 6518–6529, 2020.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.-K., and Kim, S.: Estimation of surface-level NO_2 and O_3 concentrations using TROPOMI data and machine learning over East Asia, Environ. Pollut., 288, 117711, https://doi.org/10.1016/j.envpol.2021.117711, 2021.

Kim, M., Brunner, D., and Kuhlmann, G.: Importance of satellite observations for high-resolution mapping of near-surface NO_2 by machine learning, Remote Sens. Environ., 264, 112573, https://doi.org/10.1016/j.rse.2021.112573, 2021.

Lee, M., Lin, L., Chen, C.-Y., Tsao, Y., Yao, T.-H., Fei, M.-H., and Fang, S.-H.: Forecasting air quality in Taiwan by using machine learning, Sci. Rep.-UK, 10, 4153, 2020.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367–371, 2015.

Li, H., Yang, Y., Jin, J., Wang, H., Li, K., Wang, P., and Liao, H.: Climate-driven deterioration of future ozone pollution in Asia predicted by machine learning with multi-source data, Atmos. Chem. Phys., 23, 1131–1145, https://doi.org/10.5194/acp-23-1131-2023, 2023.

Li, T., Wang, Y., and Yuan, Q.: Remote sensing estimation of regional NO_2 via space-time neural networks, Remote Sens.-Basel, 12, 2514, 2020.

Liang, Y.-C., Maimury, Y., Chen, A. H.-L., and Juarez, J. R. C.: Machine learning-based prediction of air quality, Appl. Sci.-Basel, 10, 9151, 2020.

Lin, X., Trainer, M., and Liu, S.: On the nonlinearity of the tropospheric ozone production, J. Geophys. Res.-Atmos., 93, 15879–15888, 1988.

Liu, Y., Wang, P., Li, Y., Wen, L., and Deng, X.: Air quality prediction models based on meteorological factors and real-time data of industrial waste gas, Sci. Rep.-UK, 12, 9253, 2022.

Lou, S., Liao, H., Yang, Y., and Mu, Q.: Simulation of the interannual variations of tropospheric ozone over China: Roles of variations in meteorological parameters and anthropogenic emissions, Atmos. Environ., 122, 839–851, 2015.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56–67, 2020.

Malashock, D. A., DeLang, M. N., Becker, J. S., Serre, M. L., West, J. J., Chang, K.-L., Cooper, O. R., and Anenberg, S. C.: Estimates of ozone concentrations and attributable mortality in urban, peri-urban and rural areas worldwide in 2019, Environ. Res. Lett., 17, 054023, https://doi.org/10.1088/1748-9326/ac66f3, 2022.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst. Sci. Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., and Nauss, T.: Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation, Environ. Modell. Softw., 101, 1–9, 2018.

Nussbaumer, C. M. and Cohen, R. C.: The role of temperature and NO_x in ozone trends in the Los Angeles basin, Environ. Sci. Technol., 54, 15652–15659, 2020.

Osses, M., Rojas, N., Ibarra, C., Valdebenito, V., Laengle, I., Pantoja, N., Osses, D., Basoa, K., Tolvett, S., Huneeus, N., Gallardo, L., and Gómez, B.: High-resolution spatial-distribution maps of road transport exhaust emissions in Chile, 1990–2020, Earth Syst. Sci. Data, 14, 1359–1376, https://doi.org/10.5194/essd-14-1359-2022, 2022.

Pisoni, E., Albrecht, D., Mara, T. A., Rosati, R., Tarantola, S., and Thunis, P.: Application of uncertainty and sensitivity analysis to the air quality SHERPA modelling tool, Atmos. Environ., 183, 84–93, 2018.

Pusede, S. E. and Cohen, R. C.: On the observed response of ozone to NO_x and VOC reactivity reductions in San Joaquin Valley California 1995–present, Atmos. Chem. Phys., 12, 8323–8339, https://doi.org/10.5194/acp-12-8323-2012, 2012.

Pusede, S. E., Gentner, D. R., Wooldridge, P. J., Browne, E. C., Rollins, A. W., Min, K.-E., Russell, A. R., Thomas, J., Zhang, L., Brune, W. H., Henry, S. B., DiGangi, J. P., Keutsch, F. N., Harrold, S. A., Thornton, J. A., Beaver, M. R., St. Clair, J. M., Wennberg, P. O., Sanders, J., Ren, X., VandenBoer, T. C., Markovic, M. Z., Guha, A., Weber, R., Goldstein, A. H., and Cohen, R. C.: On the temperature dependence of organic reactivity, nitrogen oxides, ozone production, and the impact of emission controls in San Joaquin Valley, California, Atmos. Chem. Phys., 14, 3373–3395, https://doi.org/10.5194/acp-14-3373-2014, 2014.

Qu, Z., Jacob, D. J., Silvern, R. F., Shah, V., Campbell, P. C., Valin, L. C., and Murray, L. T.: US COVID-19 shutdown demonstrates importance of background NO_2 in inferring NO_x emissions from satellite NO_2 observations, Geophys. Res. Lett., 48, e2021GL092783, https://doi.org/10.1029/2021GL092783, 2021.

S5P: HCHO Readme, S5P Mission Performance Centre Formaldehyde [L2 HCHO] Readme, https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Formaldehyde-Readme.pdf (last access: 10 March 2023), 2022a.

S5P: NO_2 Readme, S5P Mission Performance Centre Nitrogen Dioxide [L2 NO_2] Readme, https://sentinel.esa.int/documents/247904/3541451/Sentinel-5P-Nitrogen-Dioxide-Level-2-Product-Readme-File (last access: 10 March 2023), 2022b.

S5P: O_3 Readme, S5P Mission Performance Centre Readme OFFL Total Ozone, https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Readme-OFFL-Total-Ozone.pdf (last access: 10 March 2023), 2022c.

Sicard, P., Paoletti, E., Agathokleous, E., Araminienė, V., Proietti, C., Coulibaly, F., and De Marco, A.: Ozone weekend effect in cities: Deep insights for urban air pollution control, Environ. Res., 191, 110193, https://doi.org/10.1016/j.envres.2020.110193, 2020.

Sillman, S.: The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments, Atmos. Environ., 33, 1821–1845, 1999.

Sillman, S., Logan, J. A., and Wofsy, S. C.: The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes, J. Geophys. Res.-Atmos., 95, 1837–1851, 1990.

Singh, J., Singh, N., Ojha, N., Sharma, A., Pozzer, A., Kiran Kumar, N., Rajeev, K., Gunthe, S. S., and Kotamarthi, V. R.: Effects of spatial resolution on WRF v3.8.1 simulated meteorology over the central Himalaya, Geosci. Model Dev., 14, 1427–1443, https://doi.org/10.5194/gmd-14-1427-2021, 2021.

Trombetti, M., Thunis, P., Bessagnet, B., Clappier, A., Couvidat, F., Guevara, M., Kuenen, J., and López-Aparicio, S.: Spatial intercomparison of Top-down emission inventories in European urban areas, Atmos. Environ., 173, 142–156, 2018.

Vlasenko, A., Matthias, V., and Callies, U.: Simulation of chemical transport model estimates by means of a neural network using meteorological data, Atmos. Environ., 254, 118236, https://doi.org/10.1016/j.atmosenv.2021.118236, 2021.

Wang, W., van der A, R., Ding, J., van Weele, M., and Cheng, T.: Spatial and temporal changes of the ozone sensitivity in China based on satellite and ground-based observations, Atmos. Chem. Phys., 21, 7253–7269, https://doi.org/10.5194/acp-21-7253-2021, 2021.

Xie, X., Wang, T., Yue, X., Li, S., Zhuang, B., Wang, M., and Yang, X.: Numerical modeling of ozone damage to plants and its effects on atmospheric CO_2 in China, Atmos. Environ., 217, 116970, https://doi.org/10.1016/j.atmosenv.2019.116970, 2019.

Zaini, N., Ean, L. W., Ahmed, A. N., Abdul Malek, M., and Chow, M. F.: $PM_{2.5}$ forecasting for an urban area based on deep learning and decomposition method, Sci. Rep.-UK, 12, 17565, https://doi.org/10.1038/s41598-022-21769-1, 2022.

Zhang, J., Chen, Q., Wang, Q., Ding, Z., Sun, H., and Xu, Y.: The acute health effects of ozone and $PM_{2.5}$ on daily cardiovascular disease mortality: A multi-center time series study in China, Ecotox. Environ. Safe., 174, 218–223, 2019.

Zhao, Z., Wu, J., Cai, F., Zhang, S., and Wang, Y.-G.: A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the COVID-19 pandemic, Sci. Rep.-UK, 13, 1015, 2023.

Zhu, Q., Bi, J., Liu, X., Li, S., Wang, W., Zhao, Y., and Liu, Y.: Satellite-Based Long-Term Spatiotemporal Patterns of Surface Ozone Concentrations in China: 2005–2019, Environ. Health Persp., 130, 027004, https://doi.org/10.1289/EHP9406, 2022.

Zong, R., Yang, X., Wen, L., Xu, C., Zhu, Y., Chen, T., Yao, L., Wang, L., Zhang, J., Yang, L., Wang, X., Shao, M., Tong, Z., Xue, L., and Wang, W.: Strong ozone production at a rural site in the North China Plain: Mixed effects of urban plumesand biogenic emissions, J. Environ. Sci., 71, 261–270, 2018.