Environment International 142 (2020) 105856

Contents lists available at ScienceDirect

Environment International

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

Evaluation of different methods and data sources to optimise modelling of NO_2 at a global scale

Meng Lu[a,*], Oliver Schmitz[a], Kees de Hoogh[b,c], Qin Kai[d], Derek Karssenberg[a]

[a] Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[b] Swiss Tropical and Public Health Institute, Basel, Switzerland
[c] University of Basel, Basel, Switzerland
[d] China University of Mining and Technology, Xuzhou, China

ARTICLE INFO

Handling editor: Xavier Querol

Keywords:
Air pollution
Global scale
High resolution
Statistical learning
Temporal
TROPOMI

ABSTRACT

Background: In countries where air pollution stations are unavailable or scarce, station measurements from other countries and atmospheric remote sensing could jointly provide information to estimate ambient air quality at a sufficiently fine resolution to study the relationship between air pollution exposure and health. Predicting NO_2 concentration globally with sufficient spatial and temporal resolution and accuracy for health studies is, however, not a trivial task. Challenges are data deficiency, in terms of NO_2 measurements and NO_2 predictors, and the development of a statistical model that can typify the regional and continental differences, such as traffic regulations, energy sources, and local weather.

Objective: We investigated the feasibility of mapping daytime and nighttime NO_2 globally at a high spatial resolution (25 m), by including TROPOMI (TROPOspheric Monitoring Instrument) data and comparing various statistical learning techniques.

Method: We separated daytime (7:00 am - 9:59 pm) and nighttime (10:00 pm - 6:59 am) based on the local times. To study if one should build models for each country separately, national models in 4 selected countries (the US, China, Germany, Spain) were developed. We build the models for 2017 and used 3636 stations. Seven statistical learning techniques were applied and the impact of the predictors, model fitting, and predicting accuracy was compared between different techniques, national models, national and global models, and models with and without including the NO_2 vertical column density retrieved from TROPOMI.

Result and conclusion: The ensemble tree-based methods obtained higher accuracy compared to the linear regression-based methods in national and global models. The global tree-based methods obtained similar accuracy to national models. Different spatial prediction patterns are observed even when the prediction accuracy is very similar. Separating between day and night can be important for more accurate air pollution exposure assessment. The TROPOMI variable is ranked as one of the most important variables in the statistical learning techniques but adding it to global models that contain other precedent remote sensing products does not improve the prediction accuracy.

1. Introduction

NO_2 is an important risk factor for respiratory (Chauhan et al., 1998) and cardiovascular diseases (Collart et al., 2018). Global NO_2 maps are required to estimate the global burden of disease (Burnett et al., 2018), which has been conducted at country level (Chen et al., 2010), and, more recently, at the neighbourhood level (Anenberg et al., 2018; Achakulwisut et al., 2019). In addition, modelling air pollution at the global scale enables consistent comparisons of relations between air pollution and health in different countries using a universal air pollution dataset.

As NO_2 is highly traffic-related and localised, mapping it at a high resolution in space and time is needed for the assessment of personal outdoor exposure, in particular in areas close to sources of the pollutant such as primary roads. In addition, high resolution maps are a requirement for the incorporation of human spatiotemporal activity patterns in assessment of personal exposures (Lu et al., 2019; Park and Kwan, 2017) and for calculating address-based health outcomes (home or hospital locations) (Morgenstern et al., 2007).

NO_2 can be modelled at high resolution using dispersion models

* Corresponding author.
E-mail address: m.lu@uu.nl (M. Lu).

https://doi.org/10.1016/j.envint.2020.105856
Received 28 November 2019; Received in revised form 16 April 2020; Accepted 2 June 2020
Available online 25 June 2020

(Holmes and Morawska, 2006; Institute, 2010), statistical models (Chen et al., 2019a), or a combination of both (Mölter et al., 2010). Dispersion models simulate the emission, transforming, transportation, and deposition of the pollutant, but require detailed emission inventory data, which is not available globally. Statistical models utilise geospatial predictors that are related to the emission sources (e.g. road network) and dispersion processes (e.g. meteorological data) of the pollutants (Briggs et al., 2000), which are available globally and thus can provide an estimation of global NO_2 at a high spatial resolution (Larkin et al., 2017).

Statistical models for air pollution mapping are under continuous development. Hoek et al. (2008) evaluates 25 Land Use Regression (LUR) studies, with a focus on national and regional models. The Kriging method and Bayesian models were combined to fully assess the uncertainty (Adam-Poupart et al., 2014). Zhan et al. (2018) integrates spatial correlations of residuals from a random forest model. Dispersion models have been integrated into LUR models to account for the mechanisms of pollutant dispersal (Marshall et al., 2008; Beelen et al., 2010; Dijkema et al., 2010; Akita et al., 2014). Statistical learning techniques have been shown to obtain higher (Kerckhoffs et al., 2019) or similar (Chen et al., 2019a) prediction accuracy compared to LUR. Wang et al. (2013) and Geddes et al. (2016) evaluated temporal stability of spatial structures. Chen et al. (2010) evaluated various temporal aggregations in air pollution modelling. Eeftens et al. (2012) and Vienneau et al. (2010) developed LUR models for multiple countries. Hoek et al. (2015) showed that using the column density of satellite measured NO_2 could increase the prediction accuracy of LUR models.

Most statistical air pollution models are developed at intra-urban or urban scales. Relatively few studies have investigated large-scale air pollution mapping. Table 1 summarises NO_2 studies at country and larger scales. The study in China by Zhan et al. (2018) focuses on integrating remote sensing measurements at a relatively coarse resolution (0.1 degrees). De Hoogh et al. (2018) develops hybrid LUR models for west European countries and used Kriging to model LUR model residuals. The stability of the spatial structure of models over time is assessed by comparing with models developed separately for each year.

At the global scale, data deficiency and heterogeneity of air pollution emitting sources are two challenges. As will be shown in our study, the same road length within a certain buffer may lead to different concentration levels between countries, possibly caused by different fuel and car types and the filter system used in cars. For instance, in Brazil, ethanol fuels are prevalent due to its agricultural-industrial technology and massive arable lands (Sperling and Gordon, 2010). The combustion of ethanol does not produce NO_2, which may result in different road effects of NO_2 compared to other countries. In addition, street characteristics such as the density of buildings and trees, and the oxidant capacity of the local atmosphere, e.g., the amount of O_3 (Han et al., 2011), also affect the NO_2 distribution. Besides the sparsity of monitoring stations in many areas, another data constraint for large-scale spatial air pollution prediction is the sometimes limited availability of air pollution predictors, such as the road traffic intensity. Reliability and the resolution or detail of the predictors may decrease when considering larger areas. The heterogeneity of pollutant-emitting sources and dispersing pathways from region to region call for a statistical model that can characterise the differences without loss of generality. Thus far, the global model study of Larkin et al. (2017) used the highest number of NO_2 measurement stations compared to other large-scale studies listed in Table 1, and has been the only NO_2 global model at a fine resolution (100 m). A Lasso regression was used in combination with predefined rules to select the predictor variables in multiple linear regression (LM) model to fit the regression coefficients. The global model is well-validated and is used in current global health studies (Achakulwisut et al., 2019).

Table 1
Large-scale statistical NO_2 modelling studies. "Stations" indicates number of sensor locations used.

Area	Resolution	Stations	Reference
US	100 m	68	Novotny et al. (2011)
US	25 km	423	Young et al. (2016)
Canada	0.1 degree	134	Hystad et al. (2011)
Australia	100 m	68	Knibbs et al. (2014)
China	1 degree	744–1604	Zhan et al. (2018)
West European countries	100 m	1426	de Hoogh et al. (2016a)
West European countries	100 m	2399	De Hoogh et al. (2018)
Global	100 m	5220	Larkin et al. (2017)

The studies discussed above are valuable for large-scale air pollution mapping. However, it is unknown how different statistical methods and spatial stratification techniques for NO_2 prediction can address the problems of ground measurement data deficiency and the heterogeneous air pollutant-predictor relationships at the global scale. Most of the studies selected a single technique and study area and did not evaluate how various techniques perform at different scales and in different areas. In addition, it may be worthwhile separating between day and night NO_2 models for three reasons: 1) temporal air pollution modelling is expected to lead to improved exposure assessment as it enables integration of exposure over multiple space–time activities, 2) the statistical relationships between predictors and NO_2 may be better modelled by separating day and night, and 3) the photochemical oxidation reactions that determine the NO_2 levels are different between day and night. Lastly, the TROPOMI (Tropospheric monitoring instrument, NSO and ESA, 2019) onboard the Sentinel 5p satellite provides the column density of NO_2 and other gaseous pollutants with greatly improved resolution. It remains unclear how can the retrievals from TROPOMI (called TROPOMI measurements) contribute to high-resolution global NO_2 mapping.

This study contributes to global high-resolution NO_2 mapping and added in a temporal component by (1) comparing various linear regression-based and ensemble tree-based (e.g. Random Forest, Breiman, 2001) techniques, to investigate if more flexible models can better capture non-linear relationships between predictors and NO_2, (2) comparing global models with national models, to understand the effects of the heterogeneity between countries, and (3) separating between day and night. While this study has a focus on predicting the spatial variation of long-term NO_2 concentrations, the successful prediction of long-term day- and night NO2 merits further research in even finer temporal resolutions, for example during rush-hour. We evaluate the TROPOMI measurements to understand the role of it in current model configurations.

We selected statistical methods to evaluate two classes of statistical learning methods: regression-based methods which fit one model to the entire range of each predictor, and ensemble tree-based methods which build on subsets of data and sub-ranges of predictors. These methods are representative for the techniques that are evaluated in the most recent air pollution modelling (Chen et al., 2019a; Kerckhoffs et al., 2019). We developed models globally and in four countries, the US (United States of America), Germany, China, and Spain. These countries were selected because they (1) have different industrial development and environmental policies and may, therefore, have different distributions of NO_2 monitoring networks, (2) represent different climates and pollution levels across the globe, and (3) have monitoring networks sufficiently large to apply statistical learning techniques.

2. Materials and methods

We define daytime in our study as the time span between 7:00 am and 9:59 pm (local time) and nighttime as the remaining hours, based on the amount of time a person on average stays indoor and traffic patterns of several countries (Brasche and Bischof, 2005; Schweizer et al., 2007; Kwon et al., 2008). The separation between the daytime and nighttime is not based on the amount of daylight because the exact

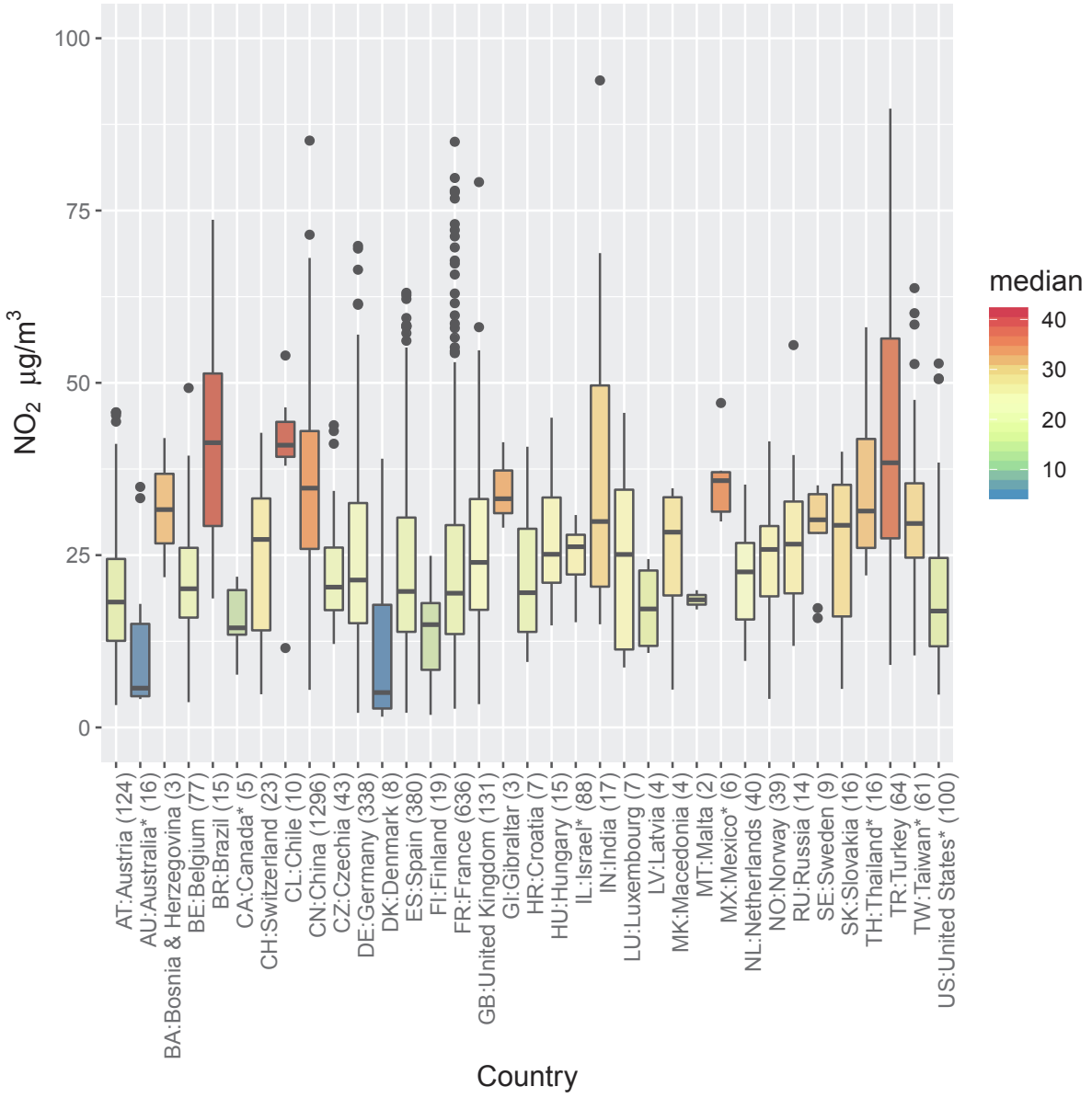

Fig. 1. Annual mean NO_2 concentrations over all stations within a country. The star behind the country name indicates that NO_2 is measured in *ppm* and converted to $\mu g/m^3$. The number of ground stations used in the study is shown between brackets. The dots indicate extreme values and the line in the centre of each box gives the median.

relationship between human space–time activity and daylight is unknown.

2.1. Air quality measurement data

The station measurements are obtained from the OpenAQ (2019) open science community which gathers air pollution measurements world-wide. The official Chinese air quality data was not available on OpenAQ and was obtained from the Chinese environmental institute (CNEMC, 2019). As most of the OpenAQ measurements come from official national ground monitor networks, we considered the quality of the data from OpenAQ and CNEMC (2019) the same. A limitation is that OpenAQ does not provide information about instrument types, which is discussed in Section 4. We used data from 2017, as in this year the number of data included in OpenAQ was considerably increased. All of the stations used have hourly measurements. We removed stations in Andorra, Argentina, Serbia, and Poland due to insufficient temporal coverage (less than 6 months) in 2017. Fig. 1 shows summary statistics of the data set used for each country that are used in our global model. Table 2 presents summary statistics of the NO_2 measurements and the number of stations in the US, China, Germany, Spain, and globally. The temporal data are aggregated at 3 levels: annual mean, annual mean over daytime, and annual mean over nighttime. The geographical distribution of the stations is shown in a digital supplement ("oaqmean-NO2.html"). Concentration levels for countries that measure NO_2 in *ppm* (parts per million) were converted to $\mu g/m^3$ (micrograms per cubic meter air) using $NO_2(\mu g/m^3) = NO_2(ppm)$* 1.91 * 1000, assuming an average temperature of 20 °C and pressure of 1013 mb. The coefficient, 1.91, is subject to an error range of 2% - 7% over a range of temperatures between 0 °C and 25 °C.

2.2. Predictor variables

The candidate predictors were calculated at 25 m resolution. They are either spatial attributes aggregated within a circular ring centred at each sensor or prediction location, called buffered predictors, or values

Table 2
Summary statistics of the 2017 annual mean, daytime mean, and nighttime mean NO_2 ($\mu g/m^3$) of ground station measurements in the US, China, Germany, Spain, and over the globe. Q.: Quartile.

		min.	1st Q.	median	3rd Q.	max.	ground sensors (n)
US	Mean	5.68	11.77	16.89	24.59	52.81	100
	Day	5.02	10.50	16.03	22.94	51.11	
	Night	6.17	14.08	19.92	27.45	58.15	
China	Mean	11.44	25.60	34.58	42.93	85.14	1, 296
	Day	13.46	30.84	40.44	49.72	88.71	
	Night	9.40	19.49	25.69	32.28	78.24	
Germany	Mean	4.53	15.12	21.38	32.56	69.87	338
	Day	4.25	15.11	21.75	35.50	84.14	
	Night	4.98	14.64	20.21	27.55	56.91	
Spain	Mean	3.25	13.86	19.73	30.45	63.07	380
	Day	3.34	13.81	20.59	30.51	67.53	
	Night	2.85	12.29	18.47	29.34	69.33	
Global	Mean	4.42	16.80	25.25	36.19	93.88	3, 636
	Day	4.25	17.66	27.43	40.81	94.26	
	Night	4.10	15.02	21.94	29.48	100.58	

of the spatial attribute at the observation or prediction location (Table 3 and 4).

For buffered predictors, roads and industrial areas were extracted from OpenStreetMap data (OpenStreetMap contributors, 2019) using 'highway' and 'landuse' key values. The road types we refer to in this paper are the same as the road types classified and tagged in OpenStreetMap. For the main road types, the link roads were also included. The secondary, tertiary, and residential roads were combined to a new road type called 'local roads'.

We first created raster maps with the total road lengths (in metres) and industrial areas (in m^2) for each 25 m cell, and afterwards aggregated these values using various buffer sizes as follows: for each 10×10 km area where station data was available, we projected the OpenStreetMap data from WGS84 to Lambert Azimuthal Equal-Area (LAEA) using the centroid of the 10×10 km area as the projection centre. We then assigned 25 m grids to each of the LAEA areas. For industrial areas, we used these grids to assign the land use class to cell values of raster maps. For road lengths, we intersected each road type with the 25 m grids, calculated the length of the resulting road segments in each grid cell, summed up the lengths for each cell and assigned those values to raster maps. For population, we used grid data of the year 2015 from the Global Human Settlement Layer population grid (JRC, 2015). Finally, we used the raster maps to calculate aggregated values using various buffer radii (see Table 3). The processing of the buffered candidate predictors was done using Python (2019), PCRaster (Karssenberg et al., 2010) and GDAL (GDAL Development Team, 2018).

As gridded predictors, i.e. predictors not aggregated over a buffer area but directly retrieved at the location of the sensor, we used monthly average temperature at 2 m height and wind speed at 10 m height of the year 2017 from ERA-Interim (Dee et al., 2011). The elevation was extracted from the ETOPO1 1 arc-minute global relief dataset (Amante and Eakins, 2009). Two satellite NO_2 vertical column density products were used: the OMI (Ozone Monitoring Instrument) 2017 annual average which is preprocessed to the level 3 (gridded) product Earthdata (2019) and the monthly mean tropospheric NO_2 from TROPOMI from Feb 2018 (the earliest data available) to Jan 2019 (TEMIS, 2019b). For both OMI and TROPOMI, negative values and values larger than 20×10^{15} (mol/cm^2) were treated as missing (TEMIS, 2019a). A GEOS-CHEM (Bey et al., 2001; GEOS-CHEM, 2019) annual NO_2 surface concentration product (Geddes et al., 2016) was included in the modelling. The product (Geddes et al., 2016) is only available till 2012 and is based on coarser-resolution products from remote sensing instruments SCIAMACHY and GOME-2. However, we anticipate that a surface concentration product may relate more directly to NO_2 measured by ground stations compared to the column densities retrieved from TROPOMI and OMI. The TROPOMI data (2018–2019) and GEOS-CHEM concentration product (2012) are from different years as the modelling period (2017); however, we expect that the spatial pattern of NO_2 concentrations is quite stable over these years (De Hoogh et al., 2018) and therefore these two predictors may improve the mapping of NO_2.

The variables are grouped into road length, population, industrial, elevation, meteorological, and remote sensing products. In addition, buffered variables with radii of 25–300 m are referred to as emission-related variables, and the remaining as background variables. The scatterplots of predictor variables and NO_2 concentrations are shown in the supplement (SF.1 - SF.4).

Table 3
Buffered predictors, "_buf" indicates the buffer radius. The buffered predictors with buffer radii of 25 m, 50 m, 100 m, 300 m, 500 m, 800 m, 1000 m, 3000 m, 5000 m are calculated.

Candidate predictor	Variable name	Unit	Source
Total length of highway	road_1_buf	m	polygon, lineString
Total length of primary roads	road_2_buf	m	polygon, lineString
Total length of local roads	road_M345_buf	m	polygon, lineString
Area of industry	I_1_buf	m^2	polygon, lineString

2.3. Statistical learning methods

Two different types of statistical learning methods are tested to explain the spatial variability of NO_2. The first group consists of linear regression model-based approaches, including LM (multiple linear regression), Lasso (Least Absolute Shrinkage and Selection Operator) regression, Ridge regression, and ElasticNet (Zou and Hastie, 2005). The other group consists of ensemble tree-based approaches (called tree-based methods), including Random Forest (RF, Breiman, 2001), Stochastic Gradient Boosting (SGB, Friedman, 2002), and Extreme (also called regularised) Gradient Boosting (XGB, Chen and Guestrin, 2016). The equations and algorithms of these methods are given in the supplement. Table 5 lists the features of these methods and their implementations.

Among the statistical model-based methods, ElasticNet, Lasso, and Ridge are regularised to penalise complex models. The L2 norm penalty that Lasso uses allows for shrinking the variable coefficients to zero, which enables "variable selection". This feature of Lasso makes it a popular method in LUR modelling (Larkin et al., 2017). Ridge regression does not select variables but may predict faster and more accurately (James et al., 2013). ElasticNet is theoretically the most flexible in fitting, in addition to being able to reduce variable space, the penalty of it can be tuned by cross-validation.

The tree-based methods are robust to the inclusion of less relevant features and can represent non-linear relationships (Friedman et al., 2010a). RF grows each tree independently, which may be less prone to over-fitting and is relatively easy to set up. XGB and SGB grow trees subsequently, are more likely to model artefacts when the training set is relatively small. XGB attempts to improve from SGB using a regularization term to penalize complex models. However, this means that XGB may require an even larger set of training samples to achieve its full performance. Also, it is the most difficult to obtain the optimal set up due to the larger number of hyperparameters.

2.3.1. Model setting, hyperparameter tuning, implementation

The hyperparameters of ElasticNet and ensemble tree-based methods were set before the training process. For ElasticNet, the regularization term (`alpha`) is set to 0.2. We tuned the hyperparameters of tree-based methods using 10-fold cross-validation (with the R `caret`

(Kuhn, 2018) and gbm (Greenwell et al., 2019) package), and for SGB and XGB with additional manual tuning. All the records of annual mean NO_2 measurements are used for hyperparameter tuning. The tuning parameters for RF are the number of trees (ntrees) and the number of variables sampled (mtry) for each tree. Among them, the RF is more sensitive to mtry. A large mtry may not contribute to the reduction of model variation, while a small mtry may miss important variables and may not capture the interactive effects of predictors to the response. In contrast, after reaching an optimal ntree value, further increasing the ntree may not affect model performance, only add computations. Based on the cross-validation results, we finally set the ntree to 2000 and the mtry to 33.

For SGB, the parameters we tuned are the learning rate (shrinkage), the maximum depth of the trees (maximum.depth) and the ntree. Among the three, SGB is the most sensitive to shrinkage and maximum.depth, a high value of shrinkage or maximum.depth may improve the model performance, but they both depend on sufficient training samples and computation power. Compared to RF, SGB may be more sensitive to large ntree values; as the trees grow subsequently, a high ntree value may make the model prone to overfitting. The value of ntree was tuned using cross-validation. The shrinkage and the maximum.depth were then tuned manually with the optimum ntree until the training RMSE (root mean squared error) remained stable. The ntree was set to 2000, maximum.depth to 6 and the shrinkage to 0.01.

For XGB, we tuned the shrinkage, maximum.depth, and the ntrees. The sensitivity of XGB to the shrinkage may be lower than the SGB and higher for the ntrees. The cross-validation suggests using 0.4 for the shrinkage and 140 for the ntrees; however, we found that with these parameter settings the model did not converge with some bootstrapped samples. These two parameters were therefore tuned manually until the training RMSE remained stable. This resulted in the eta being 0.02, maximum.depth being 4, and ntrees being 1000. To make it more comparable to the SGB, the ntrees was set to 2000.

2.3.2. Accuracy assessment

As an accuracy indicator, the RMSE (Root Mean Squared Error) provides general insight into the variance and magnitude of the error. In addition, we calculated the MAE (Mean Absolute Error) for the magnitude of the error and the IQR (Inter-Quartile Range) for the variance of the error. To make the accuracy assessed at different study areas and between day and night comparable, we calculated the RRMSE (relative RMSE), RMAE (relative MAE), rIQR (relative IQR), and the R-squared (R^2). The RRMSE was calculated by dividing the RMSE by the mean of observations. The rMAE was calculated by dividing the MAE by the mean of observations and the rIQR was calculated by dividing the IQR by the median of observations.

We used 80% of the dataset for modelling and 20% for validation. A 20-time bootstrapping procedure was used. Thus the accuracy measures described above were calculated on validation datasets 20 times, and the median of each was used as the final accuracy measure. For the global dataset which includes all countries, the data was sampled proportional to the station measurements of each country, that is, in each country, the stations were split into 4:1 for model fitting and validation.

2.3.3. Variable importance

The predictors are ranked by their importance calculated in the tree-based methods in the modeling process. This means that each variable is ranked 20 times (Section 2.3.2). The median of the 20 variable importance rankings is used for the final ranking. The proxy of variable importance for RF is the mean decrease in prediction error using the out-of-bag (OOB) permutation test, for SGB and XGB it is the mean decrease in prediction error using the permutation test of the entire dataset. These rankings are compared between different methods, day and night (only with XGB), and with the Lasso selected variables using the penalty (commonly denoted as λ) which gives the model whose error is within one standard error of the minimum mean cross-validated error. The Lasso selected variables are the variables that are selected more than 5 times out of the 20 times bootstrapping.

2.4. Model comparison

2.4.1. Comparison between global and national models for different techniques

We compared the prediction accuracy and the spatial prediction patterns of global and national daytime and nighttime RF, XGB, and Lasso models. To evaluate prediction patterns, spatial predictions were made over 10×10 km^2 areas that were randomly selected from a city of each country.

2.4.2. Comparison between global annual and day-night models

To make our global model comparable with other studies that do not separate between day and night, and to show the added value of separating between day and night in global modelling, we developed a RF model with the same settings as our global models but without separating between day and night. We compared the spatial prediction patterns and the prediction accuracy of the global annual model and our daytime and nighttime models.

2.4.3. Contributions of TROPOMI measurements in prediction accuracy

The variable importance analysis (Section 2.3.3) will show the impact of the TROPOMI measurements in different models. We assessed the contribution of the TROPOMI measurements by comparing if the involvement of it as a predictor in current model settings can improve the model prediction accuracy. We compared the RMSE of our prediction models (with all the predictors) and models with the same settings but without using the TROPOMI measurements as a predictor.

2.5. Spatial prediction patterns by settlement type and land use

We analysed the model performance for different settlement types (rural, urban) and land use types (industry). We applied a global model to predict a $1^\circ \times 1^\circ$ (approximately 111 km × 111 km) region centred at a US city, Phoenix, Arizona, and zoom into industrial, urban, and

Table 4
Gridded predictors. "mon" indicates months, (mon = 1,…,12).

Predictor	Variable name	Unit	Resolution
Monthly wind speed measured at 10 m altitude, for each month of the year.	Wind_speed_10m_mon	km/hr	10 km
Monthly temperature measured at 2 m altitude, for each month of the year.	temperature_2m_mon	Celsius	10 km
OMI 2017 annual mean vertical column density (level 3 product)	OMI_mean_filt; OMI	mol/cm^2	0.25 degree
TROPOMI vertical column density (level 3 product), mean of monthly average from 2018/02–2019/01	Tropomi_2018; Tropomi	mol/cm^2	0.125 degree
Remote sensing product generated from SCIAMACHY, GOME-2, global GEOS-Chem, 2011 (Geddes et al., 2016)	RSp	$\mu g/m^3$	10 km
Population 5 km resolution	pop5k	count	5 km
Population 3 km resolution	pop3k	count	3 km
Population 1 km resolution	pop1k	count	1 km

Table 5
The statistical learning methods used, their features, implementations, and the software used in this study ("R Package"). LM (Multiple Linear Regression), Lasso (Lasso regression), Ridge (Ridge regression), RF (Random Forest), SGB (Stochastic Gradient Boosting), XGB (Regularised Gradient Boosting). OLS: Ordinary Least Squares.

Method	Feature	R package	Implementation
LM	OLS fitting of all the variables	`base`	-
Lasso	L2 norm to regularize coefficient sizes	`glmnet`	Friedman et al. (2010b,c, 2011)
Ridge	L1 norm to regularize coefficient sizes	`glmnet`	Friedman et al. (2010b,c, 2011)
ElasticNet	Combining the L1 and the L2 norms for regularization	`glmnet`	Friedman et al. (2010b,c, 2011)
RF	Averaging over trees growing from sampled predictors and observations	`ranger`	Wright and Ziegler (2017)
SGB	Trees are grown subsequently from results of the previous tree	`gbm`	Greenwell et al. (2019)
XGB	Regularized SGB	`xgboost`	Chen et al. (2019b)

Table 6
The top 10 most important variables that are agreed by at least two ensemble tree-based methods. For the variable names please refer to Tables 3 and 4.

	US	China	Germany	Spain	Global
1	Tropomi_2018	Tropomi_2018	ROAD_M345_3000	pop3k	Tropomi_2018
2	ROAD_1_1000	ROAD_1_5000	ROAD_2_50	pop5k	pop3k
3	ROAD_1_100	temperature_2m_10	pop1k	Tropomi_2018	pop5k
4	ROAD_1_300	temperature_2m_9	ROAD_M345_300	ROAD_M345_500	ROAD_2_50
5	ROAD_1_800	RSp	ROAD_M345_25	ROAD_1_3000	ROAD_2_100
6	ROAD_1_500	OMI_mean_filt	pop3k	I_1_3000	temperature_2m_7
7	elevation	I_1_5000	ROAD_1_5000	ROAD_M345_5000	wind_speed_10m_6
8	pop3k	ROAD_1_3000	ROAD_M345_50	ROAD_M345_300	pop1k
9	I_1_300	pop1k	ROAD_M345_5000	ROAD_M345_3000	wind_speed_10m_3
10	pop5k	temperature_2m_1	pop5k	ROAD_M345_50	ROAD_M345_100

suburban areas. To further investigate the road effects, the predictor local roads (Table 4) was classified into secondary, tertiary, and small local roads as predictors.

3. Results

We compare for each model the variable importance and accuracy measures, and the contribution of TROPOMI measurements. Table 6 shows the top 10 most important variables that are agreed by at least two ensemble tree-based methods. It shows that important variables differ between geographical areas. The US model selected mostly highway road density variables. The model for China selected mostly background variables, notably the temperature variables. The model for Germany selected emission-related variables, i.e. primary and local road density. The model for Spain selected emission-related variables representing local road density. The global model selected a combination of emission-related and background variables partially overlapping with the four national models. The subsections below give an analysis of the variable importance and cross-validation results (Tables 7 and 8) for the national and global models. The relative accuracy measures of Tables 7 and 8 are shown in bar charts (supplement SF.9 to SF.13) to facilitate visual comparison.

3.1. National models

The US: Fig. 2 ranks the variables according to the variable importance calculated from the XGB for the daytime. The RF and SGB ranked the three local road emission-related variables lower and replaced them with background variables. The Lasso selected variables mostly coincide with highly ranked variables of the tree-based methods (Fig. 2, supplement SF. 7). The most notable difference between the ranking of the variables of XGB models for daytime and nighttime is the lower-ranked emission-related variables and higher-ranked background variables, which may be explained by fewer motor vehicles on roads at nighttime.

The cross-validation results (Tables 7 and 8, supplement SF. 7) indicate similar performance obtained by regularised linear regression and ensemble tree-based methods. The negative R^2 of the LM model indicates that the linear model prediction is worse than predicting using the sample mean. The higher R^2 and lower relative indices RRMSE, rIQR, and rMAE indicate better fitting at the nighttime. The relatively low performance of XGB among tree-based methods and Ridge among regularised regression-based methods may be caused by the low number of observations in the US. This shows that when the data is sparse, the regularised regression method which allows variable selection could further prevent over-fitting.

China: Both Lasso and ensemble tree-based methods (Fig. 3) rank background variables as the most important. Two temperature variables are ranked top by all the methods for the daytime but lower for the nighttime. Results of the cross-validation (Tables 7 and 8, supplement SF. 8) show that the tree-based methods have considerably better performance compared to the regression-based methods. The lower values of relative indices and higher R^2 indicate the tree-based and regularised regression-based methods perform better for the daytime. The dissimilarity between the tree-based methods is negligible.

Germany: Both RF and SGB rank the primary road length within 100 m high and the local road length within 50 m lower (Fig. 4). In addition, the RF ranked the primary road length within 25 m high, which is ranked low by XGB and SGB. For the nighttime, the TROPOMI is ranked higher and the local road length within 25 m buffer is ranked lower. The Lasso regression does not select local road length in the 25 and 50 m buffers, but emission-related highway and primary roads variables. The primary road length within 25 m is considered an important variable by Lasso and RF.

The prediction accuracy measures obtained (Tables 7 and 8, supplement SF. 9) are notably different between daytime and nighttime. With all the techniques the relative indices are all much higher for the daytime compared to the nighttime, indicating the model performances are better for the nighttime. However, except for the LM, the differences in R^2 between day and night are not that large. This might be caused by larger variations of NO_2 at daytime. The RF and XGB performed the best, but the differences between the methods are not that large except for the LM, indicating redundant NO_2 predictors. The differences between regularization-based linear regression methods are small.

Spain: RF ranked the emission-related variables that are highly ranked by XGB lower and replaced them with background variables

Table 7
Cross-validation results of different methods for the daytime. RMSE (root mean square error, μg/m^3), RRMSE (relative RMSE), IQR (Inter-quartile range, μg/m^3), rIQR (relative IQR), MAE (mean absolute error, μg/m^3). rMAE (relative MAE). US: United States of America, CN: China, DE: Germany, ES: Spain.

		RMSE	RRMSE	IQR	rIQR	MAE	rMAE	R^2
US day	LM	18.62	0.98	19.38	1.17	13.90	0.73	− 3.14
	Lasso	7.09	0.40	6.36	0.42	5.32	0.30	0.52
	Ridge	7.96	0.45	8.66	0.57	6.17	0.35	0.36
	ElasticNet	7.42	0.41	7.34	0.48	5.62	0.32	0.46
	RF	6.70	0.37	6.05	0.39	4.99	0.28	0.56
	SGB	6.78	0.38	7.03	0.46	5.18	0.29	0.51
	XGB	7.15	0.40	6.88	0.45	5.28	0.30	0.49
CN day	LM	9.29	0.27	8.56	0.25	7.52	0.22	0.64
	Lasso	8.69	0.21	11.42	0.28	6.81	0.17	0.58
	Ridge	8.70	0.21	10.89	0.27	6.78	0.17	0.58
	ElasticNet	8.67	0.21	11.22	0.28	6.77	0.17	0.59
	RF	6.93	0.17	8.27	0.20	5.23	0.13	0.73
	SGB	6.94	0.17	8.59	0.21	5.30	0.13	0.73
	XGB	7.15	0.18	8.77	0.22	5.47	0.13	0.72
DE day	LM	10.48	0.44	11.17	0.53	7.67	0.32	0.31
	Lasso	10.08	0.39	10.77	0.51	7.58	0.30	0.56
	Ridge	10.07	0.39	10.64	0.50	7.65	0.30	0.56
	ElasticNet	10.00	0.39	10.69	0.50	7.56	0.30	0.57
	RF	9.44	0.37	9.29	0.44	6.92	0.27	0.61
	SGB	10.49	0.41	10.89	0.51	7.67	0.30	0.52
	XGB	9.83	0.38	9.45	0.44	7.05	0.27	0.58
ES day	LM	7.51	0.33	8.71	0.44	5.79	0.25	0.65
	Lasso	7.57	0.32	8.65	0.42	5.81	0.25	0.65
	Ridge	7.59	0.32	8.55	0.41	5.77	0.25	0.65
	ElasticNet	7.61	0.32	8.66	0.42	5.80	0.25	0.65
	RF	7.02	0.30	7.83	0.37	5.27	0.23	0.70
	SGB	6.99	0.30	6.96	0.33	5.12	0.22	0.70
	XGB	7.16	0.31	7.75	0.37	5.31	0.23	0.69
global day	LM	9.19	0.33	9.85	0.38	7.20	0.26	0.60
	Lasso	10.76	0.35	13.48	0.47	8.35	0.27	0.54
	Ridge	10.74	0.35	13.64	0.48	8.34	0.27	0.54
	ElasticNet	10.76	0.35	13.66	0.48	8.36	0.27	0.54
	RF	8.69	0.28	9.96	0.35	6.48	0.21	0.70
	SGB	8.65	0.28	9.58	0.34	6.39	0.21	0.70
	XGB	8.73	0.28	10.01	0.35	6.49	0.21	0.70

Table 8
Cross-validation results of different methods for the nighttime. RMSE (root mean square error, μg/m^3), RRMSE (relative RMSE), IQR (Inter-quartile range, μg/m^3), rIQR (relative IQR), MAE (mean absolute error, μg/m^3). rMAE (relative MAE). CN: China, DE: Germany, ES: Spain.

		RMSE	RRMSE	IQR	rIQR	MAE	rMAE	R^2
US night	LM	19.76	0.91	18.78	0.94	14.41	0.66	− 3.21
	Lasso	7.36	0.34	8.33	0.42	5.79	0.27	0.54
	Ridge	7.97	0.37	8.18	0.41	6.20	0.29	0.45
	ElasticNet	7.69	0.35	8.47	0.43	5.98	0.28	0.50
	RF	6.78	0.31	6.98	0.35	5.25	0.24	0.61
	SGB	6.46	0.30	6.64	0.33	5.02	0.23	0.62
	XGB	7.16	0.33	8.14	0.41	5.71	0.26	0.55
CN night	LM	6.07	0.23	6.96	0.27	4.47	0.17	0.60
	Lasso	6.65	0.25	8.12	0.32	5.01	0.19	0.51
	Ridge	6.69	0.25	8.13	0.32	5.04	0.19	0.51
	ElasticNet	6.63	0.25	8.06	0.31	4.99	0.19	0.52
	RF	5.39	0.20	6.23	0.24	3.97	0.15	0.68
	SGB	5.50	0.21	6.34	0.25	4.07	0.15	0.67
	XGB	5.64	0.21	6.42	0.25	4.17	0.16	0.65
DE night	LM	6.20	0.29	7.03	0.35	4.56	0.22	0.52
	Lasso	5.72	0.27	7.19	0.36	4.40	0.21	0.60
	Ridge	5.62	0.27	6.66	0.33	4.33	0.21	0.61
	ElasticNet	5.69	0.27	7.09	0.35	4.39	0.21	0.60
	RF	5.30	0.25	6.71	0.34	4.17	0.20	0.65
	SGB	5.50	0.26	6.75	0.34	4.29	0.20	0.63
	XGB	5.34	0.25	6.62	0.33	4.18	0.20	0.65
ES night	LM	7.98	0.37	10.23	0.55	6.33	0.29	0.62
	Lasso	7.99	0.37	9.50	0.51	6.16	0.28	0.62
	Ridge	8.01	0.37	9.31	0.50	6.09	0.28	0.62
	ElasticNet	8.03	0.37	9.57	0.51	6.17	0.28	0.62
	RF	7.25	0.33	8.06	0.43	5.48	0.25	0.69
	SGB	7.45	0.34	8.19	0.44	5.54	0.25	0.67
	XGB	7.74	0.35	8.51	0.46	5.77	0.26	0.64
global night	LM	7.00	0.30	7.66	0.34	5.09	0.22	0.61
	Lasso	7.42	0.31	8.45	0.38	5.48	0.23	0.56
	Ridge	7.45	0.32	8.53	0.38	5.51	0.23	0.55
	ElasticNet	7.44	0.32	8.53	0.38	5.49	0.23	0.55
	RF	6.29	0.27	6.51	0.29	4.47	0.19	0.68
	SGB	6.31	0.27	6.81	0.31	4.54	0.19	0.68
	XGB	6.33	0.27	6.78	0.30	4.52	0.19	0.68

(Fig. 5). Compared to the tree-based methods, the Lasso regression selected more emission-related variables associated with the primary roads. The differences between day and night are not large for top-ranked important variables, except for the primary road length within 300 m buffer which is ranked as less important for the nighttime.

The cross-validation results (Tables 7 and 8, supplement SF. 10) show that the differences between the prediction accuracy for daytime and nighttime are smaller compared to China and Germany, with higher accuracy obtained for the daytime. At nighttime, there is a notable increase in rIQR, which may indicate the need of additional predictors to account for potentially different emission sources or land use categories that influence the transportation process in some areas of Spain at nighttime. In general, the tree-based methods obtained higher accuracy compared to the regression-based methods, especially the rIQR for the nighttime. The reason may be that the tree-based methods are more flexible to identify the non-linear relationships between predictors and NO_2.

3.2. Global models

For the global XGB model, 30% of the top 15 ranked variables are emission-related variables (Fig. 6), The top 9 ranked important variables are consistent between the tree-based methods. The RF ranked the local road length 100 m buffers and highway length in 100 m buffer variables lower. The important variable ranking differences between day and night are mainly in the low ranking of the meteorological variables for the nighttime.

The cross-validation results (Tables 7 and 8, supplement SF. 11) show that lower relative indices are obtained for the night than the day for all methods. The tree-based methods, however, obtained higher R^2 for the daytime. The tree-based methods obtained considerably lower RMSE, MAE, and IQR, and higher R^2 compared to regression-based methods. There is no outstanding method within the ensemble tree-based methods (RF, SGB, and XGB) and the same applies to the regression-based methods.

3.3. Comparison of global and national models and techniques

Fig. 7 shows that the differences in the RMSE between the global and national models are small for XGB and RF. For some countries, global tree-based methods even outperform national models, for instance for Spain (daytime and nighttime) and US (nighttime). This indicates that tree-based methods in some cases gain from a large number of data points used in the global models. In contrast, Lasso mostly produces better results when applied to the national models, particularly for the daytime. This indicates geographically heterogeneous relationships between predictors and NO_2 concentrations at the global scale. At the nighttime, the differences between the global and national models for the US, Germany, and Spain become considerably smaller, which may be caused by the reduction of traffic-related emissions at

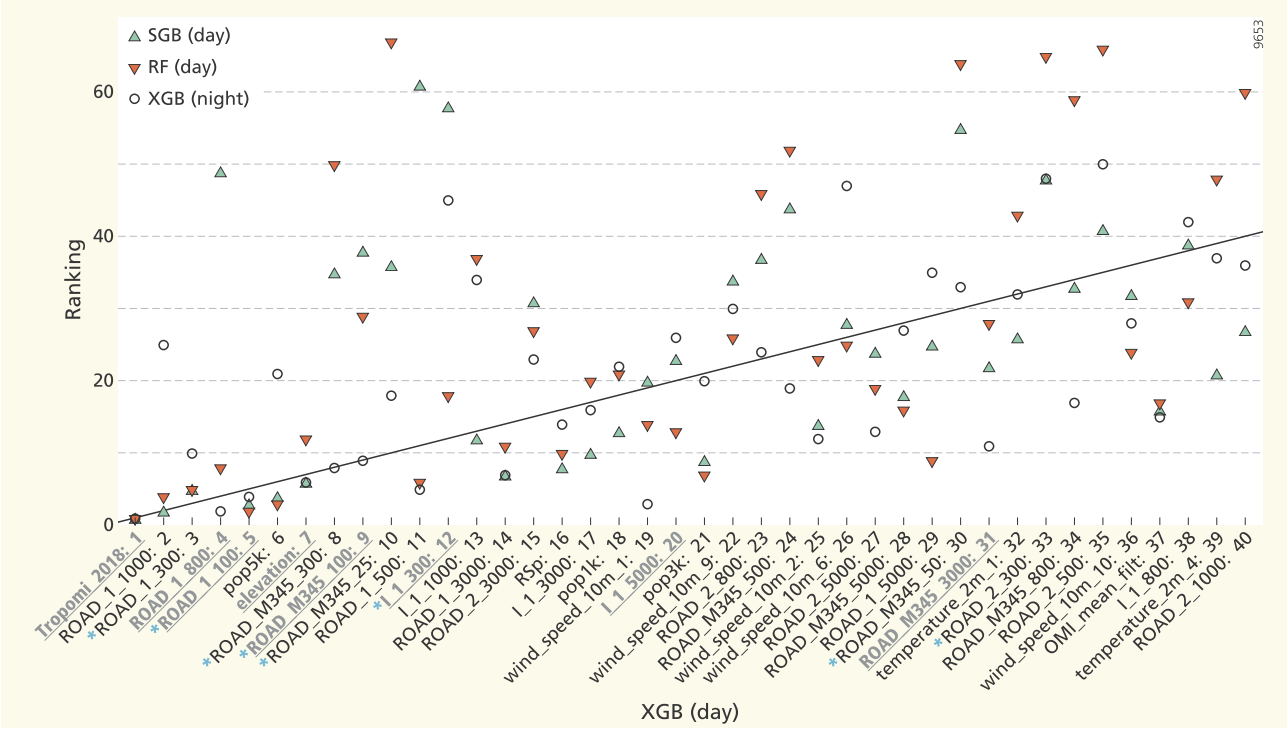

Fig. 2. Variable importance rankings and variables selected by Lasso, the US. The variables are sorted along the x-axis according to the rank order of the variables for the daytime XGB model. For each variable, the y-axis gives the rank order for the SGB daytime model, the RF daytime model, and the XGB nighttime model, using different symbols for each model. The variables that are selected by the Lasso regression for the daytime are in grey, bold, and with an underline. The blue star tagged variables on the x-axis indicated the variables that are emission-related. For explanation of variable names refer to Tables 3 and 4. For better visualisation, only the variables ranked with the importance top 40 using the XGB are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

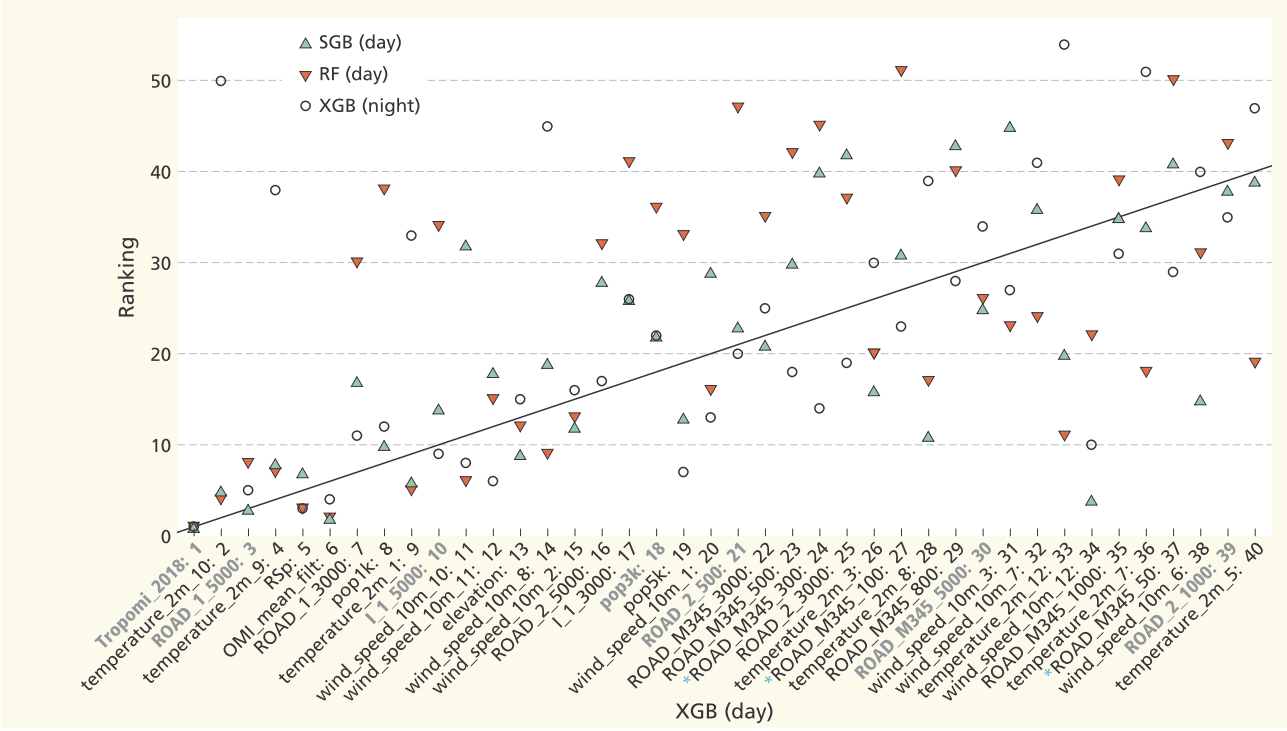

Fig. 3. Variable importance rankings and variables selected by Lasso, China. The variables are sorted along the x-axis according to the rank order of the variables for the daytime XGB model. For each variable, the y-axis gives the rank order for the SGB daytime model, the RF daytime model, and the XGB nighttime model, using different symbols for each model. The variables that are selected by the Lasso regression for the daytime are in grey, bold, and with an underline. The blue star tagged variables on the x-axis indicated the variables that are emission-related. For explanation of variable names refer to Tables 3 and 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

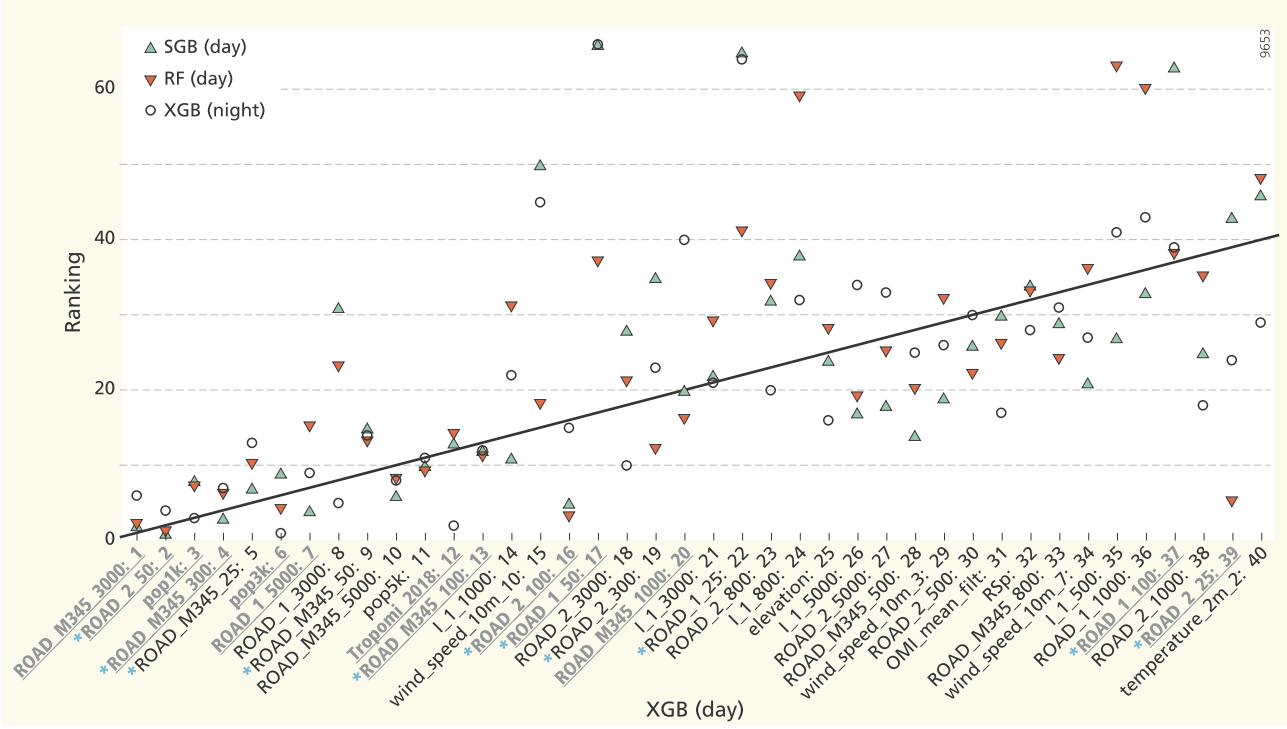

Fig. 4. Variable importance rankings and variables selected by Lasso, Germany. The variables are sorted along the x-axis according to the rank order of the variables for the daytime XGB model. For each variable, the y-axis gives the rank order for the SGB daytime model, the RF daytime model, and the XGB nighttime model, using different symbols for each model. The variables that are selected by the Lasso regression for the daytime are in grey, bold, and with an underline. The blue star tagged variables on the x-axis indicated the variables that are emission-related. For explanation of variable names refer to Tables 3 and 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

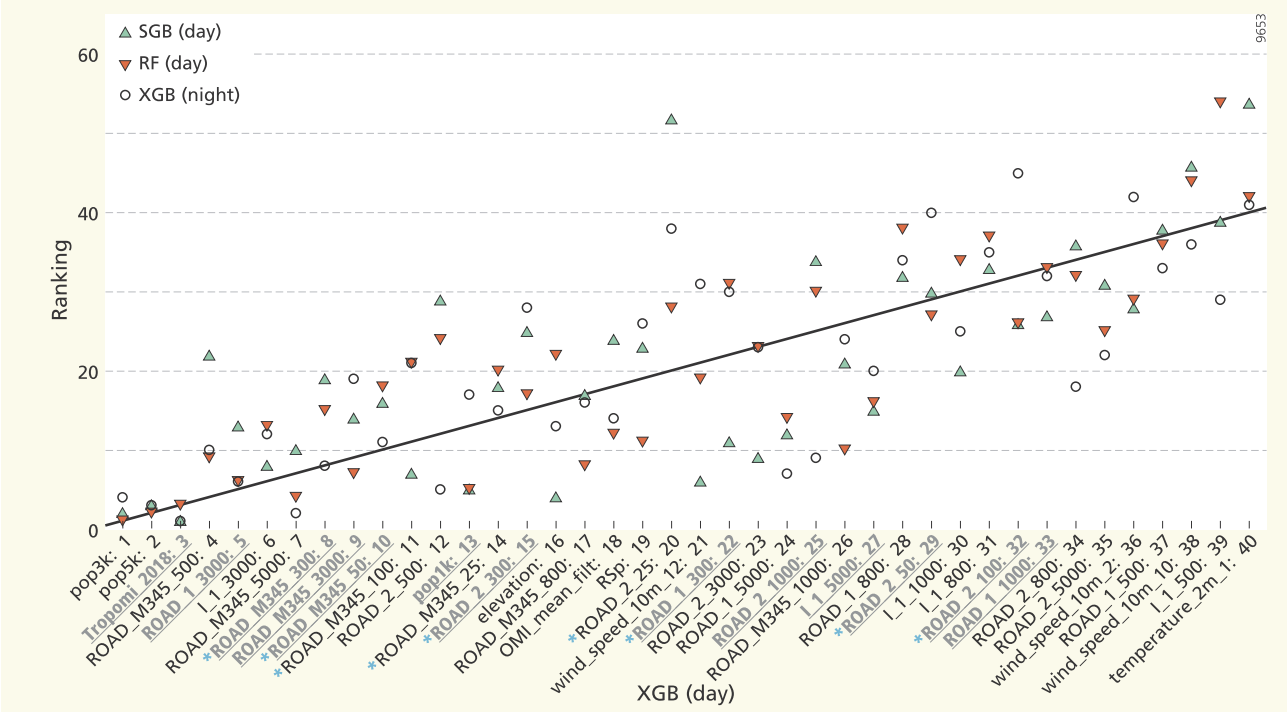

Fig. 5. Variable importance rankings and variables selected by Lasso, Spain. The variables are sorted along the x-axis according to the rank order of the variables for the daytime XGB model. For each variable, the y-axis gives the rank order for the SGB daytime model, the RF daytime model, and the XGB nighttime model, using different symbols for each model. The variables that are selected by the Lasso regression for the daytime are in grey, bold, and with an underline. The blue star tagged variables on the x-axis indicated the variables that are emission-related. For explanation of variable names refer to Tables 3 and 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

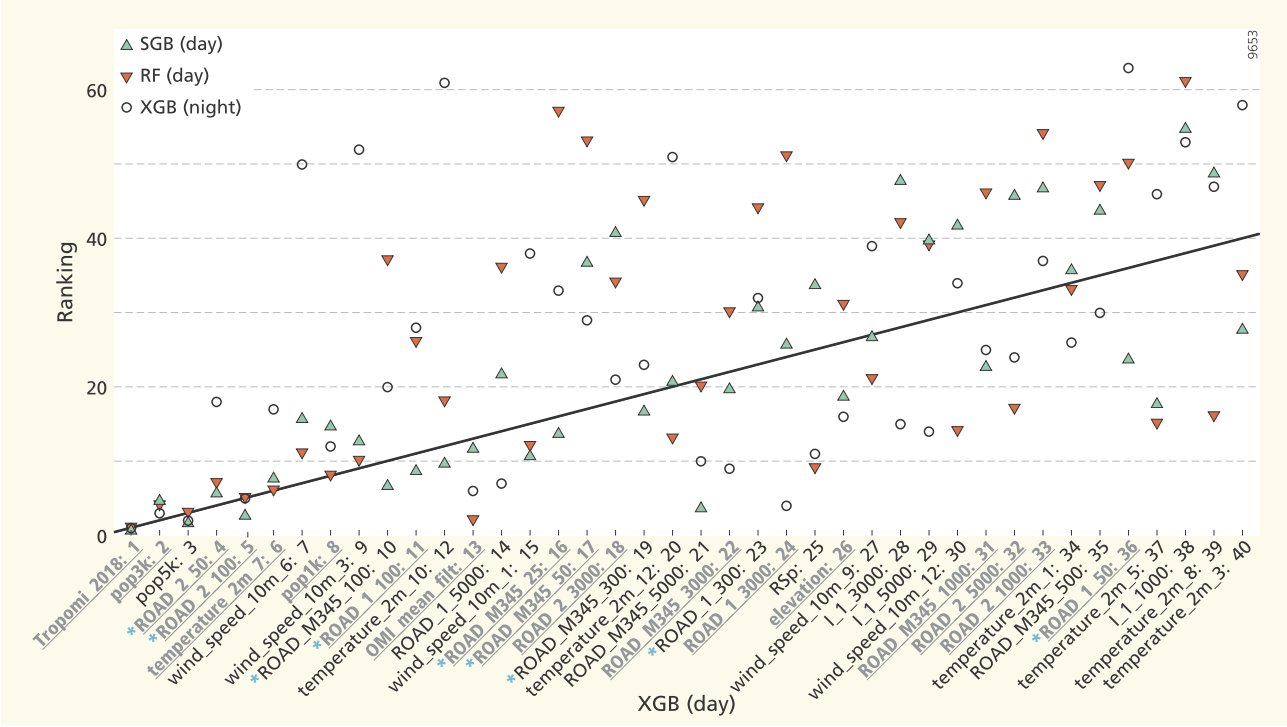

Fig. 6. Variable importance rankings and variables selected by Lasso, global. The variables are sorted along the x-axis according to the rank order of the variables for the daytime XGB model. For each variable, the y-axis gives the rank order for the SGB daytime model, the RF daytime model, and the XGB nighttime model, using different symbols for each model. The variables that are selected by the Lasso regression for the daytime are in grey, bold, and with an underline. The blue star tagged variables on the x-axis indicated the variables that are emission-related. For explanation of variable names refer to Tables 3 and 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

night, which is also reflected in the daytime and nighttime variable importance.

3.3.1. Spatial prediction patterns

The RF model predictions are shown for an area of each country (Fig. 8). Road infrastructures of the corresponding tiles are shown in the supplement (SF.5). For all countries, when the global model is used, the predicted daytime NO_2 concentrations are high along roads, reflecting recognisable patterns of the highway, primary and local roads. For the nighttime, the NO_2 pattern is smoothed along the local roads but highways and primary roads are identifiable. The predictions at nighttime for all global models are generally lower and smoother, which indicates the emission-related predictors are less influential in the global nighttime models.

To quantify the disparity between the national and global model predictions, we calculated the R^2 of the linear regression between national and global RF model predictions in the RF predicted tiles (Fig. 11). The national and global models for China are the least correlated. For all countries except China, the correlations between national and global models are higher for the nighttime, which is consistent with the finding that the global and local RF models obtained similar accuracy for the nighttime (Fig. 7). Note that the R^2 calculated here is only for a sampled region of each country, the R^2 between the full extent of the national and global model predictions will be different.

To compare the spatial predictions from different techniques and to study the influence of variable importance rankings on spatial patterns, we additionally show the spatial prediction using XGB (Fig. 9) and Lasso (Fig. 10). Compared to the RF predictions, the XGB predictions show more details spatially but some details are possibly artefacts. As XGB is regularised, it is expected to be more robust to over-fitting compared to non-regularised methods. A possible explanation is that the higher-order of gradient descent it uses may facilitate an optimal solution to be reached with fewer trees. The Lasso predictions do not reveal local roads patterns. The US, Germany and Spain models predictions with RF, XGB, and Lasso show diminishing NO_2 concentrations away from dense road networks. The same pattern was observed with the global model predictions of RF and XGB but is much less obvious with Lasso. The lack of strength of Lasso to reveal road-related emission effects again indicates the necessity of modelling the non-linear relationships between predictors and NO_2 concentrations.

For the US, all the models predict high NO_2 concentrations along highways at daytime. At nighttime, the NO_2 is distributed more evenly around the highway. These patterns are consistent with the finding that emission-related highway variables are ranked high in importance with RF and XGB for the daytime and low for the nighttime (Fig. 2).

For China, the effect of high NO_2 concentrations around primary roads shown by the RF, XGB, and Lasso global model predictions disappear when the national model is used, which is consistent with the fact that background variables rank high in the national model (Fig. 3).

For Germany, the XGB and RF national model predictions show strong local roads effects, which is not observed in the Lasso national model prediction. Except for the RF national model, all models show high NO_2 concentrations along the highway. The highest NO_2 is predicted by the XGB, RF, and Lasso models in the city centre along a primary road. Primary road length within 25 m, 50 m and 100 m buffers are ranked very high in the RF daytime model (Fig. 4). The RF and XGB models have similar strength in estimating road effects, despite the XGB ranked local road length in small buffers lower.

For Spain, the RF and Lasso national model predictions show high NO_2 at the city centre and diminishing NO_2 away from the city centre, for both daytime and nighttime. The XGB national model predictions show local road effects at the suburban area, where higher NO_2 concentrations are predicted at nighttime compared to daytime. This counter-intuitive spatial pattern could be caused by model over-fitting.

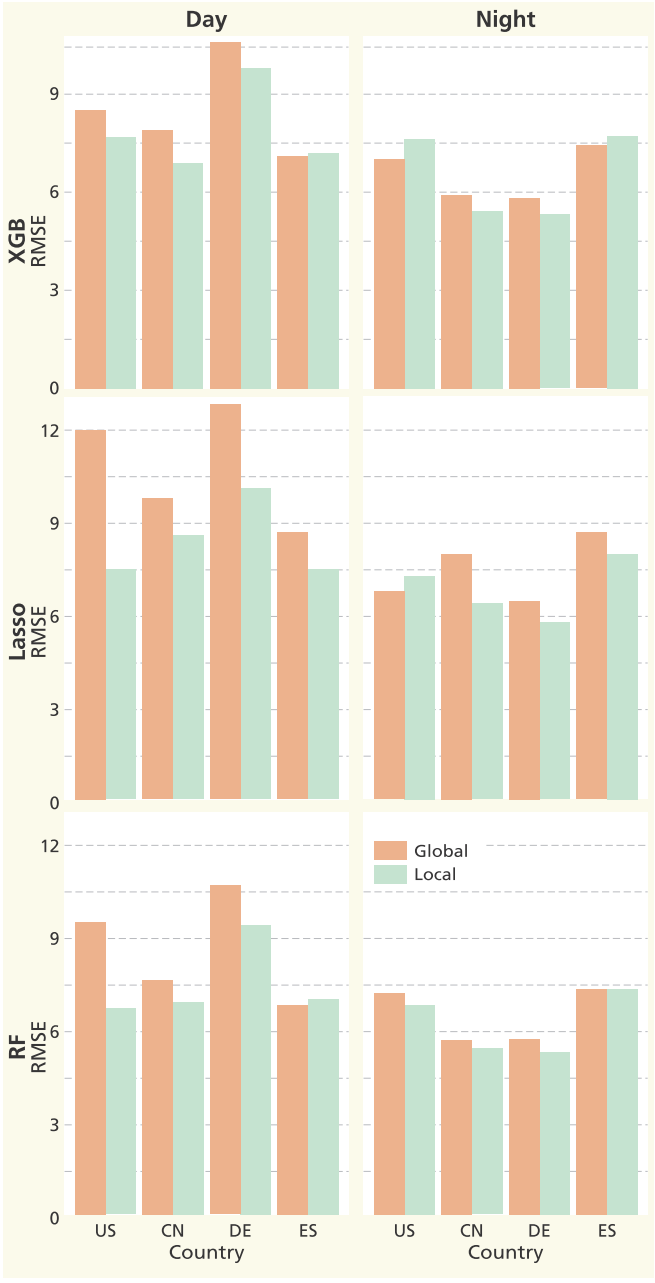

Fig. 7. RMSE (μg/m^3) of global and local extreme gradient boosting (XGB), Lasso and random forest (RF) models developed for each country, for daytime and nighttime.

3.4. Global annual versus day and night models

A comparison of the predictions from the annual global RF model (Fig. 12) and the daytime and nighttime models (Fig. 8) shows that the spatial prediction of an annual model is not a simple aggregation of the daytime and nighttime models. The annual model prediction in Spain shows higher values further away from the roads, compared to the daytime and nighttime models. For the tiles in Germany and the US, the annual model prediction is more smoothed along the highway and the primary roads and is more smoothed over the local roads compared to the daytime model. The RMSE of the annual global RF model is 7.43 (μg/m^3) and the R^2 is 0.7.

3.5. Contribution of TROPOMI measurements as a predictor

For all the national and global models, the tree-based methods ranked the TROPOMI variable as an important variable (Tables 7 and 8). If the TROPOMI measurements are not used as a predictor, the RSp (GEOS-Chem and satellite-based surface concentration product) and OMI variables become top-ranked for the China, Spain, and global models, while with TROPOMI, the RSp and OMI are less important (Table 9). For the model of Germany, which includes TROPOMI as an important variable, the RSp and OMI predictors are unimportant in the models which exclude the TROPOMI variable. The lowest RMSE that is achieved in each study area with and without including the TROPOMI variable is shown in Fig. 13. The RMSE that is obtained using other tree-based methods is shown in Tables 7 and 8 for the standard model and in the supplement (ST.1) for models without including the TROPOMI variable. Including the TROPOMI variable increased the prediction accuracy in the US for the nighttime and in Spain for the daytime, whereas the differences for other countries are negligible.

3.6. Spatial predictions for different land conditions

The spatial predictions of the RF global model for a larger area reveal a more gradual decline of NO_2 from the city centre to suburban areas compared to predictions of the Lasso global model (Fig. 14). This is likely attributed to the flexible tree-based ensembles of RF. Consistent with Section 3.3.1, the XGB model predictions decline less rapidly towards the suburban areas. In urban areas, Lasso and XGB predicted higher NO_2 compared to RF. Lasso predicted high NO_2 at highways junctions and highway-primary roads, indicating that the predictions made by Lasso are highly influenced by the highway and primary road length in small buffers but less by local roads in large buffers. This may also explain the higher NO_2 concentration in the suburban areas from the RF model predictions compared to the Lasso prediction. XGB shows rapidly declining NO_2 predictions away from the highway and primary roads, but high NO_2 close to local roads, which is likely caused by the model over-fitting.

4. Discussion

This study evaluates methods to develop a global high-resolution NO_2 model to facilitate health studies. We compared models that were different regarding (1) geographical extent (global versus national models), (2) temporal resolution (nighttime and daytime models), (3) statistical learning technique, and (4) satellite products. In this section, we discuss these comparisons and compare our results with the global NO_2 model of Larkin et al. (2017) and other national models, and discuss challenges for global mapping.

Geographical extent: The important variables that are selected by the ensemble tree-based models differ among models of different geographical extent. Each national model includes a set of predictors that is mainly related to a particular determinant of air pollution, while the global model has a more varied set of important predictor variables. Emission-related variables are important in Germany and Spain, but are respectively related to primary and local roads, indicating the necessity of separating different types of roads between countries. For the China national daytime models, climate variables have high importance, which is not the case for other national models. The strong correlation between climate and NO_2 is most likely due to the fact that China spans different climate zones and latitudes and the climate pattern may coincide with the spatial patterns of heavy industry (e.g. in the North) and economic development (e.g. along the coast) of the country, which may strongly relate to NO_2. The population variable is consistently an important variable in all models.

The outperformance of ensemble tree-based methods relative to Lasso in global modelling is reflected in the additional spatial detail in the predictions and RMSE values that are equal to or even lower than

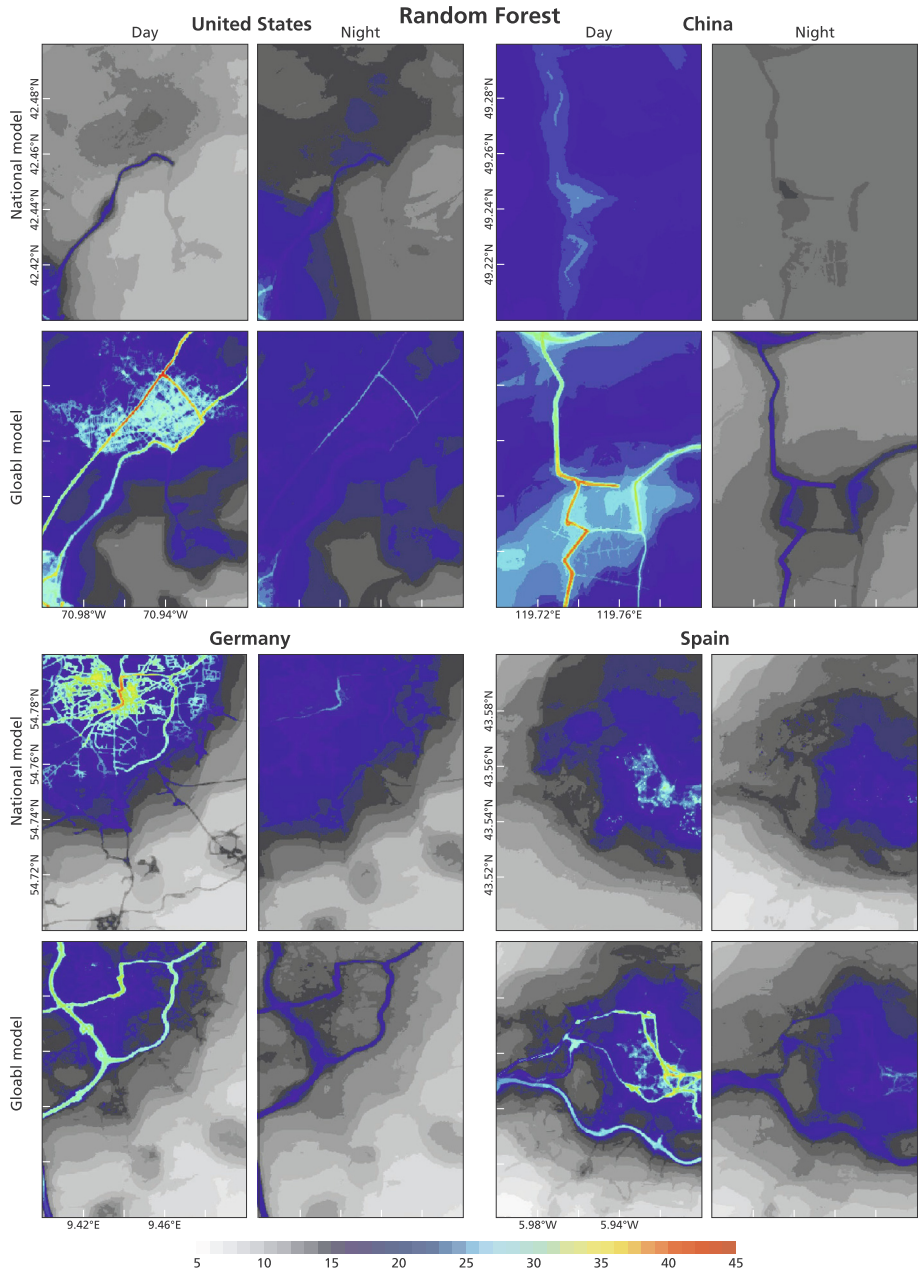

Fig. 8. Spatial predictions of NO_2 (μg/m³) for the four countries using random forest. For each country, predictions are made using the global and national models separately for daytime and nighttime. The tile in the US covers a Massachusetts city called Lynn (42.47° N, 70.94° W). The tile in China covers an inner-Mongolian city called Hulunbuir (49.23° N, 119.76° E). The tile of Germany covers a city in the north of Schleswig–Holstein called Flensburg (54.79° N, 9.44° E). The tile of Spain covers a city in the region of Asturias called Aviles (43.55° N, 5.92° W). The road structures of the corresponding tiles are shown in the supplement (SF.5).

those found by the national models. Thus, tree-based methods are capable of preserving local relationships in global mapping. The local road effects that are found in the national models, for example the strong local road effects appearing in the model of Germany, are also present in the global RF model, but not in the global Lasso model. Road effects in national models are preserved in the tree-based global models most likely because tree-based methods consider joint effects of background variables such as climate, elevation, population, and satellite measurements and interactions between background and emission-related predictors. By using ground monitor measurements available globally, data-driven methods can benefit from a larger number of observations to derive the NO_2 emission and dispersion dynamics.

To understand the effect of the road variables on the accuracy measures, we subdivided the global validation dataset in station measurements close to roads (stations located less than 25 m distance away from local roads or less than 100 m from highway or primary roads, 396 stations) and station measurements away from roads (the remaining stations in the validation dataset, 269 stations). We then calculated performance measures for each subset using the global daytime and nighttime RF models. The R^2 obtained for the close-to and away-from roads subsets were respectively 0.64 and 0.83 for the daytime and approximately 0.70 for both subsets for the nighttime. The results indicate that the daytime prediction accuracy may increase with distance away from roads, possibly due to the higher variation of NO_2 close to roads. In our future research, we will carry out a comprehensive accuracy assessment process that not only considers the accuracy assessed with point measurements but also is constraint by spatial prediction patterns. We compared spatial patterns of four relatively small cities to avoid the nonrepresentativeness of big cities of general national spatial patterns. The comparison between big cities may lead to additional

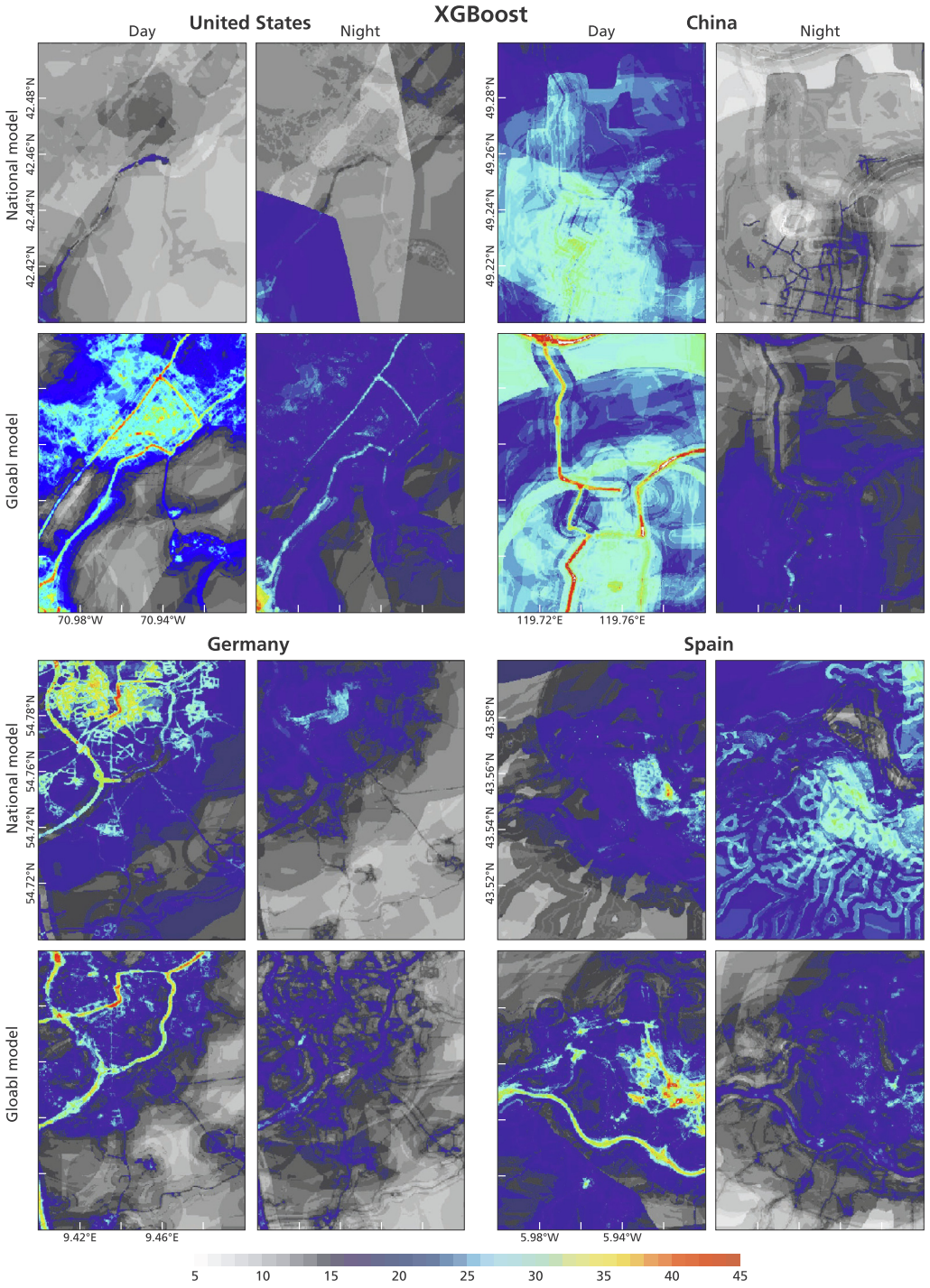

Fig. 9. Spatial predictions of NO_2 (μg/m³) for the four countries using XGB (Extreme Gradient Boosting). For each country, predictions are made using the global and national models and separately for daytime and nighttime. For the geographical information of each tile, please refer to the caption of Fig. 8.

findings and is planned in our future work.

Temporal resolution:

Picornell et al. (2019) highlighted that a home-based static exposure assessment may provide unrealistic estimates for population exposure and emphasised the need to integrate human activity patterns in exposure assessment. Separating day and night in air pollution mapping, as was done in our study, has important implications in exposure assessment as human space–time activities are commonly different between day and night. In addition, our results show differences in spatial prediction pattern between an annual global model and global models that separate between day and night. This indicates that separating between day and night may allow for improved representation of the relationships between predictors and NO_2. Future studies need to evaluate different criteria to split the day and night and consider finer temporal resolutions and perform analysis for representative hours for various periods, for example, 4:00 am to 5.00 pm for atmospheric stability (Perrino et al., 2001).

Statistical learning techniques: The bootstrapped cross-validation accuracy differs slightly between the tree-based methods, as well as between the regularised linear regression methods. For the global models, XGB, RF, and SGB obtained the same RMSE, but the spatial predictions from XGB show more artefacts. We suspect that the cause of

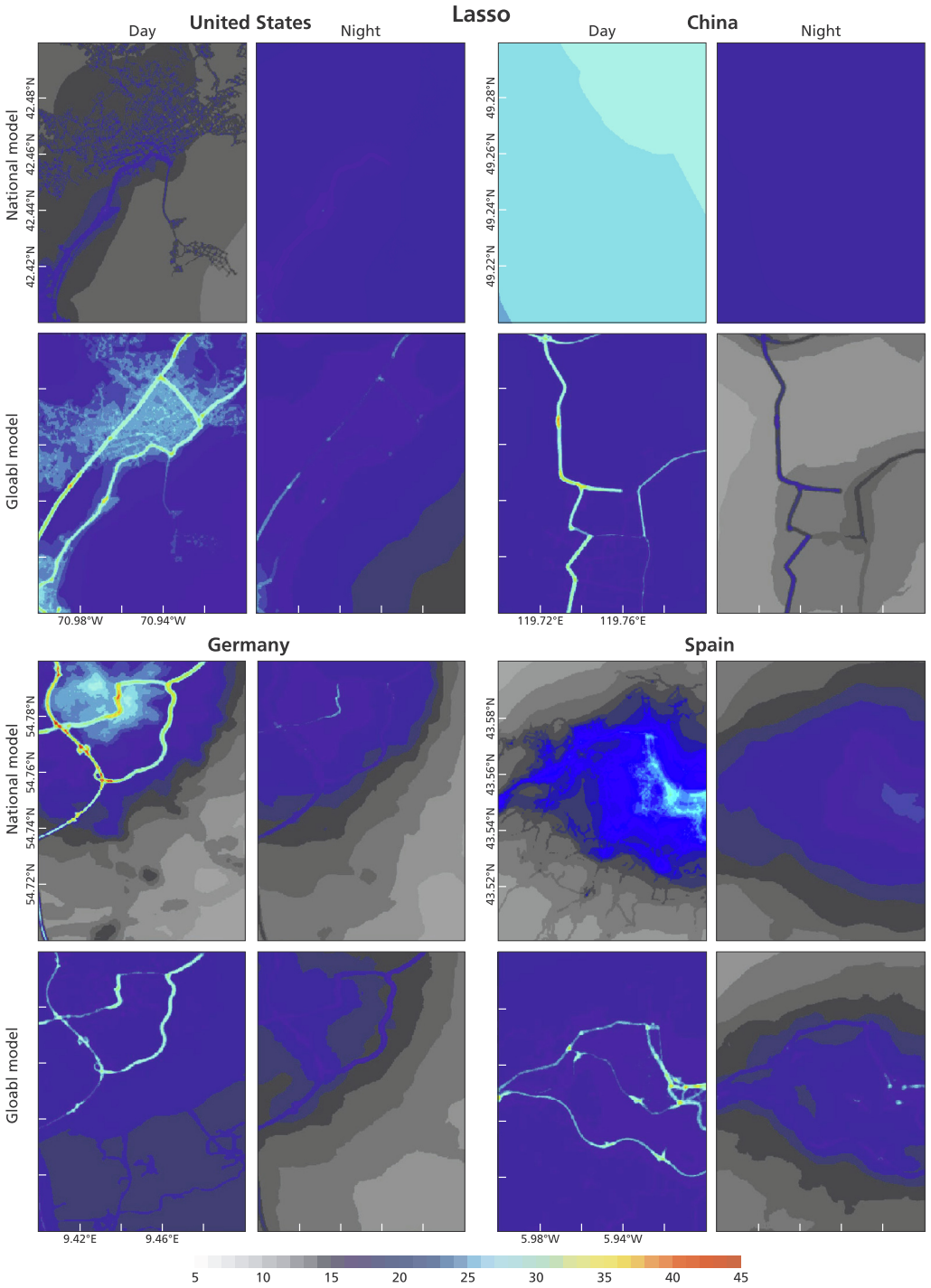

Fig. 10. Spatial predictions of NO_2 ($\mu g/m^3$) for the four countries using Lasso. For each country, predictions are made using the global and national models and separately for daytime and nighttime. For the geographical information of each tile, please refer to the caption of Fig. 8.

this may be the high number (2000) of trees that is used for XGB to make the XGB model consistent regarding methodology with RF and SGB. We, therefore created an additional XGB model by reducing the number of trees to 1000, which is the optimal setting found in the hyperparameter tuning procedure (Section 2.3.1). The resulting models showed similar accuracy measures and spatial patterns compared to our standard national and global models. This experiment indicates that the number of trees may not be the main cause of the artefacts in the XGB predictions. In future studies, XGB, therefore, needs to be further optimised, for example by testing the use of lower learnig rates, alternative regularizations, and more observations.

For the national models of China and Spain, as well as for the global models, the tree-based methods obtained higher accuracy compared to the linear regression-based methods and the differences in terms of cross-validation accuracy between global and national models are smaller for the tree-based methods compared to linear regression-based methods. These results indicate that the non-linear relationship between NO_2 and the predictors could be better modelled by the tree-based methods. The rankings of the variable importance (even for variables that are not highly correlated) are different between techniques but for the top-ranked variables are mostly consistent.

Satellite products: Excluding the TROPOMI variable does not

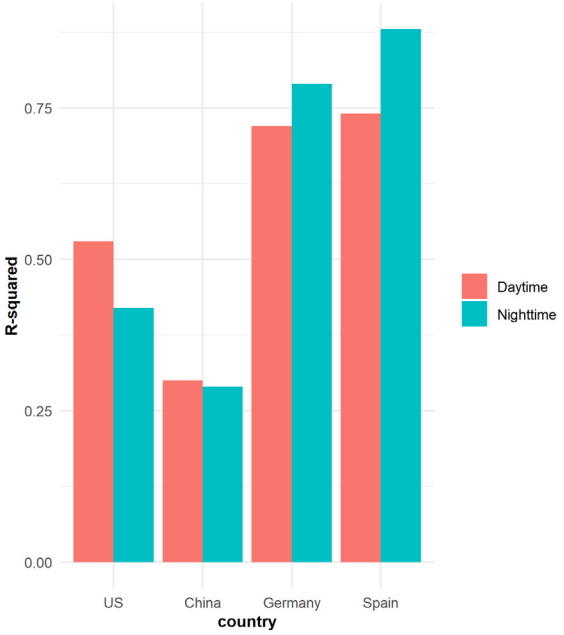

Fig. 11. R^2 of the linear regression between national and global random forest model predictions on the tiles of Fig. 8.

decrease the RMSE in the global model and most of the national models. However, as the TROPOMI variable is ranked as one of the most important predictors in most national models and the global model, in regions where ground monitors are sparse and the predictors are less reliable, the TROPOMI measurements may provide essential information. Future studies should evaluate the value of TROPOMI also in areas with less (or lower quality) ground station monitoring data as in these areas the relative contribution may be larger.

Comparison with other global and national models: The study of Larkin et al. (2017) obtained an RMSE of 5 ppm (about 9.5 μg/m^3) and an R^2 of 0.54. In comparison, our RF global model that does not separate between day and night obtained a lower RMSE (7.4 μg/m^3) and a higher R^2 (0.7).

For national models, Cuevas et al. (2014) studied the long-term (1996–2012) trends in NO_2 over Spain using satellite imagery. Though not completely comparable with our Spain model, Cuevas et al. (2014) indicated high NO_2 concentrations at densely populated areas, which is consistent with our finding that population is an influential predictor in Spain, which is also reflected in the spatial prediction map. Xu et al. (2019) provided a LUR model for China at a 1 × 1 km grid for the year 2014–2015. This study obtained a slightly higher R^2 (0.73–0.78) compared to our model (0.70), which could be caused by the use of

Table 9
Variable importance ranked by the XGB for the daytime using the models that have been developed in this study (called "Standard model") and models that leave out the TROPOMI measurements (called "Without TROPOMI"). OMI: OMI_mean_filt, TROPOMI: Tropomi_2018.

	Standard model			Without TROPOMI	
	TROPOMI	RSp	OMI	RSp	OMI
the US	1	16	37	16	23
China	1	5	6	1	2
Germany	12	32	31	31	24
Spain	3	19	18	7	8
Global	1	25	13	9	1

variables that are not yet available globally in the Chinese study.

Other technical challenges: This study used the OpenAQ database, which required a rigorous quality control before application. We compared the locations of the records with the stations from the Airbase project (European Environment Agency, 2018; de Hoogh et al., 2016b) and found that most of the official national public measurements are in OpenAQ. This indicates that the NO_2 measurements from the OpenAQ community have reliable data sources for conducting multi-country studies. However, important meta-information, such as the instrument types, is missing. In addition, some measurement data available in Airbase (e.g. for Italy, Portugal, and Ireland) were missing in the OpenAQ. The differences in the locations of the stations between our dataset and the Airbase dataset is provided in a digital supplement ("airbasevsoaq.html"). For future studies, the Airbase dataset needs to be combined with the OpenAQ database and more station measurements from US, Australia, South America, and Africa need to be included from other datasets.

As we do not have access to an external dataset, we used the same dataset for hyperparameter tuning and model training. One could argue this leads to possible information leakage from the validation set. However, it has been proved in Chen et al. (2019a), who applied the same hyperparameter tuning and training procedure, that validation results on an external dataset are similar to the cross-validation results. This indicates that the model accuracy we obtained is valid. Also, François (2017) (page 97) shows that if a very small number of hyperparameters (in our study 1–3) is tuned, and only once, it is valid to assess accuracy with the data set also used for hyperparameter tuning. In addition, Chen et al. (2019a) used 5-fold cross validation for both hyperparameter tuning and model training. We used 20-times bootstrapping, which means that even if there is information leakage, it is minimized. Kerckhoffs et al. (2019) used default hyperparameter settings; however, not tuning hyperparameter may lead to a higher risk of model over-fitting, and unreliable comparisons between machine learning models.

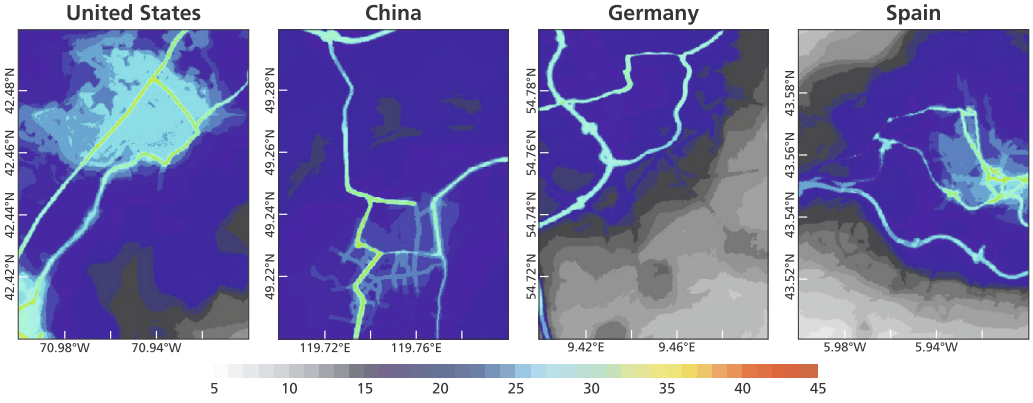

Fig. 12. Spatial predictions of NO_2 (μg/m^3) for the four countries from the annual global random forest model (i.e. not separating between day and night). For the locations of each tile, see the caption of Fig. 8.

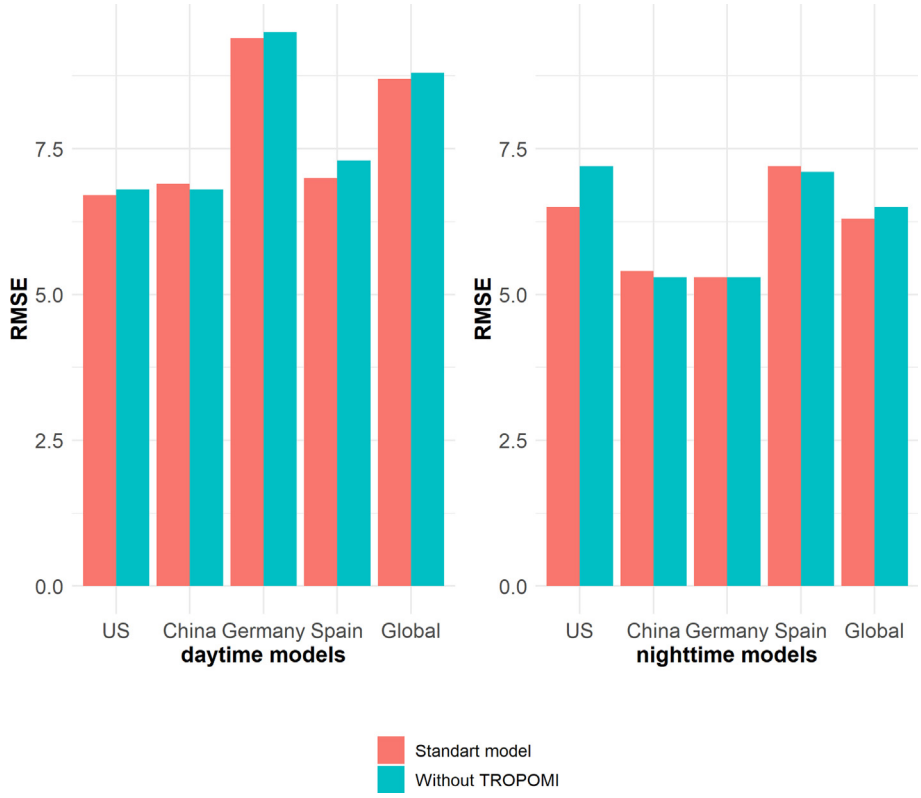

Fig. 13. The lowest RMSE predicted using the models that have been developed in this study (called "Standard model") and models that leave out the TROPOMI measurements (called "Without TROPOMI").

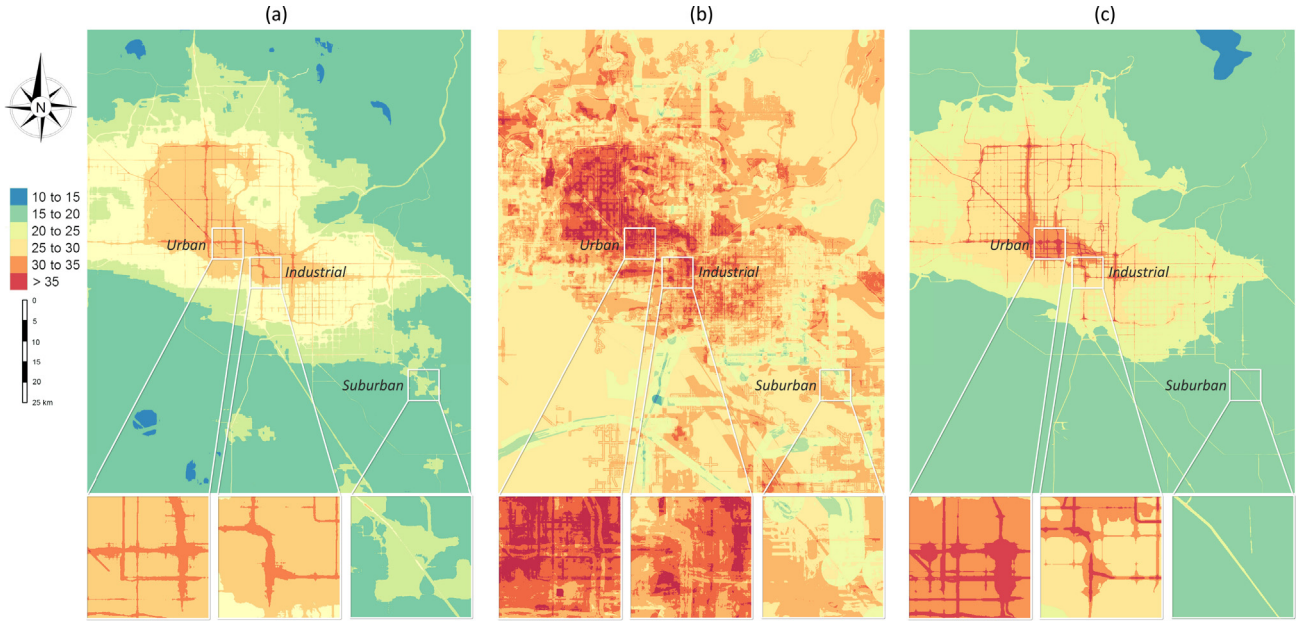

Fig. 14. Spatial predictions of annual mean NO_2 (μg/m^3) using (a) RF, (b) XGB, and (c) Lasso for a 1 ° × 1° (approximately 111 km × 111 km) region centred at Phoenix, Arizona, US (33.5 ° N, 112.09° W). Details are shown for each method for urban, industrial, and suburban areas. Figure SF. 8 shows roads and industrial areas.

Lastely, prediction accuracy may be improved by downscaling the coarse-resolution predictors, e.g. using area-to-point Kriging (Yoo and Kyriakidis, 2006), before reading the values at the station location, as this will smooth the boundary effects between grid cells. Also, other global land-use variables may be included, such as remotely sensed light at night, biomass, and land cover classes.

5. Conclusion

This study evaluated global and sub-domain (US, China, Germany and Spain) models, different statistical learning techniques, and the use of the TROPOMI measurements for high-resolution global NO_2 mapping. Cross-validation accuracies of the tree-based methods are in general higher than those of linear regression-based methods for the

national models and the global model. We found that for the ensemble tree-based methods, the global models are almost as good as national models in terms of the cross-validation RMSE, which is not the case with the linear regression-based methods. This may imply that due to improved learning from a larger number of observations, global ensemble tree-based methods are capable of incorporating patterns resulting from underlying emission and dispersal dynamics, providing useful information to map areas with deficient NO_2 ground measurements. Moreover, the RF and XGB models reveal more spatial details compared to the Lasso models. In addition, techniques may have obtained similar accuracy in terms of validation against point measurements (e.g., different tree-based techniques) while their spatial prediction pattern may be notably different. Based on the spatial patterns, we favour RF over XGB because of the artefacts that are modelled by the latter. When more monitor stations are available, and through better optimization of hyperparameters, XGB may be recommended. Lastly, we found that despite being an important predictor, adding the TROPOMI vertical column density on top of OMI measurements and the satellite-based surface product (Geddes et al., 2016) does not improve the RMSE in the current model setup and countries used for validation. Future studies need to evaluate TROPOMI with a particular focus on countries with limited ground station data.

Declaration of Competing Interest

None.

Acknowledgement

This research is funded by the Global Geo Health Data Centre (Utrecht University) and the Startimpulsprogramma Meten en Detecteren van Gezond Gedrag (Dutch Science Foundation). We are thankful to data providers of OpenAQ and the Chinese environmental institute. The authors appreciate Ton Markus for his advises and contributions on improving the figures. The authors are grateful to the editors and reviewers for their contributions.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.envint.2020.105856.

References

Achakulwisut, P., Brauer, M., Hystad, P., Anenberg, S.C., 2019. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: estimates from global datasets. Lancet Planet. Health 3 (4), e166–e178.

Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., Smargiassi, A., 2014. Spatiotemporal modeling of ozone levels in Quebec (Canada): a comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy–LUR approaches. Environ. Health Perspect. 122 (9), 970.

Akita, Y., Baldasano, J.M., Beelen, R., Cirach, M., De Hoogh, K., Hoek, G., Nieuwenhuijsen, M., Serre, M.L., De Nazelle, A., 2014. Large scale air pollution estimation method combining land use regression and chemical transport modeling in a geostatistical framework. Environ. Sci. Technol. 48 (8), 4452–4459.

Amante, C., Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA.

Anenberg, S.C., Henze, D.K., Tinney, V., Kinney, P.L., Raich, W., Fann, N., Malley, C.S., Roman, H., Lamsal, L., Duncan, B., et al., 2018. Estimates of the global burden of ambient $PM_{2.5}$, ozone, and NO_2 on asthma incidence and emergency room visits. Environ. Health Perspect. 126 (10), 107004.

Beelen, R., Voogt, M., Duyzer, J., Zandveld, P., Hoek, G., 2010. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a dutch urban area. Atmos. Environ. 44 (36), 4614–4621.

Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B.D., Fiore, A.M., Li, Q., Liu, H.Y., Mickley, L.J., Schultz, M.G., 2001. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. J. Geophys. Res.: Atmos. 106 (D19), 23073–23095.

Brasche, S., Bischof, W., 2005. Daily time spent indoors in German homes–baseline data for the assessment of indoor exposure of German occupants. Int. J. Hygiene Environ. Health 208 (4), 247–253.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci. Total Environ. 253 (1–3), 151–167.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenthal, S., Coggins, J., Di, Q., Brunekreef, B., Frostad, J., Lim, S.S., Kan, H., Walker, K.D., Thurston, G.D., Hayes, R.B., Lim, C.C., Turner, M.C., Jerrett, M., Krewski, D., Gapstur, S.M., Diver, W.R., Ostro, B., Goldberg, D., Crouse, D.L., Martin, R.V., Peters, P., Pinault, L., Tjepkema, M., van Donkelaar, A., Villeneuve, P.J., Miller, A.B., Yin, P., Zhou, M., Wang, L., Janssen, N.A.H., Marra, M., Atkinson, R.W., Tsang, H., Quoc Thach, T., Cannon, J.B., Allen, R.T., Hart, J.E., Laden, F., Cesaroni, G., Forastiere, F., Weinmayr, G., Jaensch, A., Nagel, G., Concin, H., Spadaro, J.V., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc. Nat. Acad. Sci. 115 (38), 9592–9597.

Chauhan, A., Krishna, M., Frew, A., Holgate, S., 1998. Exposure to nitrogen dioxide (NO_2) and respiratory disease risk. Rev. Environ. Health 13 (1–2), 73–90.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., et al., 2019a. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ. Int. 130, 104934.

Chen, L., Bai, Z., Kong, S., Han, B., You, Y., Ding, X., Du, S., Liu, A., 2010. A land use regression for predicting NO_2 and PM_10 concentrations in different seasons in Tianjin region, China. J. Environ. Sci. 22 (9), 1364–1373.

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, pp. 785–794. ACM.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2019b. xgboost: Extreme Gradient Boosting. R package version (82), 1.

CNEMC 2019. China national environmental monitoring centre. http://www.cnemc.cn/. last accessed: Aug 23, 2019.

Collart, P., Dubourg, D., Levêque, A., Sierra, N.B., Coppieters, Y., 2018. Short-term effects of nitrogen dioxide on hospital admissions for cardiovascular disease in Wallonia, Belgium. Int. J. Cardiol. 255, 231–236.

Cuevas, C.A., Notario, A., Adame, J.A., Hilboll, A., Richter, A., Burrows, J.P., Saiz-Lopez, A., 2014. Evolution of NO_2 levels in Spain from 1996 to 2012. Scient. Rep. 4, 5887.

De Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., et al., 2018. Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe-Evaluation of spatiotemporal stability. Environ. Int. 120, 81–92.

de Hoogh, K., Gulliver, J., Van Donkelaar, A., Martin, R.V., Marshall, J.D., Bechle, M.J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., et al., 2016a. Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10.

de Hoogh, K., Gulliver, J., van Donkelaar, A., Martin, R.V., Marshall, J.D., Bechle, M.J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., et al., 2016b. Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. Roy. Meteorol. Soc. 137 (656), 553–597.

Dijkema, M.B., Gehring, U., van Strien, R.T., Van Der Zee, S.C., Fischer, P., Hoek, G., Brunekreef, B., 2010. A comparison of different approaches to estimate small-scale spatial variation in outdoor NO_2 concentrations. Environ. Health Perspect. 119 (5), 670–675.

Earthdata. GES DISC. last assessed May 21, 2019.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., de Nazelle, A., et al., 2012. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46 (20), 11195–11205.

European Environment Agency 2018. Air Quality e-Reporting (AQ e-Reporting). Last accessed 10, Sep. 2019.

François, C., 2017. Deep learning with python.

Friedman, J., Hastie, T., Tibshirani, R., 2010a. Regularization paths for generalized linear models via coordinate descent. J. Stat. Software 33 (1), 1.

Friedman, J., Hastie, T., Tibshirani, R., 2010b. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

Friedman, J., Hastie, T., Tibshirani, R., 2010c. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.

GDAL Development Team 2018. GDAL - Geospatial Data Abstraction Library. Accessed 1 May 2018.

Geddes, J.A., Martin, R.V., Boys, B.L., van Donkelaar, A., 2016. Long-term trends worldwide in ambient NO_2 concentrations inferred from satellite observations. Environ. Health Perspect. 124 (3), 281.

GEOS-CHEM 2019. GEOS-CHEM. <http://acmg.seas.harvard.edu/geos/>. last accessed:

Nov 22, 2019.
Greenwell, B., Boehmke, B., Cunningham, J., Developers, G., 2019. gbm: Generalized Boosted Regression Models. R package version 2 (1), 5.
Han, S., Bian, H., Feng, Y., Liu, A., Li, X., Zeng, F., Zhang, X., 2011. Analysis of the relationship between O_3, no and NO_2 in Tianjin, China. Aerosol Air Qual. Res 11 (2), 128–139.
Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 7561–7578.
Hoek, G., Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Boersma, K.F., Veefkind, P., 2015. Satellite NO_2 data improve national land use regression models for ambient NO_2 in a small densely populated country. Atmos. Environ. 105, 173–180.
Holmes, N.S., Morawska, L., 2006. A review of dispersion modelling and its application to the dispersion of particles: an overview of different dispersion models available. Atmos. Environ. 40 (30), 5902–5928.
Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., van Donkelaar, A., Lamsal, L., Martin, R., Jerrett, M., Demers, P., 2011. Creating national air pollution models for population exposure assessment in Canada. Environ. Health Perspect. 119 (8), 1123–1129.
Institute, H.E., 2010. Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects, number 17. Health Effects Institute.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning,, volume 112 Springer.
JRC 2015. GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network.
Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K., Bierkens, M.F.P., 2010. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. Environ. Model. Software 25 (4), 489–502.
Kerckhoffs, J., Hoek, G., Portengen, L., Brunekreef, B., Vermeulen, R.C., 2019. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. Environ. Sci. Technol. 53 (3), 1413–1421.
Knibbs, L.D., Hewson, M.G., Bechle, M.J., Marshall, J.D., Barnett, A.G., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ. Res. 135, 204–211.
Kuhn, M., 2018. caret: Classification and Regression Training. R package version 6.0-81.
Kwon, S.-B., Cho, Y., Park, D., Park, E.-Y., 2008. Study on the indoor air quality of Seoul metropolitan subway during the rush hour. Indoor Built Environ. 17 (4), 361–369.
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 (12), 6957–6964.
Lu, M., Schmitz, O., Vaartjes, I., Derek, K., 2019. Activity-based air pollution exposure assessment: differences between homemakers and cycling commuters. Health & Place 60, 102233. https://doi.org/10.1016/j.healthplace.2019.102233.
Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: comparison of estimation methods. Atmos. Environ. 42 (6), 1359–1369.
Mölter, A., Lindley, S., de Vocht, F., Simpson, A., Agius, R., 2010. Modelling air pollution for epidemiologic research—part i: A novel approach combining land use regression and air dispersion. Sci. Total Environ. 408 (23), 5862–5869.
Morgenstern, V., Zutavern, A., Cyrys, J., Brockow, I., Gehring, U., Koletzko, S., Bauer, C.-P., Reinhardt, D., Wichmann, H.-E., Heinrich, J., 2007. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup. Environ. Med. 64 (1), 8–16.
Novotny, E.V., Bechle, M.J., Millet, D.B., Marshall, J.D., 2011. National satellite-based land-use regression: NO_2 in the united states. Environ. Sci. Technol. 45 (10), 4407–4414. PMID: 21520942.
NSO and ESA 2019. TROPOMI. <http://www.tropomi.nl/>. last accessed: May 23, 2019.
OpenAQ. Fighting Air Inequality With Open Data and Community. last assessed May 1, 2019.
OpenStreetMap contributors 2019. Planet dump 7 Jan 2019 retrieved from https://planet.osm.org.
Park, Y.M., Kwan, M.-P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health & Place 43, 85–94.
Perrino, C., Pietrodangelo, A., Febo, A., 2001. An atmospheric stability index based on radon progeny measurements for the evaluation of primary urban pollution. Atmos. Environ. 35 (31), 5235–5244.
Picornell, M., Ruiz, T., Borge, R., García-Albertos, P., de la Paz, D., Lumbreras, J., 2019. Population dynamics based on mobile phone data to improve air pollution exposure assessments. J. Exposure Sci. Environ. Epidemiol. 29 (2), 278.
Python 2019. Python Software Foundation website. last accessed Sep 16, 2019.
Schweizer, C., Edwards, R.D., Bayer-Oglesby, L., Gauderman, W.J., Ilacqua, V., Jantunen, M.J., Lai, H.K., Nieuwenhuijsen, M., Künzli, N., 2007. Indoor time–microenvironment–activity patterns in seven regions of europe. J. Exposure Sci. Environ. Epidemiol. 17 (2), 170.
Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization paths for cox's proportional hazards model via coordinate descent. J. Stat. Softw. 39 (5), 1–13.
Sperling, D., Gordon, D., 2010. Two billion cars: driving toward sustainability. Oxford University Press.
TEMIS 2019. Nitrogen dioxide (NO_2) Background and applications. last assessed May 21, 2019.
TEMIS 2019. TEMIS air pollution monthly NO_2 column density. last assessed May 27, 2019.
Vienneau, D., De Hoogh, K., Beelen, R., Fischer, P., Hoek, G., Briggs, D., 2010. Comparison of land-use regression models between Great Britain and the Netherlands. Atmos. Environ. 44 (5), 688–696.
Wang, R., Henderson, S.B., Sbihi, H., Allen, R.W., Brauer, M., 2013. Temporal stability of land use regression models for traffic-related air pollution. Atmos. Environ. 64, 312–319.
Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77 (1), 1–17.
Xu, H., Bechle, M.J., Wang, M., Szpiro, A.A., Vedal, S., Bai, Y., Marshall, J.D., 2019. National PM2. 5 and NO2 exposure models for China based on land use regression, satellite measurements, and universal kriging. Sci. Total Environ. 655, 423–433.
Yoo, E.-H., Kyriakidis, P.C., 2006. Area-to-point kriging with inequality-type data. J. Geogr. Syst. 8 (4), 357–390.
Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. Environ. Sci. Technol. 50 (7), 3686–3694. PMID: 26927327.
Zhan, Y., Luo, Y., Deng, X., Zhang, K., Zhang, M., Grieneisen, M.L., Di, B., 2018. Satellite-based estimates of daily NO_2 exposure in China using hybrid random forest and spatiotemporal kriging model. Environ. Sci. Technol. 52 (7), 4180–4189.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) 67 (2), 301–320.