

Satellite-based estimation of NO₂ concentrations using a machine-learning model: A case study on Rio Grande do Sul, Brazil

Adriana BECERRA-RONDÓN^{1*}, Jorge DUCATI¹ and Rafael HAAG²

¹ Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre, RS, Brazil.

² Universidade Estadual do Rio Grande do Sul, Av. Bento Gonçalves, 8855, Porto Alegre, RS, Brazil.

*Corresponding author; email: abecerrarondon@gmail.com

Received: August 26, 2021; Accepted: March 10, 2022

RESUMEN

El dióxido de nitrógeno (NO₂) es uno de los contaminantes atmosféricos más importantes que afecta la salud humana (mayor propensión a infecciones respiratorias) y el medio ambiente (acidificación del suelo y agua). En muchas regiones de Brasil las mediciones de NO₂ a nivel del suelo presentan dificultades debido a que la red de estaciones de monitoreo es escasa y está desigualmente espaciada. Las observaciones satelitales combinadas con modelos de aprendizaje automático pueden mitigar esta falta de datos. Este artículo es el resultado de una investigación sobre la capacidad de un enfoque de aprendizaje automático (un algoritmo estadístico no lineal de bosques aleatorios, en adelante RF) para ejecutar una reconstrucción espacio-temporal de NO₂ a nivel del suelo, de 2006 a 2019, utilizando como parámetros de entrada datos de NO₂ recuperados del sensor del Instrumento de Monitoreo de Ozono (OMI) a bordo del satélite Aura, además de covariables meteorológicas y mediciones localizadas de NO₂ a nivel del suelo. Los resultados muestran que el modelo de RF predice el NO₂ con una precisión expresada por una correlación $R^2 = 0.68$ basada en una validación cruzada de 10 iteraciones. El modelo también predijo una concentración media de NO₂ de $18.73 \pm 3.86 \mu\text{g m}^{-3}$. La concentración total de NO₂ en toda la región analizada mostró una tendencia decreciente ($2.9 \mu\text{g m}^{-3} \text{yr}^{-1}$) entre 2006 y 2017. Este estudio demuestra que las estadísticas no lineales empleadas por el algoritmo de RF pueden ser herramientas complementarias a las observaciones in situ y por satélite para predecir NO₂.

ABSTRACT

Nitrogen dioxide (NO₂) is one of the most important atmospheric pollutants, affecting human health (increasing susceptibility to respiratory infections) and the environment (soil and water acidification). In many regions of Brazil, NO₂ measurements at ground level meet difficulties because monitoring stations are few and unevenly distributed. Satellite observations combined with machine learning models can mitigate this lack of data. This paper report results from an investigation on the ability of a machine learning approach (a non-linear statistical Random Forest algorithm, hereafter RF) to reconstruct the long-term spatiotemporal ground-level NO₂ from 2006 to 2019 using as input parameters NO₂ data retrieved from the Ozone Monitoring Instrument (OMI) sensor aboard Aura satellite, besides meteorological covariates and localized ground-level NO₂ measurements. Results show that the RF model predicts NO₂ with an accuracy expressed by an $R^2 = 0.68$ correlation based on a 10-fold cross-validation. The model also predicted a mean NO₂ concentration of $18.73 \pm 3.86 \mu\text{g m}^{-3}$. The total NO₂ concentration over the entire region analyzed showed a decreasing trend ($2.9 \mu\text{g m}^{-3} \text{yr}^{-1}$), being 2006 the year with the higher concentrations and 2017 with the lowest. This study suggests that non-linear statistical algorithm reconstructions using RF can be complementary tools to in situ and satellite observations for NO₂ mapping.

Keywords: nitrogen dioxide (NO₂), Random Forest algorithm, OMI sensor, southern Brazil.

1. Introduction

The industrial revolution changed dramatically the planet's atmospheric composition bringing air pollution and making it one of the biggest ongoing threats facing global public health. It is estimated that ~ 90% of the world's population lives in areas where levels of air pollution are above limits deemed safe for human health (WHO, 2018). Worldwide, these levels are influenced by pressures for economic development (combustion processes for energy generation using technologies not internationally recommended) and, regionally, by each area's topography and climate (Parra et al., 2009). This last factor, climate, acts from large to small scales, modulating emissions by increasing or decreasing them. The nature of emissions depends on their source and can be primary air pollutants (emitted directly into the air from a source) with direct impacts, or precursors for secondary air pollutants formed through reactions in the atmosphere (often in the presence of sunlight) such as oxides of nitrogen (Bimbaitė and Girždienė, 2007; Schnitzhofer et al., 2008).

There are many chemical species of nitrogen oxides (NO_x), but the air pollutant of most interest is nitrogen dioxide (NO_2), not only because of its effects on human health but also because (a) it absorbs visible solar radiation and contributes to impaired atmospheric visibility; (b) as an absorber of visible radiation it can have a potential direct role in global climate change if its concentrations become high enough; (c) it is, along with nitric oxide (NO), a chief regulator of the oxidizing capacity of the troposphere by controlling the build-up and fate of radical species, including hydroxyl radicals, and (d) it plays a key role in atmospheric chemistry related to ozone formation and destruction, this being the main vector to absorption of ultraviolet radiation reaching the planet's surface, whether in polluted or unpolluted atmospheres (WHO, 2000). As a significant contributor to air pollution, NO_2 is released into the atmosphere from both natural and anthropogenic sources, being the major ones fossil fuel combustion, biomass burning, lightning, and ammonia oxidation. It is estimated that the lifetime of NO_2 in the atmosphere varies from 6 h in summer to 18–24 h in winter (Seinfeld and Pandis, 1998; Bond et al., 2001; Beirle et al., 2003; Zhang et al., 2003). Its concentration level is influenced by

complex variables such as wind, temperature, burning material, besides policies and other anthropogenic factors of each city (e.g., urban areas) (Liu et al., 2016; Wang and Su, 2020).

The growing trend in the emission of atmospheric nitrogen dioxide (NO_2) has been a major concern around the world, the developing countries being at larger risk owing to fewer resources and socioeconomic vulnerability (Munzi et al., 2009). Traditionally, this trace gas is measured through a network of monitoring stations in the ground, but such an extensive network may not exist in all areas of interest and particularly in lower-income countries. An alternative approach to overcome these limitations is to use satellite data, which provide spatiotemporal data from regional to global scales under widely different atmospheric conditions and offer expeditious, accurate information about changes in air quality and their human impacts (Bais et al., 2007; Liu et al., 2016; Mostafa et al., 2021).

In South America, studies on air pollution are limited to a few cities, because air quality monitoring stations tend to be expensive, and not all air pollutants are monitored (Réquia et al., 2015; Fajersztajn et al., 2017; Silva et al., 2020). An example of the lack of an extensive network for environmental monitoring is the state of Rio Grande do Sul, located in south Brazil, with only nine stations within its borders. Compared to other regions of the country, it is exposed to higher levels of ultraviolet radiation for being closer to the Antarctic ozone hole (Kirchhoff et al., 2000; Guarnieri et al., 2004), and is strongly influenced by transient meteorological systems (cold and hot fronts), isentropic transport (between the tropical stratosphere reservoir, polar vortex, middle-latitude) and exchange processes (between upper-lower stratosphere) (Reboita et al., 2010; Bittencourt et al., 2019).

Recently, spaceborne observations were combined with statistical methods to produce spatiotemporal predictions on NO_2 concentrations, supplementing existing ground-based stations and providing coverage where ground-based data are unavailable (Hoek et al., 2015; Larkin et al., 2017; Qin et al., 2020). These predictions are based on machine learning models ranging from linear regression models to non-linear, being the latter the ones that perform better since they consider that there is a non-linear

relationship between predictor variables and dependent variables (Zhan et al., 2018a). Predictive variables of models for air pollutants usually include satellite data, meteorology, population density, elevation, and land-use type variables (Zhan et al., 2018b; Pan et al., 2021).

Random Forest (RF) is a non-linear model simple in structure, requires less computational resources, and has strong interpretation capabilities. Besides giving importance weights for each of the included variables to select the model predictors, it is also less prone to overfitting and less sensitive to outliers (Pan et al., 2021). Many of the NO₂ studies based on this model have focused on short and long-term spatiotemporal changes (Hoek et al., 2015; Araki et al., 2018; Zhan et al., 2018a, b; Kamińska, 2019; Zhu et al., 2019; Qin et al., 2020), showing good prediction performance, with cross-validation R² from 0.69 to 0.84. In addition, models in such studies exhibit higher accuracies when the dataset integrates temporal and spatial information.

Given that the existing environmental monitoring stations in Rio Grande do Sul are sparse and unevenly distributed, the objective of this study is to reconstruct the long-term spatiotemporal environmental NO₂ concentrations from 2006 to 2019 across the state, through the RF machine learning model, using as input parameters data retrieved from the Ozone Monitoring Instrument (OMI) sensor aboard Aura satellite, besides meteorological covariates and ancillary data.

2. Material and methods

2.1 Study area

Rio Grande do Sul is the Brazilian southernmost state, having international borders with Argentina to the west and Uruguay to the south (Fig. 1). Its area is 281 707 km², and with more than 11.5 million inhabitants it is the fifth most populated state in the country; about 85% of its population lives in urban areas (Fig. 2). The metropolitan area of its capital, Porto Alegre, concentrates an important fraction of the state's population and economic activities. Regarding the consumption of fossil fuels, the state has about 7.5 million vehicles (IBGE, 2021). The region has a humid subtropical climate with a large seasonal variation with hot summers and well-defined, cold winters. Mean temperatures vary from 15 to 18 °C,

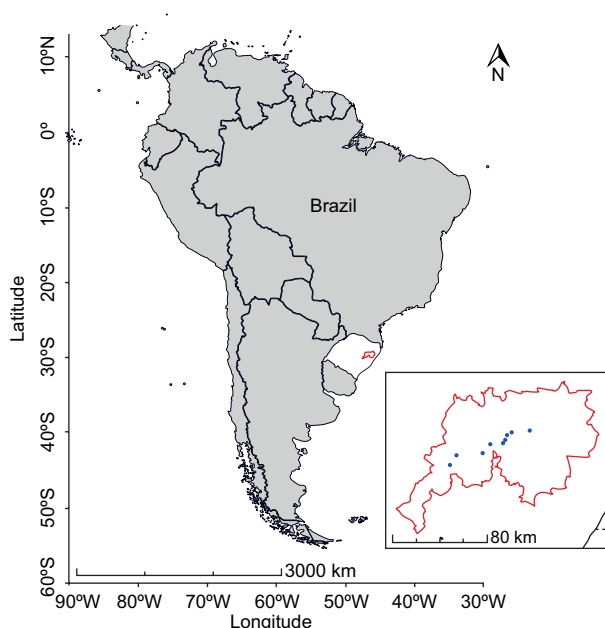


Fig. 1. Map with Rio Grande do Sul's (Brazil) location (in white) and its metropolitan area (red contour) and the positions of NO₂ monitoring sites (blue points).

with lows of as much as −10 °C (June and July) and highs going up to 40 °C (December to March) (Livi, 2002). The surface elevation ranges from sea level up to 1400 m, with the highest points northeast of the state (Fig. 2).

2.2 Data set

2.2.1 Ground-level NO₂ data

The ground-level NO₂ data was acquired by nine monitoring sites (Fig. 1) managed by the Fundação Estadual de Proteção Ambiental (the state's Environmental Protection Foundation, FEPAM), measured hourly by the chemiluminescence method from 2006 to 2019, with the occurrence of some discontinuous periods (FEPAM, 2021). For this research, we integrated the hourly average concentrations to daily values for all stations, and then all data were interpolated to a 0.25° × 0.25° grid. This spatial resolution was the standard in this study, and 420 such cells were necessary to cover the state. After this pre-processing, a total of 2229 daily measurements of ground-level NO₂ were used to develop to model. This number represents 43.5% of the time series of 14 years if data were acquired every day during the studied period.

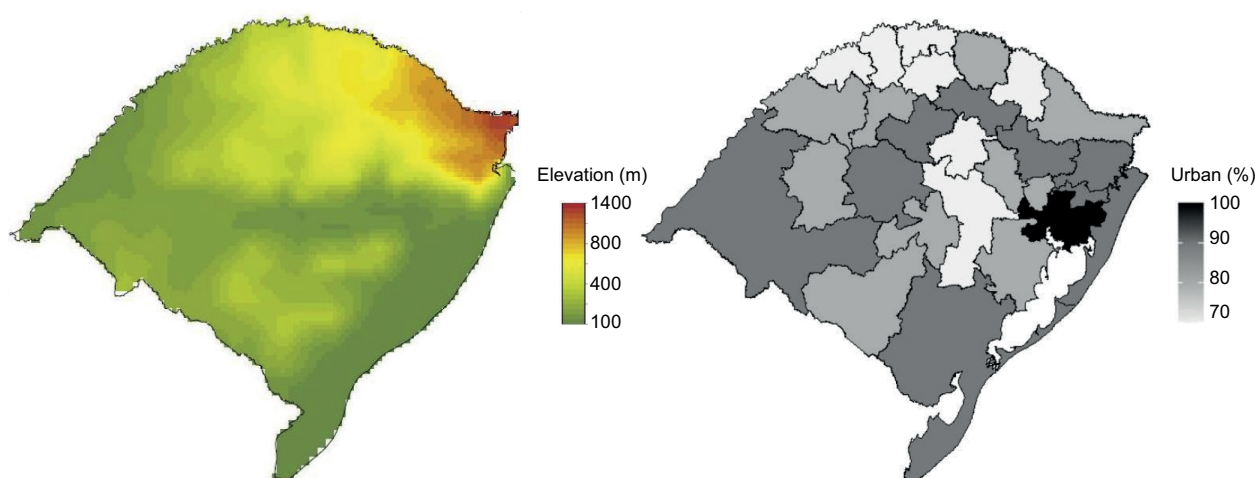


Fig. 2. Topographic and demographic representations of Rio Grande do Sul, Brazil. Elevation (left), urbanization (right) (COREDES, 2010).

2.2.2 Satellite NO_2 data

The orbital data used in this study came from the OMI sensor onboard the Aura satellite. This instrument is equipped with a spectrometer pointed to the nadir which measures the ultraviolet light (264–504 nm) coming from the Sun and back-scattered by the atmosphere. The Differential Optical Absorption Spectroscopy (DOAS) algorithm was developed to derive NO_2 (product OMNO2d) (Krotkov et al., 2006). The data provided in this product are expressed in molecules cm^{-2} units daily at $0.25^\circ \times 0.25^\circ$ (latitude/longitude) spatial resolution, these column densities being adequate to track spatial patterns for ground-level NO_2 . Therefore, it is unnecessary to convert vertical column density to ambient concentrations in the input model (Bechle et al., 2015). These data are represented in a time series of 14 years (2006–2019), with 5092 daily images (99% of the series) acquired from the data provider GES-DISC (NASA, 2021a).

2.2.3 Meteorological data

The data of meteorological covariates were acquired from the Modern-Era Retrospective analysis for Research and Applications (MERRA-2) produced by the NASA Global Modeling and Assimilation Office (GMAO), which provides a global atmospheric reanalysis beginning in 1980. These data were initially produced on a $0.5^\circ \times 0.666^\circ$ global grid and were re-gridded by the POWER project to the $0.5^\circ \times$

0.5° (latitude/longitude) bilinear interpolated spatial grid with daily resolution. The daily meteorological conditions used in the model include surface temperature, wind speed, humidity, surface pressure, all variables to 2 m (Rienecker et al., 2011). This data are represented in a time series of 14 years (2006–2019) with 5113 images for each variable with daily measurements (100% of the series) acquired from the data provider POWER (NASA, 2021b).

2.2.4 Ancillary data

To support our model, data on elevation were included on the form of 90 m vertical resolution measurements collected by the Shuttle Radar Topography Mission (STRM) and acquired from the Federal University of Rio Grande do Sul (UFRGS, 2021), from which a mean elevation $0.25^\circ \times 0.25^\circ$ grid was compiled.

Information on the erythemal daily dose and ozone were also added, provided by satellite OMI through product OMUVBG in J m^{-2} units with a daily spatial resolution of $0.25^\circ \times 0.25^\circ$ (latitude/longitude). Daily data from the product OMTO3d (ozone total column density) were also included, but with a $1.0^\circ \times 1.0^\circ$ (latitude/longitude) spatial resolution in Dobson units (where $1 \text{ DU} = 2.7 \times 10^{18} \text{ molecules O}_3 \text{ cm}^{-3}$) (Levelt et al., 2006; Tanskanen et al., 2006). These data are represented in a time series of 14 years (2006–2019) with 5092 daily images (99% of the series), acquired from the data provider GES-DISC (NASA, 2021a).

2.3 Statistical model

Predicting NO₂ concentration with sufficient spatial and temporal resolution and accuracy is not a simple task. Although satellite data can contribute to this problem, they are usually limited to a single measurement per day, and their units in vertical column density are difficult to compare with ground data (Liu et al., 2021). Nevertheless, there is a growing interest in combining ground-based data and satellite observations to produce information comparable with measurements at surface level (Lu et al., 2020).

Modeling concentrations, like those of NO₂ presently in focus, can be achieved through a diversity of techniques, among which we can mention the Gradient Boosting Decision Tree (GBTD) and the XGBoost algorithms, besides support vector machine (SVM) supervised learning algorithms. For this study, modeling using the RF technique was chosen due to its potential to direct implementation and for its robustness, being less prone to overfitting and less sensitive to outliers (Breiman, 2001). Therefore, it constitutes a viable option to be applied to complex atmospheric processes, which provides a spatiotemporal estimate of air pollutants by using a simple regression model (Qin et al., 2020; Chan et al., 2021). The RF regression model consists of a collection of regression trees trained by bootstrap samples that splits each node in the regression tree according to the best of a subset of randomly chosen predictors (Liaw and Wiener, 2002). Each tree is a regression function on its own, and the final output is the average of the individual tree outputs. In this study, from the collected dataset of ground-level NO₂ concentrations and various predictors such as satellite NO₂ retrievals and meteorological variables (temperature, wind speed, relative humidity, and atmospheric pressure), plus the ancillary data, ground NO₂ concentrations were simulated with the following expression:

$$\begin{aligned} NO_2Ground_{ij} \sim & NO_2Satellite_{ij} + Wind_{ij} + \\ & Temperature_{ij} + Pressure_{ij} + Humidity_{ij} + \\ & Erythemal\ Dose_{ij} + Ozone_{ij} + Elevation_j + \\ & Year + Julian\ day \end{aligned} \quad (1)$$

where j is the cell and i is the day.

This regression model was applied in three phases:

1. Training set, where we used 70% of the database with the ability to predict concentrations observed in monitoring sites, based on a set of predictors. At the end of this phase the relative importance of the predictors is directly provided by the RF model.
2. Performance, where, according to our objective of estimating ground-level NO₂ concentrations over places with no monitoring stations and from limited input data, we applied two validation methods: 10-fold cross-validation (over the training set), which is a re-sampling procedure used to evaluate if our model is under-fitting/over-fitting; and test set validation (with 30% of the database), used to the evaluation of the final model fit on the training dataset, with an assessment of the discrepancies between predictions and observations of the monitoring sites. These comparisons were summarized in terms of R^2 (percentage of explained variance), root mean square error (RMSE), and mean absolute error (MAE). The two latter metrics explain the differences between predicted and true values; the smaller the RMSE and MAE, the smaller the error between these two.
3. Generalization, which estimates the concentrations in the cells where no ground-level observations were available (Liaw and Wiener, 2002; Zhu et al., 2019; Gariazzo et al., 2020; Dou et al., 2021).

2.4. Analysis

The machine learning method used to estimate ground-level NO₂ concentrations in this study area is summarized in Figure 3, which shows collected and standardized multi-source data used to obtain the desired dataset with uniform spatial-temporal resolution ($0.25^\circ \times 0.25^\circ$ [latitude/longitude], daily resolution). The Pearson correlation coefficient was calculated for all predictors and ground observation data before running the model to evaluate the effect of these parameters.

To predict in all 420 cells through Random Forest, we used the R caret package (Kuhn, 2008; Wright y Ziegler, 2017), which is an interface that unifies several functions from different packages under the same framework, facilitating all the stages of pre-processing, training, optimization, and validation of predictive models. For the Random Forest function, the user can only adjust two parameters (ntrees and mtry), considered by developers to have the greatest

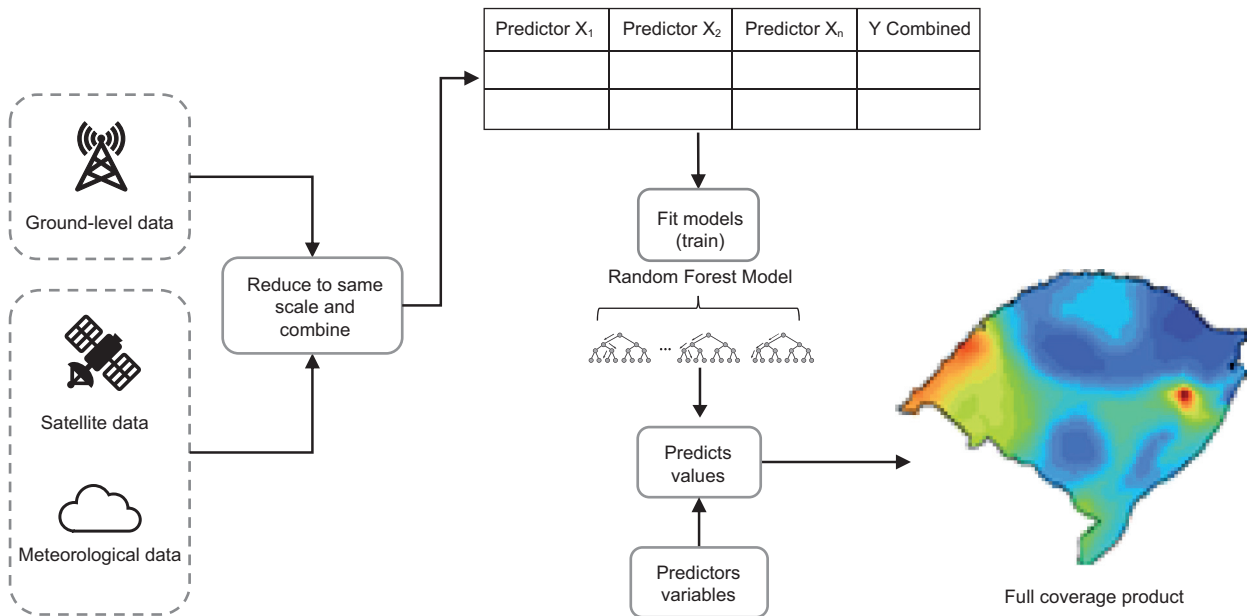


Fig. 3. Flow diagram of this study: daily ground-level NO₂ concentrations based on the Random Forest model.

effect on the accuracy of the model. Accordingly, we adjusted the number of trees (ntrees) and, to avoid a suboptimal number of variables (mtry), we indicated a random search on the training dataset for tuning. This random search was done by 10-fold cross-validation (random sampling method), which involved dividing the dataset into folds of equal sizes, performing the analysis on one subset (training), and validating on the other subset (validation). To reduce variability, multiple iterations of cross-validation were performed using different partitions, the results being combined (e.g. averaged) to give an estimate of the model tuning and performance. Based on these cross-validation results, we finally set for the final model the ntree to 500 and the mtry to 10. All maps have been produced with the R statistical software, version 4.0.5 (<http://R-project.org>).

The evaluation of the long-term trend (2006 to 2019) of ground-level NO₂ concentrations was made using the predicted values, calculating the cell averages for all 420 cells of the study area, for the whole period (historical series), as well as for month and season. The averages were calculated using the daily values for each of these periods, as follows:

$$Ave_{i=} = \frac{\sum_{t \text{ end date}}^{\text{start date}} Ave_{i,t}}{n} \quad (2)$$

where Ave_i is the average of the variable in each cell for the respective period; start date and end date correspond to the first and last date of each period; n is the number of days in the period.

The total average concentrations of ground-level NO₂ were evaluated by summarizing average concentrations across all cells that represent the State throughout each period.

3. Results and discussion

3.1 Model performance assessment

Relationships between ground-level NO₂ and each independent attribute are shown in Figure 4. The variables with the strongest correlation were: Year ($r = -0.44$), Satellite NO₂ ($r = 0.24$), Erythemat daily dose ($r = -0.21$), and Wind speed ($r = -0.17$). It can be noted that the dependent variable is influenced by interannual variations, dispersion, and photochemistry; however, variables with lower correlations also play a role in the model construction, as shown in Figure 5, which gives a perception on the contribution of variables to the final model. The categorical variable Year is the most important with a relative value of 1, followed by Elevation (0.5), Wind speed (0.26), Julian day (0.26), Erythemat daily dose (0.24), and sixth but not least, Satellite NO₂ (0.20). The remaining

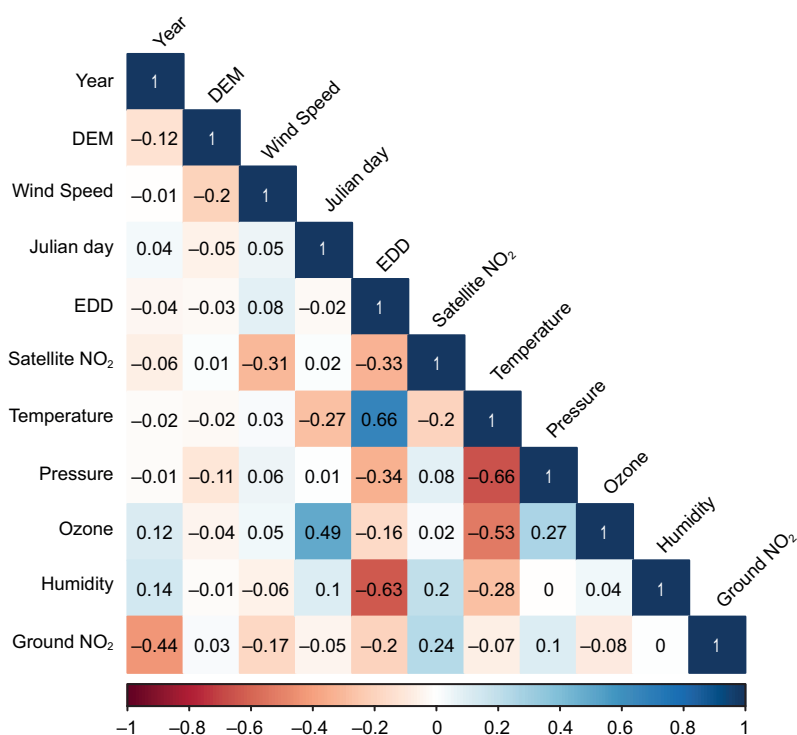


Fig. 4. Correlation matrix of the variables used in the model.

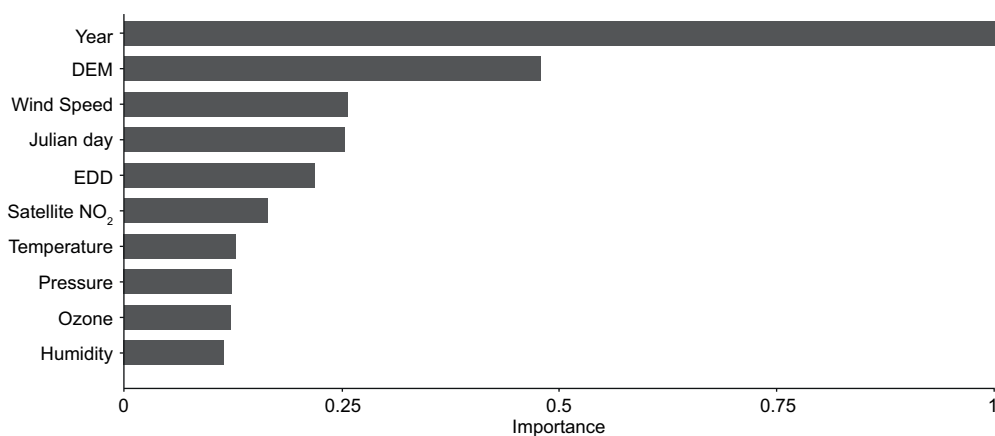


Fig. 5. Relative importance of predictor variables in the Random Forest model. Variables are listed in order of importance from top to bottom. The horizontal axis represents the measure of importance.

variables are ranked as less important. It has been reported that the satellite NO₂ (vertical column density) variable can be ranked as one of the most important in most models that use machine learning techniques to predict NO₂ (Araki et al., 2018; Zhu et al., 2019; Qin et al., 2020; Dou et al., 2021; Chan et al., 2021), but some of these studies have shown that

this importance varies between geographical areas and it may turn into one of the least important variables, contributing in some cases with only 1% of the prediction. This variation can occur for two reasons, the first one being due to the spatial resolution of this product since variations of NO₂ were averaged within each cell. The second reason is that the sensitivity

of the OMI sensor and of any satellite-based NO_2 measurement increases with altitude, implying that measurements at ground-level are less sensitive due to scattering of radiation at the surface and through the atmosphere (Lu et al., 2020; Penn and Holloway, 2020; Di et al., 2020).

Results expressing the model performance are presented in Figure 6 as a scatter plot, where the 10-fold cross-validation method yielded $R^2 = 0.68$, $\text{RMSE} = 6.43$, and $\text{MAE} = 4.37$. The test set validation method used to evaluate the model's performance yielded $R^2 = 0.70$, indicating the ability of the model to predict unknown data. The model metrics presently reported express a performance similar to those obtained by Qin et al. (2020) and Dou et al. (2021), who also adopted $0.25^\circ \times 0.25^\circ$ spatial resolutions to estimate ground-level NO_2 . We note, however, that in regression models the adopted spatial resolution of the predictors will depend on the objectives pursued, and for many applications data a finer ($\leq 1 \text{ km}^2$) spatial resolution is necessary to capture environmental variability that can be partly lost at lower resolutions (Hijmans et al., 2005).

In models like the one presently developed it has been shown that analysis of the spatiotemporal distribution of a variable with spatial dependencies can suffer from a misinterpretation of certain predictor variables (e.g., coordinates and elevation), an issue

that has been reported as a disadvantage for flexible algorithms (e.g., Random Forest) when predicting beyond the location of the training data (Meyer et al., 2019). In such cases, the predictive performance becomes more sensitive to the sample sizes or absolute values of dependent variables (Zhan et al., 2018a; Li et al., 2020). In the case presently investigated, Elevation is one of the most important variables to the final model, which was built from data by nine ground monitoring stations with elevations between 20 to 85 m, while the whole study area displays elevations between 0 and 1400 m; this fact possibly represents a limitation to this analysis, on an extent that has to be evaluated by further studies. However, an additional assessment of this model performance, in the form of a comparison between observed and predicted ground concentrations for the area covered by the nine monitoring stations, suggests that deviations due to elevation would have low weight, since, as it is presented in Figure 7a, the modeled concentrations closely follow the measurements over the studied period.

3.2 Spatial and temporal distribution of ground-level estimated NO_2

Table I presents the total estimated ground-level NO_2 concentrations over the entire state for the period 2006 to 2019. From the RF model, the predicted

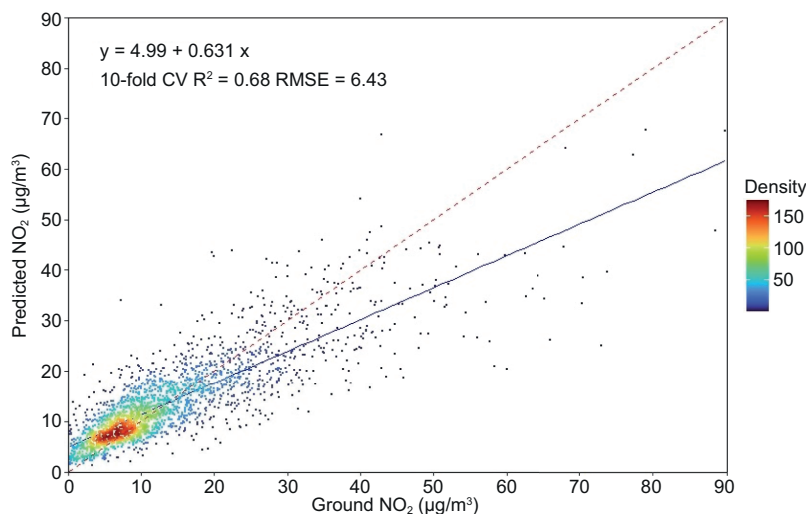


Fig. 6. Scatter plot result for the Random Forest model 10-fold cross-validation. The red dash line is the 1:1 reference line, the blue solid line is the trend line. The color bar indicates the number of data points.

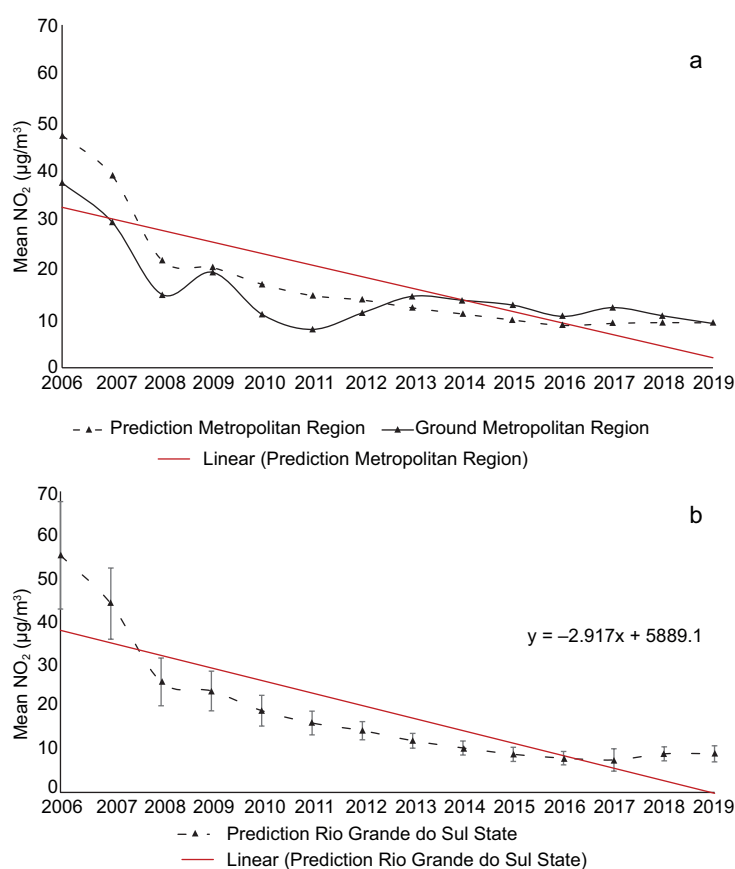


Fig. 7. Time series analysis of the annual mean surface NO₂ from 2006 to 2019 predicted by the model: (a) area covered by the ground monitoring stations; (b) all study area. Red lines are the trends for the period.

Table I. Historical modeled means of NO₂ from 2006 to 2019 for Rio Grande do Sul state, Brazil.

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
State	Min	18.14	18.28	9.64	9.84	8.97	8.40	8.34	8.14	6.54	6.13	5.56	5.48	4.63	3.25
	Max	65.92	50.64	31.28	28.38	23.42	19.88	17.42	17.27	16.49	15.84	14.64	20.07	14.28	13.61
	Mean	55.10	43.88	25.69	23.57	18.99	16.12	14.34	12.00	10.31	8.84	7.95	7.53	8.95	8.97
	SD	13.68	9.43	6.22	5.25	3.94	3.03	2.29	1.55	1.42	1.14	1.06	1.22	1.62	2.19

ground-level NO₂ concentrations across Rio Grande do Sul for this period have a mean of 18.73 µg m⁻³, a minimum of 8.67 µg m⁻³, and a maximum of 24.94 µg m⁻³, with a standard deviation of 3.86 µg m⁻³. From Table I and Figure 7b it can also be noted that the total modeled NO₂ concentrations over the entire state showed a decreasing trend from 2006 to 2019. This decreasing trend can be expressed by what we

presently call the interannual variability, which is the difference between the mean NO₂ for each year, minus the average for the whole period. Considering that 2006 has the highest mean concentrations, and 2017 the lowest, the interannual variability decreases almost steadily, going from as high as 36.37 µg m⁻³ in 2006 to as low as -11.20 µg m⁻³ in 2017. Considering that burning of biomass and fossil fuels is the main

source of NO_2 , this decrease could be at least partly explained, for the fossil fuels, by federal regulations of 1997 reinforced in 2008, making mandatory the installation of catalytic converters in all cars; for biomass burning, fires in crops and grazing fields were outlawed in 1992 in the State, and the legislation were slowly, but steadily enforced since then. We note that in Rio Grande do Sul state, forest fires are virtually non-existent. A final remark concerning this trend is that similar declines in NO_2 concentrations were observed in several countries, having as reference approximately the same periods (UBA, 2020; DEFRA, 2021; US-EPA, 2021), and for similar reasons.

Figures 8 to 10 show the spatial ground-level NO_2 distribution across the state in different time scales. Figure 8 shows the spatial distribution averaged over the entire 2006 to 2019 period, and to better understand this distribution we linked the information on elevation and percentage of urbanized areas in the state (Fig. 2) with the predicted concentrations in Figure 8. About elevation, we noted a pattern between of NO_2 concentrations with elevation: up to 100 m, concentrations are between $10\text{--}15\ \mu\text{g m}^{-3}$; from 400–600 m, between $15\text{--}25\ \mu\text{g m}^{-3}$; a decrease is observed between 600–800 m ($15\text{--}20\ \mu\text{g m}^{-3}$) and finally, from 800–1400 m the increase is continuous. It has been reported that, in general, higher altitudes show lower NO_2 concentrations (Zhu et al., 2019; Chan et al.,

2021), since in meteorological situations of inhibited wind transport/advection pollutants can accumulate in depressions (valleys and basins) (Giovannini et al., 2020). However, the atmospheric interactions with orography at different spatial scales influence the dispersion of atmospheric pollutants, being more complex in rough terrains. As a result, pollutants can be transported by orographic flows to mountainous areas, arriving to significant concentrations where they were not locally emitted. These constraints are reinforced by the fact, that, in Figure 8, higher NO_2 concentrations coincide either with urbanized, industrialized zones (some of which are situated in areas with elevations of about 800 m or higher), or with areas of intensive agriculture (where fertilizers are sources of NO_2 , and diesel-powered machines are heavy NO_2 producers); in some regions, both factors are summed up (Guo et al., 2020; Ai et al., 2021).

NO_2 concentrations are essentially a marker for combustion-related emissions and population, the latter having a direct relationship with urbanization levels. In this study, an inverse relationship was observed of the distribution of NO_2 concentrations with urbanization levels, highlighting the highest concentrations in less urbanized areas $\leq 70\%$, with $15\text{--}20\ \mu\text{g m}^{-3}$ averages. This apparently unexpected result is possibly due to the fact that some of the less urbanized areas are dominated by intensive agriculture, which become the major silent source of NO_x pollution, and where soil microbes convert fertilizer to nitrogen oxides, emitting about as much gases as vehicles. The emissions of these biogenic sources are larger in areas with high N fertilizer applications (Oikawa et al., 2015; Almaraz et al., 2018). In Rio Grande do Sul, Mossmann Koch et al. (2018) also highlighted that even apparently uncontaminated sites, such as small rural cities with little urbanization and family farming as main economic activity, present higher or similar concentrations than urbanized and industrialized areas. In addition, combustion emissions in rural regions are dominated by biomass burning (e.g., human-initiated burning for crop preparation) (Bechle et al., 2011).

With respect to monthly and seasonal distributions (Figs. 9 and 10), these exhibit behavior similar to other studies, with lowest concentrations in January with a mean of $15.65\ \mu\text{g m}^{-3}$ and standard deviation of $2.62\ \mu\text{g m}^{-3}$ (Table II). In spring and summer

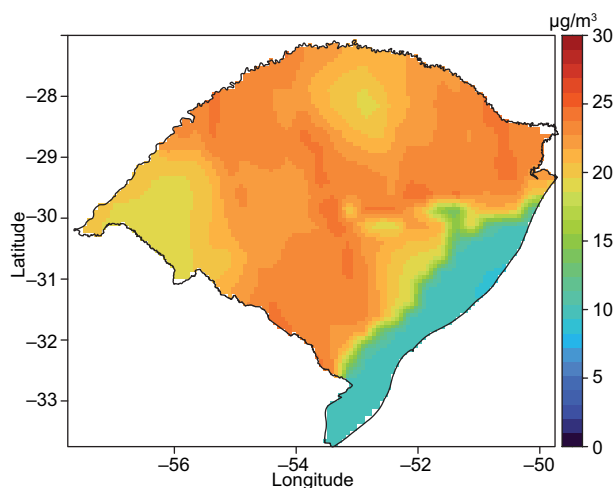


Fig. 8. Satellite-derived historical mean NO_2 concentrations ($\mu\text{g m}^{-3}$) throughout Rio Grande do Sul from 2006 to 2019.

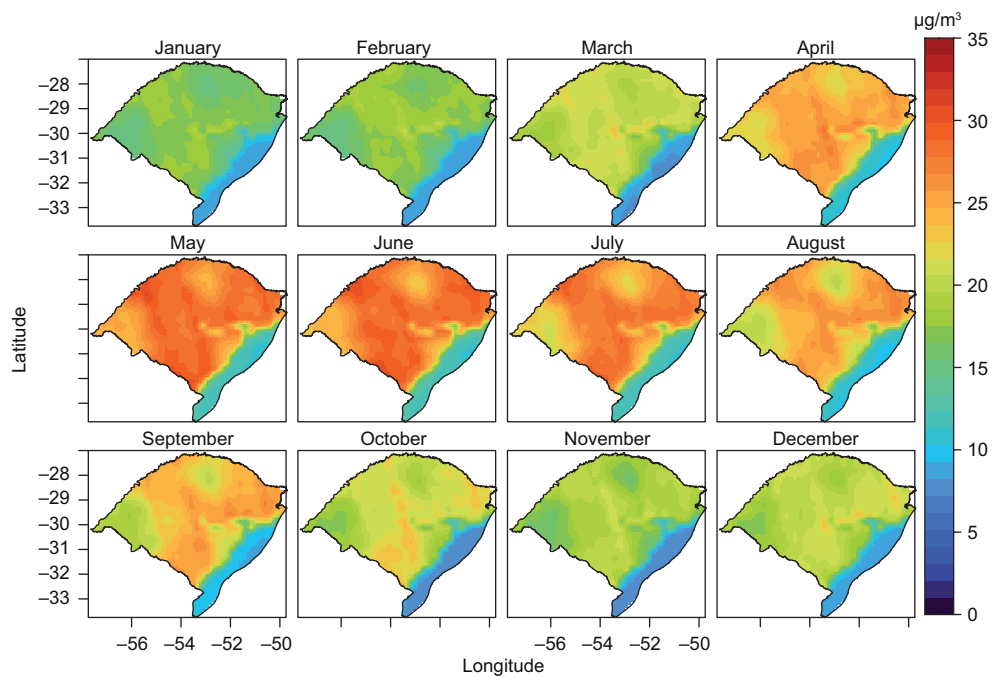


Fig. 9. Satellite-derived monthly mean NO₂ concentrations (µg m⁻³) throughout Rio Grande do Sul from 2006 to 2019.

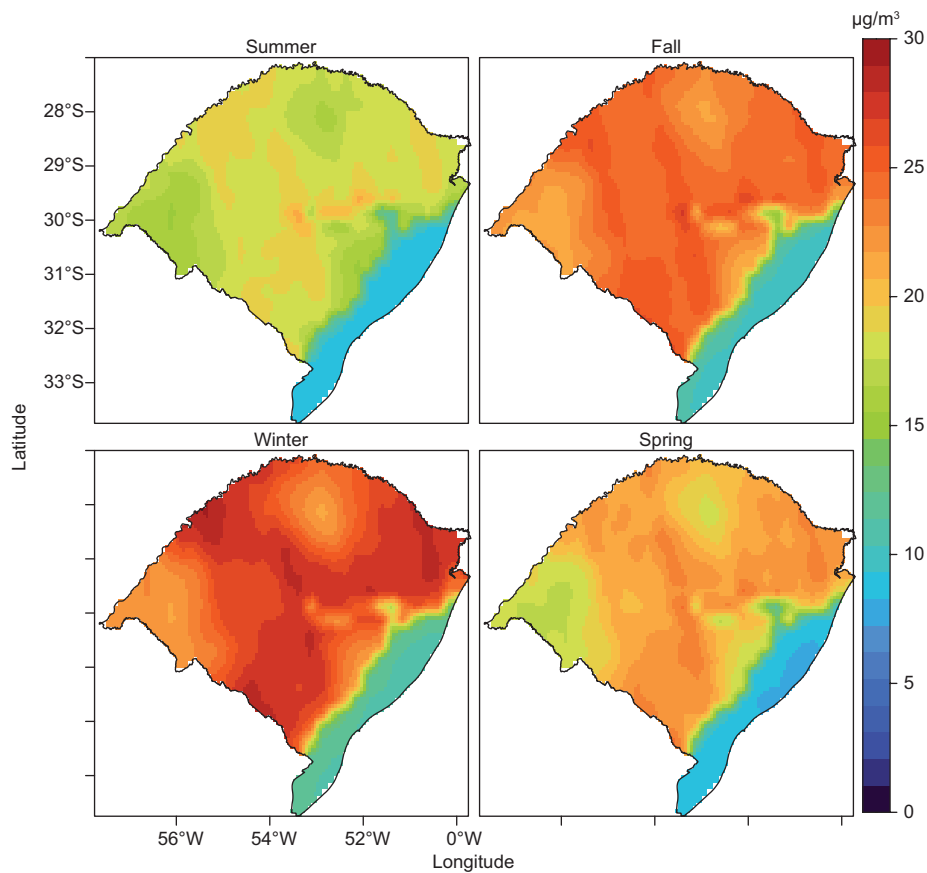


Fig. 10. Satellite-derived season mean NO₂ concentrations (µg m⁻³) throughout Rio Grande do Sul from 2006 to 2019.

Table II. Monthly means of NO₂ from 2006 to 2019 for Rio Grande do Sul state, Brazil.

		Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
State	Min	8.79	8.28	7.95	10.02	10.82	11.04	11.00	9.75	9.02	7.44	7.72	8.56
	Max	19.79	20.47	23.78	28.23	31.03	31.14	29.50	27.72	26.94	24.26	22.35	23.43
	Mean	15.65	16.08	18.66	22.65	25.44	25.39	24.27	22.20	21.51	19.01	17.32	18.47
	SD	2.62	2.98	3.98	4.57	5.24	5.22	4.94	4.68	4.74	4.44	3.69	3.73

Table III. Season means of NO₂ from 2006 to 2019 for Rio Grande do Sul state, Brazil.

		Summer	Autumn	Winter	Spring
State	Min	8.58	9.63	10.64	8.09
	Max	21.17	26.97	29.30	23.97
	Mean	16.75	22.12	23.93	19.22
	SD	3.09	4.55	4.91	4.26

concentrations assumed lower values, the latter with the lowest value (mean $16.75 \mu\text{g m}^{-3}$) (Table III). Highest concentrations were observed in May with a mean of $25.44 \mu\text{g m}^{-3}$ and standard deviation of $5.24 \mu\text{g m}^{-3}$ (Table II), while in fall and winter high concentrations were observed, the latter with a higher value (mean of $23.93 \mu\text{g m}^{-3}$) (Table III). In this context, due to its location, Rio Grande do Sul is under the action of a climatic complex whose action in spring and summer has the predominance of a tropical continental mass (a tropical Atlantic and a polar Atlantic mass), so that high humidity and high temperatures make the climatic conditions warmer, resulting in episodes of intense atmospheric convection and frequent precipitation which lead to the dispersion and wet deposition of NO₂, reducing air pollution. Additionally, in summer, with longer sunshine durations along with higher temperatures and evaporation, hydroxide (OH) concentrations are boosted, leading to a lower lifetime of NO₂, which becomes nitric acid (HNO₃). In contrast, in autumn and winter, essentially under the control of equatorial Atlantic, tropical Atlantic, continental tropical, and polar masses, there is an intensification of the action of cold fronts in the state, enhancing cold weather and reducing humidity (less precipitation and higher atmospheric pressure) with respect to other seasons, conditions that are unfavorable for pollution dispersion (Rossato, 2011; Zhu et al., 2019; Qin et al., 2020; Dou et al., 2021).

4. Conclusions

In this study, the use of Random Forest method, in combination with extensive data collection on multiple parameters from satellite and ground sources, was shown to be a valid tool for predicting NO₂ ground level concentrations at low spatial and daily temporal resolutions in Rio Grande do Sul, Brazil. However, other modeling techniques are available, and their performance will be assessed in future studies.

The modeled NO₂ seasonal variations at troposphere level displayed a higher concentration in winter, followed by autumn and spring, while summer had the lowest values during the 14 years covered by the study. In this context, such spatio-temporal trend is the combined effect of meteorological, geographical, and anthropogenic factors that, together, determine the increase or decrease of their concentrations.

Although the accuracy of modeled ground-level NO₂ still needs to be improved, especially considering elevation variations, the spatial and temporal estimates presented here can allow the formulation, planning, and implementation of environmental policies considering present and future emission sources. In addition, we highlight the need for designing studies aimed at investigating both short-term and long-term air pollution, not only in major cities but also in suburban and rural areas.

Acknowledgments

ABR acknowledges the Brazilian agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for her doctoral fellowship.

References

- Agudelo-Castañeda D, Calesso E, Norte F. 2014. Time-series analysis of surface ozone and nitrogen oxides concentrations in an urban area at Brazil. *Atmospheric Pollution Research* 5: 411-420. <https://doi.org/10.5094/APR.2014.048>
- Ai Y, Ge Y, Ran Z, Li X, Xu Z, Chen Y, Miao X, Xu X, Mao H, Shi Z, Jin T. 2021. Quantifying air pollutant emission from agricultural machinery using surveys – A case study in Anhui, China. *Atmosphere* 12: 440. <https://doi.org/10.3390/atmos12040440>
- Almaraz M, Bai E, Wang C, Trousdell J, Conley S, Faloona I, Houlton B. 2018. Agriculture is a major source of NO_x pollution in California. *Science Advances* 4: 1-8. <https://doi.org/10.1126/sciadv.aao3477>
- Araki S, Shima M, Yamamoto K. 2018. Spatiotemporal land use random forest model for estimating metropolitan NO₂ exposure in Japan. *Science of The Total Environment* 634: 1269-1277. <https://doi.org/10.1016/j.scitotenv.2018.03.324>
- Bais AF, Lubin D, Arola A, Bernhard G, Blumthaler M, Chubarova N, Erlick C, Gies HP, Krotkov NA, Lantz K, Mayer B, McKenzie RL, Piacentini RD, Seckmeyer G, Slusser JR, Zerefos CS. 2007. Surface ultraviolet radiation: Past, present, and future. In: Scientific assessment of ozone depletion: 2006. World Meteorological Organization, Geneva.
- Bechle M, Millet D, Marshall J. 2011. Effects of income and urban form on urban NO₂: Global evidence from satellites. *Environmental Science & Technology* 45: 4914-4919. <https://doi.org/10.1021/es103866b>
- Bechle M, Millet D, Marshall J. 2015. National spatiotemporal exposure surface for NO₂: Monthly scaling of a satellite-derived land-use regression, 2000-2010. *Environmental Science & Technology* 49: 12297-12305. <https://doi.org/10.1021/acs.est.5b02882>
- Beirle S, Platt U, Wenig, M, Wagner, T. 2003. Weekly cycle of NO₂ by GOME measurements: A signature of anthropogenic sources. *Atmospheric Chemistry and Physics* 3: 2225-2232. <https://doi.org/10.5194/acp-3-2225-2003>
- Bimbaitė V, Girgždienė R. 2007. Evaluation of Lithuanian air quality monitoring data applying synoptical analysis. *Journal of Environmental Engineering and Landscape Management* 15: 173-181. <https://doi.org/10.3846/16486897.2007.9636926>
- Bittencourt G, Pinheiro D, Bageston J, Bencherif H, Steffene L, Vaz Peres L. 2019. Investigation of the behavior of the atmospheric dynamics during occurrences of the ozone hole's secondary effect in southern Brazil. *Annals of Geophysics* 37: 1049-1061. <https://doi.org/10.5194/angeo-37-1049>
- Bond D, Zhang R, Tie X, Brasseur G, Huffines G, Orville R, Boccippio D. 2001. NO_x production by lightning over the continental United States. *Journal of Geophysical Research: Atmospheres* 106: 27701-27710. <https://doi.org/10.1029/2000JD000191>
- Breiman L. Random Forests. 2001. *Machine Learning* 45: 5-32. <https://doi.org/10.1023/A:1010933404324>
- Chan K, Khorsandi E, Liu S, Baier F, Valks P. 2021. Estimation of surface NO₂ concentrations over Germany from TROPOMI satellite observations using a machine learning method. *Remote Sensing* 13: 969. <https://doi.org/10.3390/rs13050969>
- CONAMA. 2018. Padrões de qualidade do ar (Resolução nº 491/2018). Conselho Nacional do Meio Ambiente. Diário Oficial da União 223: 155-156.
- COREDES. 2010. Taxa de urbanização. Conselhos Regionais de Desenvolvimento. Available at: <https://atlassocioeconomico.rs.gov.br/grau-de-urbanizacao> (accessed on July 19, 2021).
- DEFRA. 2021. National statistics: Air quality statistics in the UK, 1987 to 2020 – Nitrogen dioxide (NO₂). Department for Environment, Food and Rural Affairs. Available at: <https://www.gov.uk/government/statistics/air-quality-statistics/nitrogen-dioxide> (accessed on July 19, 2021).
- Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, Sabath MB, Choirat C, Koutrakis P, Lyapustin A, Wang Y. 2020. Assessing NO₂ concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environmental Science & Technology* 54: 1372-1384. <https://doi.org/10.1021/acs.est.9b03358>
- Dou X, Liao C, Wang H, Huang Y, Tu Y, Huang X, Peng Y, Zhu B, Tan J, Deng Z, Wu N, Sun T, Ke P, Liu Z. 2021. Estimates of daily ground-level NO₂ concentrations in China based on Random Forest model integrated K-means. *Advances in Applied Energy* 2: 100017. <https://doi.org/10.1016/j.adapen.2021.100017>

- EMBRAPA. 2003. Climate of Rio Grande do Sul. Section of Geography. Edited by the Brazilian Agricultural Research Corporation, Secretary of Agriculture, Porto Alegre (in Portuguese).
- Fajersztajn L, Saldiva P, Pereira L, Leite V, Buehler A. 2017. Short-term effects of fine particulate matter pollution on daily health events in Latin America: A systematic review and meta-analysis. *International Journal of Public Health* 62: 729-738. <https://doi.org/10.1007/s00038-017-0960-y>
- FEPAM. 2021. Rede estadual de monitoramento automático da qualidade do ar. Fundação Estadual de Proteção Ambiental Henrique Luís Roessler. Available at: <http://www.fepam.rs.gov.br/qualidade/relatorios.asp> (accessed on March 19, 2021).
- Gariazzo C, Carlino G, Silibello C, Renzi M, Finardi S, Pepe P, Radice P, Forastiere F, Michelozzi P, Viegli G, Stafoggia M. 2020. A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data. *Science of The Total Environment* 724: 138102. <https://doi.org/10.1016/j.scitotenv.2020.138102>
- Giovannini L, Ferrero E, Karl T, Rotach M, Staquet C, Trini Castelli S, Zardi D. 2020. Atmospheric pollutant dispersion over complex terrain: Challenges and needs for improving air quality measurements and modeling. *Atmosphere* 11: 646. <https://doi.org/10.3390/atmos11060646>
- Guarnieri RA, Guarnieri FL, Contreira DB, Padilha LF, Echer E, Pinheiro DK, Schuch AMP, Makita K, Schuch NJ. 2004. Ozone and UV-B radiation anticorrelations at fixed solar zenith angles in southern Brazil. *Geofísica Internacional* 43: 17-22. <https://doi.org/10.22201/igeof.00167169p.2004.43.1.209>
- Guo X, Wu H, Chen D, Ye Z, Shen Y, Liu J, Cheng S. 2020. Estimation and prediction of pollutant emissions from agricultural and construction diesel machinery in the Beijing-Tianjin-Hebei (BTH) region, China. *Environmental Pollution* 260: 113973. <https://doi.org/10.1016/j.envpol.2020.113973>
- Hijmans R, Cameron S, Parra J, P, Jarvis A. 2005. Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978. <https://doi.org/10.1002/joc.1276>
- Hoek G, Eeftens M, Beelen R, Fischer P, Brunekreef B, Folkert K, Peijnen B. 2015. Satellite NO₂ data improve national land use regression models for ambient NO₂ in a small densely populated country. *Atmospheric Environment* 105: 173-180. <https://doi.org/10.1016/j.atmosenv.2015.01.053>
- IBGE. 2021. Cidades. Rio Grande do Sul. Instituto Brasileiro de Geografia e Estatística. Available at: <http://www.cidades.ibge.gov.br> (accessed on August 10, 2021).
- Kamińska JA. 2019. A random forest partition model for predicting NO₂ concentrations from traffic flow and meteorological conditions. *Science of The Total Environment* 651: 475-483. <https://doi.org/10.1016/j.scitotenv.2018.09.196>
- Kirchhoff VWJH, Echer E, Leme NP, Silva AA. 2000. A variação sazonal da radiação ultravioleta solar biologicamente ativa. *Revista Brasileira de Geofísica* 18: 63-74. <https://doi.org/10.1590/S0102-261X2000000100006>
- Krotkov N, Carn S, Krueger A, Bhartia P, Yang K. 2006. Band residual difference algorithm for retrieval of SO₂ from the Aura Ozone Monitoring Instrument (OMI). *IEEE Transactions on Geosciences and Remote Sensing* 44: 1259-1266. <https://doi.org/10.1109/TGRS.2005.861932>
- Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28: 1-26. <https://doi.org/10.18637/jss.v028.i05>
- Larkin A, Geddes J, Martin R, Xiao Q, Liu Y, Marshall J, Brauer M, Hystad P. 2017. Global land use regression model for nitrogen dioxide air pollution. *Environmental Science & Technology* 51: 6957-6964. <https://doi.org/10.1021/acs.est.7b01148>
- Levelt P, van Den Oord G, Dobber M, Mälkki A, Visser H, de Vries J, Stammes P, Lundell J, Saari H. 2006. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing* 44: 1093-1101. <https://doi.org/10.1109/TGRS.2006.872333>
- Li R, Cui L, Liang J, Zhao Y, Zhang Z, Fu H. 2020. Estimating historical SO₂ level across the whole China during 1973-2014 using random forest model. *Chemosphere* 247: 125839. <https://doi.org/10.1016/j.chemosphere.2020.125839>
- Liaw A, Wiener M. 2002. Classification and regression by random Forest. *R News* 2: 18-22.
- Liu F, Zhang Q, van der A RJ, Zheng B, Tong D, Yan L, Zheng Y, He K. 2016. Recent reduction in NO_x emissions over China: Synthesis of satellite observations and emission inventories. *Environmental Research Letters* 11: 114002. <https://doi.org/10.1088/1748-9326/11/11/114002>

- Liu Q, Harris JT, Chiu LS, Sun D, Houser PR, Yu M, Duffy DQ, Little MM, Yang C. 2021. Spatiotemporal impacts of COVID-19 on air pollution in California, USA. *Science of The Total Environment* 750: 141592. <https://doi.org/10.1016/j.scitotenv.2020.141592>.
- Livi F. 2002. O clima em Porto Alegre no século XX: uma análise de séries temporais. M.Sc. thesis, Federal University of Rio Grande do Sul.
- Lu M, Schmitz O, de Hoogh K, Kai Q, Karssenberg D. 2020. Evaluation of different methods and data sources to optimise modelling of NO₂ at a global scale. *Environment International* 142: 105856. <https://doi.org/10.1016/j.envint.2020.105856>
- Meyer H, Reudenbach C, Wöllauer S, Nauss T. 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling* 411: 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
- Mossmann Koch N, Lucheta F, Käffer M, Martins S, Ferrão V. 2018. Air quality assessment in different urban areas from Rio Grande do Sul State, Brazil, using lichen transplants. *Annals of the Brazilian Academy of Sciences* 90: 2233-2248. <https://doi.org/10.1590/0001-3765201820170987>
- Mostafa MK, Gamal G, Wafiq A. 2021. The impact of COVID 19 on air pollution levels and other environmental indicators – A case study of Egypt. *Journal of Environmental Management* 277: 111496. <https://doi.org/10.1016/j.jenvman.2020.111496>
- Munzi S, Pisani T, Loppi S. 2009. The integrity of lichen cell membrane as a suitable parameter for monitoring biological effects of acute nitrogen pollution. *Ecotoxicology and Environmental Safety* 72: 2009-2012. <https://doi.org/10.1016/j.ecoenv.2009.05.005>
- NASA. 2021a. Earth Observatory, National Aeronautics and Space Administration. Available at: <https://urs.earthdata.nasa.gov> (accessed on March 10, 2021).
- NASA. 2021b. Earth Observatory, National Aeronautics and Space Administration. Available at: <https://power.larc.nasa.gov/> (accessed on May 01, 2021).
- Oikawa P, Ge C, Wang J, Eberwein J, Liang L, Allsman L, Grantz D, Jenerette G. 2015. Unusually high soil nitrogen oxide emissions influence air quality in a high-temperature agricultural region. *Nature Communications* 6: 8753. <https://doi.org/10.1038/ncomms9753>
- Pan Y, Zhao C, Liu Z. 2021. Estimating the daily NO₂ concentration with high spatial resolution in the Beijing-Tianjin-Hebei region using an ensemble learning model. *Remote Sensing* 13: 758. <https://doi.org/10.3390/rs13040758>
- Parra MA, Elustondo D, Bermejo R, Santamaría JM. 2009. Ambient air levels of volatile organic compounds (VOC) and nitrogen dioxide (NO₂) in a medium size city in Northern Spain. *Science of The Total Environment* 407: 999-1009. <https://doi.org/10.1016/j.scitotenv.2008.10.032>
- Penn E, Holloway T. 2020. Evaluating current satellite capability to observe diurnal change in nitrogen oxides in preparation for geostationary satellite missions. *Environmental Research Letters* 15: 034038. <https://doi.org/10.1088/1748-9326/ab6b36>
- Qin K, Han X, Li D, Xu J, Loyola D, Xue Y, Zhou X, Li D, Zhang K, Yuan L. 2020. Satellite-based estimation of surface NO₂ concentrations over east-central China: A comparison of POMINO and OMNO2d data. *Atmospheric Environment* 224: 117322. <https://doi.org/10.1016/j.atmosenv.2020.117322>
- Reboita MS, Gan MA, Rocha RPD, Ambrizzi T. 2010. Regimes de precipitação na América do Sul: uma revisão bibliográfica. *Revista Brasileira de Meteorologia* 25: 185-204. <https://doi.org/10.1590/S0102-77862010000200004>
- Réquia W, Koutrakis P, Roig H. 2015. Spatial distribution of vehicle emission inventories in the Federal District, Brazil. *Atmospheric Environment* 112: 32-39. <https://doi.org/10.1016/j.atmosenv.2015.04.029>
- Rienecker M, Suarez M, Gelaro R, Todling R, Bacmeister J, Liu E, Woollen J. 2011: MERRA –NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate* 24: 3624-3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- Rossato MS. 2011. Os climas do Rio Grande do Sul: variabilidade, tendências e tipologia. Ph.D. thesis, Federal University of Rio Grande do Sul.
- Seinfeld J, Pandis S. 1998. Atmospheric chemistry and physics: From air pollution to climate change. Wiley Interscience, New York.
- Silva J, Honscha J, Brum R, Ramires P, Tavella R, Fernandes C, Penteado J, Bonifácio A, Volcão L, Santos M, Coronas M. 2020. Air quality in cities of the extreme south of Brazil. *Ecotoxicology and Environmental Contamination* 15: 61-67. <https://doi.org/10.5132/eec.2020.01.08>
- Schnitzhofer R, Beauchamp J, Dunkl J, Wisthaler A, Weber A, Hansel A. 2008. Long-term measurements

- of CO, NO, NO₂, benzene, toluene and PM₁₀ at a motorway location in an Austrian valley. *Atmospheric Environment* 42: 1012-1024. <https://doi.org/10.1016/j.atmosenv.2007.10.004>.
- Tanskanen A, Krotkov NA, Herman JR, Arola A. 2006. Surface ultraviolet irradiance from OMI. *IEEE Transactions on Geoscience and Remote Sensing* 44: 1267-1271. <https://doi.org/10.1109/TGRS.2005.862203>
- UBA. 2020. Air quality 2019. German Environment Agency. Available at: <https://www.umweltbundesamt.de/en/press/pressinformation/air-quality-2019-trend-in-no2-decline-continues> (accessed on August 10, 2021).
- UFRGS. 2021. Shuttle Radar Topography Mission data. Federal University of Rio Grande do Sul. Available at: <https://www.ufrgs.br/labgeo/> (accessed on May 1, 2021).
- US-EPA. 2021. Nitrogen dioxide trends. United States Environmental Protection Agency. Available at: <https://www.epa.gov/air-trends/nitrogen-dioxide-trends> (accessed on July 31, 2021).
- Wang Q, Su M. 2020. A preliminary assessment of the impact of COVID-19 on environment – A case study of China. *Science of The Total Environment* 728: 138915. <https://doi.org/10.1016/j.scitotenv.2020.138915>
- WHO. 2000. Air quality guidelines. World Health Organization. Available at: https://www.euro.who.int/__data/assets/pdf_file/0017/123083/AQG2ndEd_7_1nitro-gendioxide.pdf/ (accessed on June 10, 2021).
- WHO. 2005. Air quality guidelines – Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. World Health Organization, Geneva.
- WHO. 2018. Air pollution. World Health Organization, Geneva. Available at: <http://www.who.int/airpollution/> (accessed on August 10, 2021).
- Wright MN, Ziegler A. 2017. Ranger: A fast implementation of Random Forest for high dimensional data in C⁺⁺ and R. *Journal of Statistical Software* 77: 1-17. <https://doi.org/10.18637/jss.v077.i01>
- Zhan Y, Luo Y, Deng X, Zhang K, Zhang M, Grieneisen M, Di B. 2018a. Satellite-based estimates of daily NO₂ exposure in China using hybrid random forest and spatiotemporal kriging model. *Environmental Science & Technology* 52: 4180-4189. <https://doi.org/10.1021/acs.est.7b05669>
- Zhan Y, Luo Y, Deng X, Grieneisen M, Zhang M, Di B. 2018b. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environmental Pollution* 233: 464e473. <https://doi.org/10.1016/j.envpol.2017.10.029>
- Zhang R, Tie X, Bond D. 2003. Impacts of anthropogenic and natural NO_x sources over the U.S. on tropospheric chemistry. *Proceedings of the National Academy of Sciences* 100: 1505-1509. <https://doi.org/10.1073/pnas.252763799>
- Zhu Y, Zhan Y, Wang B, Li Z, Qin Y, Zhang K. 2019. Spatiotemporally mapping of the relationship between NO₂ pollution and urbanization for a megacity in Southwest China during 2005-2016. *Chemosphere* 220: 155-162. <https://doi.org/10.1016/j.chemosphere.2018.12.095>