https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

A machine learning approach to address air quality changes during the COVID-19 lockdown in Buenos Aires, Argentina

Melisa Diaz Resquin[1,2,3,*], Pablo Lichtig[1,4], Diego Alessandrello[1], Marcelo De Oto[1], Darío Gómez[1,2], Cristina Rössler[1,5], Paula Castesana[1,4], and Laura Dawidowski[1,5,*]

[1]Comisión Nacional de Energía Atómica, Gerencia Química, Buenos Aires, Argentina.
[2]Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina.
[3]Center for Climate and Resilience Research (CR)[2], Santiago, Chile.
[4]Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina.
[5]Instituto de Investigación e Ingeniería Ambiental, Universidad Nacional de San Martín, Buenos Aires, Argentina.
[*]These authors contributed equally to this work.

Correspondence: Melisa Diaz Resquin (mdiazresquin@fi.uba.ar), Laura Dawidowski (dawidows@cnea.gov.ar)

Abstract.

The COVID-19 (COronaVIrus Disease 2019) pandemic provided the unique opportunity to evaluate the role of a sudden and deep decline in air pollutant emissions in the ambient air of numerous cities worldwide. Argentina, in general, and the Metropolitan Area of Buenos Aires (MABA), in particular, were under strict control measures from March to May 2020. Private vehicle restrictions were intense, and primary pollutant concentrations decreased substantially. To quantify the changes in CO, NO, NO_2, PM_{10}, SO_2 and O_3 concentrations under the stay-at-home orders imposed against COVID-19, we compared the observations during the different lockdown phases with both observations during the same period in 2019 and concentrations that would have occurred under a business-as-usual (BAU) scenario under no restrictions. We employed a Random Forest (RF) algorithm to estimate the BAU concentration levels. This approach exhibited a high predictive performance based on only a handful of available indicators (meteorological variables, air quality concentrations and emission temporal variations) at a low computational cost. Results during testing showed that the model captured the observed daily variations and the diurnal cycles of these pollutants with a normalized mean bias (*NMB*) of less than 11% and Pearson correlation coefficients of the diurnal variations of between 0.65 and 0.89 for all the pollutants considered. Based on the Random Forest results, we estimated that the lockdown implied concentration decreases of up to 47% (CO), 60% (NO_x) and 36% (PM_{10}) during the strictest mobility restrictions. Higher O_3 concentrations (up to 87%) were also observed, which is consistent with the response in a VOC-limited chemical regime to the decline in NO_x emissions. Relative changes with respect to the 2019 observations were consistent with those estimated with the Random Forest model, but indicated that larger decreases in primary pollutants and lower increases in O_3 would have occurred. This points out to the need of accounting not only for the differences in emissions, but also in meteorological variables to evaluate the lockdown effects on air quality. The findings of this study may be valuable for formulating emission control strategies that do not disregard their implication on secondary pollutants. The data set used in this study and an introductory machine learning code are openly available at https://data.mendeley.com/datasets/h9y4hb8sf8/1 (Diaz Resquin et al., 2021).

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Earth System Science Data
Open Access Discussions

1 Introduction

In recent times, Machine learning has proved to be an efficient approach to air quality prediction, by relying on historical data to estimate the temporal variability of different pollutants for a specific site at a low computational cost. Also, this kind of model has the ability to unravel underlying patterns in data and deal with complex interactions among predictive variables (Stafoggia et al., 2020).

During the last decade, Random Forest (RF) rose as a new method for the prediction of mean values of atmospheric pollutants (Yu et al., 2016; Feng et al., 2019; Jiang and Riley, 2015). This is a supervised machine learning method, consisting of applying multiple tree classifiers created at random using bagging (i. e., selecting samples stochastically to create new datasets, of which every classification tree is created). Many data science programming languages have libraries where RF is already efficiently implemented (e. g., `scikit-learn` in Python or `randomForest` in R). RF is faster and cheaper than other available models, such as regional chemical transport models (CTMs) in terms of computation costs, it needs less input variables and it is a useful method when information on air pollutant concentrations at a particular site is needed. According to Masih (2019), machine learning techniques may even provide better forecasting than CTMs, and out of the different existing algorithms, RF seems to stand out due to its simplicity and the quality of its results. One of the most recent applications of machine learning methods has been aimed at elucidating the interlinkage among the COVID-19 pandemic lockdown measures, human mobility and air quality (Rahman et al., 2021).

The outbreak of the COVID-19 pandemic at the end of 2019, with its devastating consequences in terms of loss of life and economic impact, has caused many governments around the world to impose different degrees of lockdown. For atmospheric scientists, it has also provided a unique opportunity to examine changes in air pollution under decreased emission levels, in an unintentional worldwide experiment. Many studies have, in general, identified significant decreases of most pollutants, except for O_3, under the stay-at-home orders imposed against COVID-19 (Muhammad et al., 2020; Faridi et al., 2021; Srivastava, 2021; Grange et al., 2021).

An early approach to analyze the changes in air quality due to the implementation of specific control measures was to comparatively assess concentrations during the lockdown with concentrations of the same period of the previous year or the mean value of a period of five years, using exclusively ground-based or satellite observations. However, the degree to which the COVID-19 lockdown influenced air quality is not only a function of emissions but also of both meteorology and physical and chemical atmospheric transformations (Kroll et al., 2020). In consequence, pure statistical tests or observational comparisons might be inadequate to have a complete understanding of what influences pollutant concentrations, since weather conditions, particle persistence, transport, radiation and seasonality affect concentrations by linear and non-linear processes (Šimić et al., 2020). Another option for such comparison is to use models to simulate a scenario in which restrictions were not implemented. Machine learning methods, including the RF algorithm, have been capable of untangling the changes in pollutant concentrations caused by the COVID-19 lockdown measures of weather-driven variability. Velders et al. (2021) have shown the efficiency of the RF method for this task. With respect to reactive species, the RF method has been also used to assess O_3 levels. For example

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Zhan et al. (2018) satisfactorily applied the RF method to predict spatio-temporal variability of daily O_3 concentrations across China using information on meteorology, elevation and emission inventories.

The drastic changes in anthropogenic emissions induced by the global pandemic are of major interest to enhance our understanding of the chemistry related to air quality, particularly when the behavior of secondary pollutants, like ozone (O_3) or components of particulate matter (PM), is explored (Gaubert et al., 2021). O_3, in particular, has a complex behavior depending on multiple factors. NO_x and VOCs play vital roles in the O_3 formation process, and its production can be either VOCs-limited or NO_x-limited (Shi and Brasseur, 2020; Liu et al., 2021; Li et al., 2019). In a VOCs-limited regime, NO_x emission reduction can promote photochemical ozone formation due to the non-linear relationships between O_3 and its precursors. On the other hand, in a NO_x-limited regime, reductions in NO_x concentrations lead to decreasing O_3 levels. Also, sometimes any change in VOCs or NO_x may alter O_3 concentrations and it could be referred to as a transitional (or mixed) regime. The goal of this study was twofold: (i) to explore the performance of the RF method in predicting the air quality situation at two monitoring sites of the Metropolitan Area of Buenos Aires (MABA), Argentina and (ii) to apply this method to estimate the changes in air pollutant concentrations under the COVID-19 control measures. We implemented the RF algorithm to estimate the concentrations of CO, NO_x, SO_2, O_3, and particles with aerodynamic diameter less or equal than 10 µm (PM_{10}) using meteorological and air quality observations, as well as the local diurnal variation of emissions as explanatory variables. Trained with data acquired in 2019 and 2020 before the start of the pandemic, the RF method can only predict concentrations under a business-as-usual (BAU) scenario. In this research, we comparatively assessed the monitored concentrations during two distinct phases of the COVID-19 lockdown with both the expected concentrations resulting from the BAU RF simulations and the observations registered during the corresponding period in 2019. We also provided the first O_3 and SO_2 observational datasets in Buenos Aires in more than a decade. In addition, we studied the responses of O_3 to the reduction in emissions of its precursors (NO_x and VOCs) because of its relevance regarding emission control and health effects.

The remainder of this paper is structured as follows. Section 2 provides a description of the studied area, the different lockdown phases, the air quality and meteorological data and the structure of the random forest model used to estimate the relative changes (RC) during the lockdown. The evaluation of the model and the analysis of the impact due to the emission reductions are in Section 3. Section 4 provides a description of the data and code availability. Finally, Section 5 presents the main conclusions of this work.

2 Material and methods

2.1 Description of the studied area

The MABA comprises the Autonomous City of Buenos Aires (ACBA) and 40 surrounding Districts of the Greater Buenos Aires (GBA). Located along the western coast of the Río de la Plata estuary, on a flat plain, the MABA is the third biggest Megalopolis of Latin America and the Caribbean. It has a population of approximately 13 millions, with a heterogeneous population density in the range 14–20 thousands $\mathrm{inhab\,km^{-2}}$. Its active fleet reached 5.4 million vehicles by 2019 (Anapolsky, 2020).

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

In terms of air pollutant emissions, although road transportation is clearly the largest contributor of CO, VOCs and PM in the area (Castesana et al., 2021), the MABA is also affected by the emissions from residential, commercial and institutional buildings, mainly based on natural gas consumption, and from three power plants, located near the shoreline of the La Plata River, which burn natural gas, gas oil and fuel oil.

2.2 Description of the lockdown for the MABA

Argentina's national government established different lockdown phases for the duration of the pandemic (Decree 297/2020, 2020). Since 80% of Argentina's COVID-19 cases were concentrated in the MABA, some policies applied to the MABA region differed from those applied to the rest of the country. Starting on 20 March 2020, strict measures were imposed to avoid a sharp increase in COVID-19 cases, emphasizing that the population should stay at home and avoid any social contact. All non-essential stores, including toys, furniture and clothing, were closed until 11 May. Table 1 provides a summary of the restrictions set for the MABA during each phase. Under severely restricted mobility, public transport and passenger car circulation decreased drastically. Local mobility dropped down 80% during the Intense lockdown phase and 65% for the Flexible lockdown phase until the end of May (Aktay et al., 2020). It is worth noting that, before the COVID-19 pandemic, 1 million vehicles entered the city of Buenos Aires from the suburbs per day.

Considering the different degrees of the restrictions imposed, we evaluated the impact of the lockdown on air quality according to two distinct periods. The first period, from 20 March to 12 April 2020, corresponded to the most restrictive lockdown (LD). The second period, from 13 April to 25 May, was denominated partial lockdown (PLD) because some restrictions were lifted. The period 1–15 March 2020, before the start of the first lockdown, was defined as BLD and was used to evaluate the model. As from 16 March, flexible restrictions started, but were optional, therefore the period 16–19 March was not considered in our research.

Being combustion the main air pollution source in the area, the significant decrease in traffic flow imposed by the lockdown led necessarily to a decrease in the emissions of traffic-related pollutants (D'Angiola et al., 2010; Puliafito et al., 2017; Diaz Resquin et al., 2018; Castesana et al., 2021).

2.3 Meteorological data and description

The atmospheric general circulation in the MABA is controlled by the influence of the semi-permanent South Atlantic High pressure system. This system influences the climate of the MABA throughout the year by bringing in moist winds from the northeast, which produce most of the precipitation in the area in the form of frontal systems, or storms produced by cyclogenesis, in autumn and winter (Barros et al., 2006). In terms of the climate conditions of the MABA, temperatures at the beginning of autumn range from warm to hot in the afternoon, but they are mild in the nights and the mornings. Later on in the season, conditions are cooler, featuring mild afternoons, and cold nights and mornings.

To identify similarities and differences between the meteorological conditions during the lockdown phases and the testing period (BLD, LD and PLD) with those of the autumn of 2019 (March, April and May, MAM2019) we carried out a meteorological analysis for all periods. We used hourly and daily data from the Buenos Aires Central Observatory (OBS: Lat: 34º 35' S

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

Table 1. Description of the lockdown phases on the MABA. NU Not used (Not included in the model)

Initial Date	Phase	Denomination	Description	Mobility
1 March 2020	Before the lockdown	BLD	Pandemic had started in South America but no restrictions were thus far implemented in Argentina.	100%
16 March 2020	School Closedown and Optional Lockdown	NU	Countrywide, all schools and universities were closed. People were advised to stay at home. Theaters and cinemas were shut. Public events over 200 people were cancelled. Implementing home-office was recommended. Gatherings were to be avoided.	90%
20 March 2020	Strict Lockdown	LD	Bars, restaurants, shopping centers and stores in general were closed, with the exception of food and medicine stores. Only essential economic activities were permitted. Circulation of passenger vehicles was only allowed with a special permit. Public transport was limited within the region. Most industrial activities were suspended. Only groceries were allowed to be delivered. These restrictions apply countrywide, regardless of the amount of cases informed. The country and the district borders were closed.	20%
31 March 2020	Flexible Lockdown I	LD	Food delivery was permited.	20%
13 April 2020	District Differentiated Lockdown I	PLD	More economic activities were permitted. More stores were permitted to open in several districts. The lockdown in the MABA continued, but people started to be less careful about the social distancing measures.	35%
6 May 2020	District Differentiated Lockdown II	PLD	All non-essential stores, including toys, furniture and clothing, were permitted to open with specific protocols. Children were allowed to go for a walk with an accompanying adult on the weekends, but no farther than 500m from home.	35%

Lon: 58° 29’ W). The site of the Meteorological Weather Service of Argentina is located in a residential area. It is representative of the meteorology of the air quality conditions under study.

Average temperatures in the BLD (24.4 °C) and in the LD (21.1 °C) were higher than that in MAM2019 (18 °C) while the average temperature in the PLD (16.8 °C) was lower than that in MAM2019 but close to the corresponding value in May 2019 (16 °C). Precipitation in March and April 2020 exceeded the accumulated values of the same months of 2019 (+60% and + 90% respectively). On the contrary, precipitation in May 2020 exhibited significantly lower values than those of 2019 (-75%).

During MAM2019, the average calm value was 6.7%, while during the BLD, the LD and the PLD, the corresponding calm values were 3.6%, 4.7% and 8.6%. Average wind velocity, within the range 7.5–8.6 $\mathrm{km\,h^{-}1}$, was similar in all periods. In

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

autumn 2020, the prevailing wind was from the NW-N sector with an average contribution of 34% against 26.5% in 2019. The LD and the PLD periods had a similar direction of prevailing winds as autumn 2019, contrarily 45% of winds during the BLD were from the NE–E sector.

Our analysis showed that there were meteorological differences in terms of temperature and precipitation between autumn 2019 and the periods analyzed in 2020 (BLD, LD, and PLD). This is indicative of the need of taking into account the influence of meteorological conditions for comparative purposes of air quality conditions that occurred in different periods.

2.4 Air quality data

We employed air quality data from two monitoring sites: Comisión Nacional de Energía Atómica (CNEA), operated by our research group, and Parque Centenario (PC), managed by the Autonomous City of Buenos Aires (described below). Both sites are mostly influenced by the emissions from mobile and residential sources, and, to a lesser extend, by the thermal power plants, located at least at 6 km from them (Diaz Resquin et al., 2018; Pineda Rojas et al., 2020).

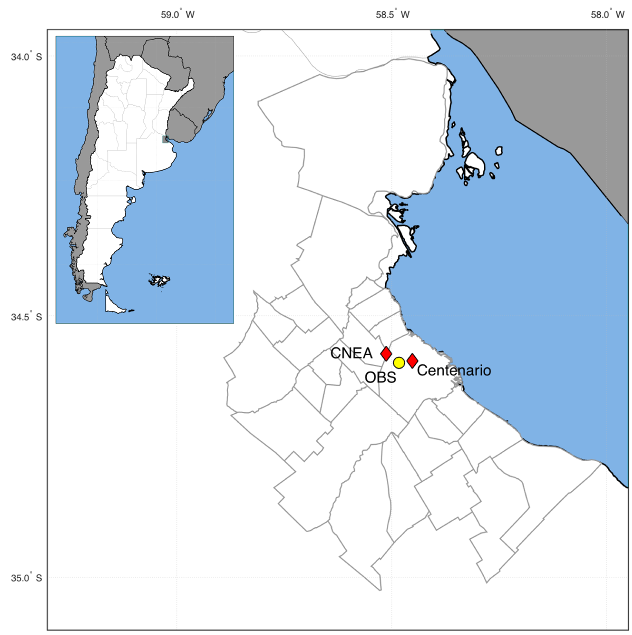

Figure 1. Location of MABA in Argentina (top left); zoom of the MABA(right). In yellow, the location of OBS, the site of the National Meteorological Service monitoring site referred in this study, and in red, the air quality monitoring sites (Shape files from IGN).

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

2.4.1 Comisión Nacional de Energía Atómica

From 23 February 2019 to 26 May 2020 a monitoring campaign was carried out in an open area (-34.57 ºS, -58.51 ºW) situated 14 km away from the Buenos Aires City center (Figure 1) to assess the levels of different gases (CO, NO, $\mathrm{NO_2}$, $\mathrm{SO_2}$, and $\mathrm{O_3}$) and their temporal variability in a residential area of the MABA.

The main goal of this monitoring campaign was to assess the temporal variability of $\mathrm{SO_2}$ and $\mathrm{O_3}$ in the area for an entire year. Although it may seem surprising, especially for a megacity like the MABA, there is scarce and fragmentary information on the concentrations of $\mathrm{SO_2}$ and $\mathrm{O_3}$ for this large urban conglomerate. Presently, $\mathrm{O_3}$ is routinely monitored in only one site of the MABA, located in an industrial area. Past data for the region are only available from a few short-time campaigns carried out in the early 2000s (Reich et al., 2006). Similarly, there is a lack of monitored $\mathrm{SO_2}$ concentrations because historical measurements carried out in the 1990s reported very low values, and therefore decision makers decided not to measure this pollutant on a regular basis. However, it has now become a pollutant of concern for local authorities that have recently decided to start monitoring $\mathrm{SO_2}$ in two of the four air quality stations of the ACBA.

Air pollutant concentrations were continuously acquired as described in Table 2. Monitors were placed at an approximate height of 10 m, and 100 m E from a main traffic artery with a high density of buses, light duty trucks and passenger cars. Another main artery is located 500 m N, having circulation of vehicles including trucks and buses in a low speed stop-and-go pattern. The international airport Jorge Newbery, two thermal power plants and the La Plata river are located within a 19 km radius of the monitoring station.

Data was registered per one minute averages. Unfortunately, from 26 May onwards, restrictions on entering our institute where the monitoring station was located led to the need to suspend the monitoring campaign.

2.4.2 Parque Centenario

To include aerosol variations in this analysis and complement the information of CNEA's site, we used $\mathrm{PM_{10}}$, CO, NO, and $\mathrm{NO_2}$ data from PC station (34.61 ºS, 58.44 ºW), one of the surface air quality sites of the Environmental protection agency of Buenos Aires city (APRA). This site is located in a residential-commercial area with medium vehicular flow and relatively low incidence of stationary sources. A monthly technical report of the hourly-average concentrations registered in PC is available at APRA website (APRA, 2020). Although the city has three other monitoring stations, at least one of the essential periods needed for this study was missing in each of them. Therefore, they did not serve our purpose.

2.4.3 Summary of the datasets

Relatively low concentration values for all the analyzed periods, with no exceedances for short term air quality standard for all the pollutants measured (Decree 1074/18, 2018; Act 1356, 2004) were registered in both sites. Air pollutants, except $\mathrm{SO_2}$, exhibited well defined diurnal cycles.

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Table 2. Description of the equipment used in CNEA site

Pollutant	Instrument	Description	Calibration*
CO	Horiba APMA-370	Sampler with a non-dispersive infrared absorption photometry sensor with a solenoid valve with cross flow modulation.	12.44 ppmV
O_3	Horiba APOA-370	Detector that operates with a cross flow modulation, ultra-violet absorption method in conjunction with the compara-tive calculation method.	0.1 ppmV. ± 0.5% diluted
SO_2	Horiba APSA-370	UV Fluorescence detector.	0.05 ppmV ± 0.6% diluted
NO/NO_2	Horiba APNA-370	Cross flow modulation type with reduced chemilumines-cence detector.	0.099 ppmV ±1.4% di-luted (NO)

* The calibration of the ambient air gases detectors was performed by following U.S. EPA regulations and Horiba standard procedures (see U.S. EPA CFR 40 Part 50, appendixes A1, C, D and F, and the corresponding user manual for the Horiba AP devices). The APMA-370, APSA-370 and APNA-370 were calibrated using EPA certified calibration gases and diluted with an Environics 6103, a NIST traceable mass flow controller dilutor, when needed.

CO and NO_x patterns were governed by traffic emissions (Figs. S1 and S2 of Supplementary Material), with the maximum values in winter. Annual mean average values of NO_x were ~ 37 ppb for both CNEA and PC. Relevant differences in CO were identified, with annual mean levels in PC doubling those measured in CNEA (0.51 ppm versus 0.26 ppm).

PM_{10}, which was only measured in PC, had a mean value of 21 $\mu g\, m^{-3}$, with the maximum values at noon.

With respect to the pollutants that were only measured in CNEA, SO_2 maximum concentrations were registered during autumn (April) with monthly averages in the range 2–2.9 ppb . In terms of O_3 concentrations, maximum daylight levels were registered during summer. The diurnal cycle presented higher levels during the afternoon and was opposite to that of NO_x.

2.5 Air pollution estimations

We used the machine learning RF method to estimate the hypothetical pollutant concentrations that would have occurred in the MABA during the LD and the PLD phases if no lockdown measures had been imposed (BAU scenario). We selected the RF algorithm based on its demonstrated ability to separate the effects of meteorology and chemical reactions from the COVID-19 driven decrease in emissions (Zhan et al., 2018; Rahman et al., 2021; Velders et al., 2021). RF requires a short training time

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

and can provide reliable information on air quality, with a strong anti-overfitting ability (Liu et al., 2021). Relations between different pollutants can also be easily included, which is of particular interest for those that have a very complex chemistry, such as O_3. It is also easy to adapt the methodology to different time periods and sites. The schematic of the model building process can be seen in Fig. 2.

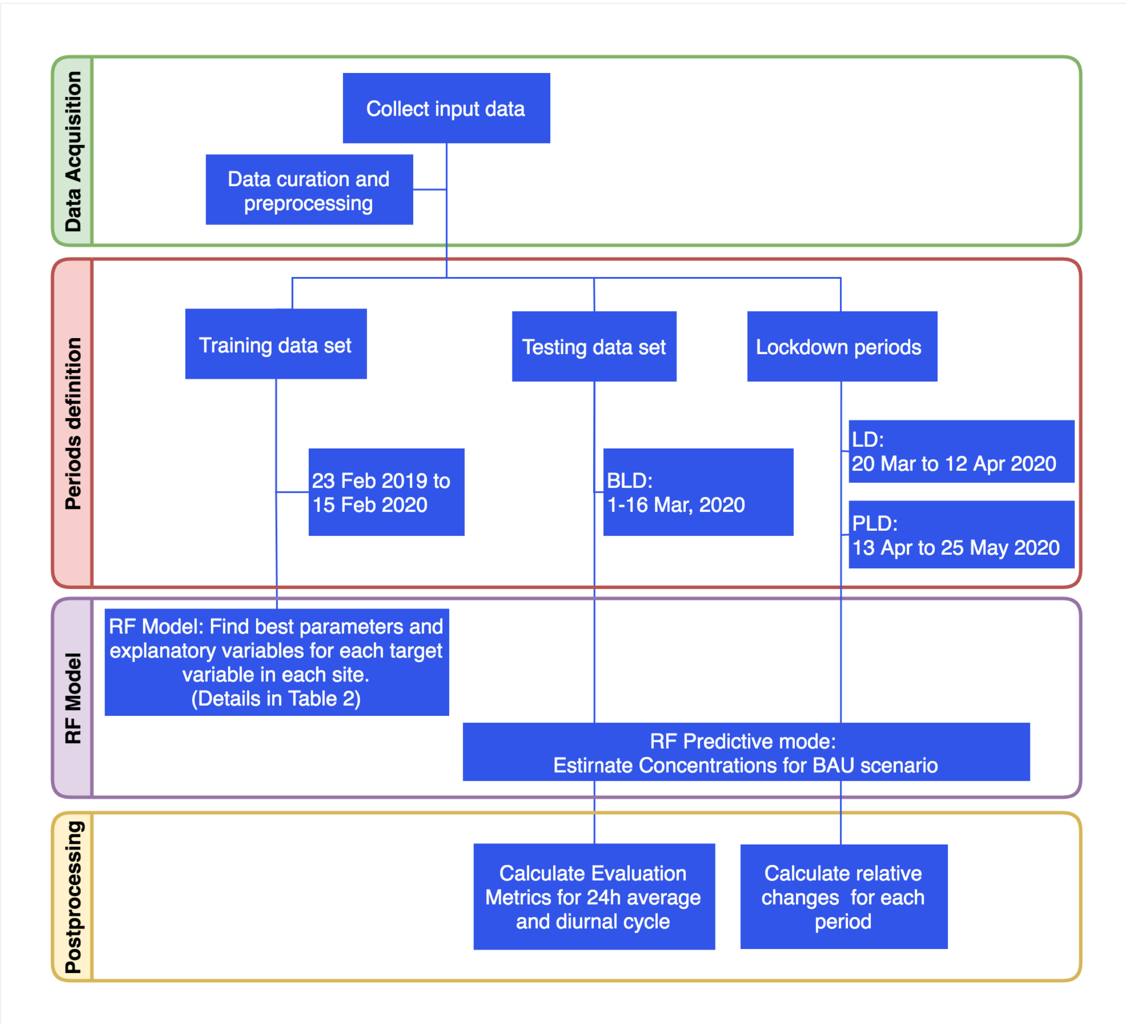

Figure 2. Schematic description of model building and evaluation.

Observations from February 2019 to May 2020 were divided in two different groups: before and after the start of the lockdown. The first group of observations, from February 2019 to February 2020, was used to train and test the RF model (80%–20% split ratio respectively). From 1 March to 25 May 2020, the model was used in predictive mode to estimate pollutant concentrations under the BAU scenario. The BLD period was established as a evaluation period in order to check the adequate

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Table 3. Random Forest model. Target variables and predictors.

Site	Target Variable	Explanatory Variables	ntree	mtry
CNEA	CO	t2, rh2, U, V, gasoline diurnal cycle	1000	3
	NO	t2, rh2, U, V, gasoline and diesel diurnal cycles	1500	3
	NO_2	t2, U, V, gasoline diurnal cycle, CO, NO	1000	3
	SO_2	wspd, daynight, CO, NO, NO_2	1000	3
	O_3	U, V, daynight, CO, NO, NO_2, SO_2	2000	3
PC	CO	t2, wspd, wdir, gasoline diurnal cycle	1000	2
	NO	t2, rh2, slp, U, V, calm, gasoline diurnal cycle	1000	3
	NO_2	t2, U, V, gasoline diurnal cycle, NO	1000	3
	PM_{10}	wspd, diesel diurnal cycle, NO, NO_2	1000	3

rh2: 2m relative humidity; slp: sea level pressure; t2: 2m temperature; U: 10m U component of winds; V: 10m V component of winds; wdir: 10m wind direction; wspd: 10m wind speed; ntree: Number of trees to grow and mtry: Number of variables randomly sampled as candidates at each split.

performance of the model. From 20 March to 25 May, RF estimates were compared to observations in order to quantify and interpret the changes during the LD and the PLD.

Measured air pollutant concentrations were the target variables for each monitoring site, namely NO, CO, NO_2, O_3, and SO_2 (CNEA) and NO, CO, NO_2, and PM_{10} (PC). The available predictive variables were wind speed, wind direction, surface temperature, sea level pressure, relative humidity, the pollutant concentrations available in each of the sites, and a variable called daynight that can have values as morning, afternoon or night. Wind speed and wind direction data were used to derive the U and V wind components to avoid model continuity errors at 0 and 360 degrees. The calm condition and the sine of wind direction were also tested as predictors. Diurnal emission cycles for pollutants associated with gasoline and diesel emissions were added to improve the agreement in the testing period with the variations throughout the day, as published by Castesana et al. (2021). Both cycles were tested for each pollutant. Weekly patterns used were taken from PREP-CHEM (Freitas et al., 2011). All meteorological and air quality variables, as well as emission cycles, were tested as explanatory variables for each pollutant, and those performing the best during testing were selected. In general, a combination of meteorological variables was identified as explanatory variables for CO, NO, and NO_2, with the subset {temperature, relative humidity, U, and V} exhibiting the most relevant role. Wind speed was only relevant for SO_2 (CNEA), CO, and PM_{10} (PC), while wind direction resulted as an explanatory variable for CO in PC only. The only meteorological variables relevant for O_3 were U,V. Noticeably, the diurnal cycle of gasoline vehicles was needed for CO, NO, and NO_2 for both sites, while the diurnal cycle of diesel was required only for NO (CNEA) and PM_{10} (PC). Not surprisingly, O_3 prediction was most dependent on the variable daynight and the simulated concentrations of the other pollutants, including SO_2, but excluding PM_{10}. Table 3 presents the final set of predictive variables used in the RF model that best reproduced the observations during the BLD for each target variable.

2.6 Random Forest model evaluation

The RF model was tested during the BLD for adequate performance, focusing on the reproduction of: (i) the mean value, (ii) the mean diurnal cycles and (iii) the 24 h average concentration.

For each pollutant, differences between the mean value resulting from the RF ($\bar{M}$) and that from the observations ($\bar{O}$) were assessed using the ratio between these two values ($\bar{M}/\bar{O}$). Diurnal cycles were comparatively assessed by graphical inspection of the temporal series of the mean values and spreads of the modelled and observed concentrations of each pollutant, as well as the Pearson correlation coefficient (r). Daily average concentrations were assessed considering the normalized mean bias (*NMB*), the Pearson correlation coefficient and the mean fractional bias (*MFB*).

$$NMB = \frac{1}{N}\sum_{k=1}^{N}\frac{M_k - O_k}{O_k} \times 100 \quad (1)$$

$$r = \frac{1}{N-1}\sum_{k=1}^{N}\left(\frac{M_k - \bar{M}}{\sigma_M}\right)\left(\frac{O_k - \bar{O}}{\sigma_O}\right) \quad (2)$$

$$MFB = \frac{2}{N}\sum_{k=1}^{N}\left(\frac{O_k - M_k}{O_k + M_k}\right) \times 100 \quad (3)$$

The *NMB* is useful for comparing pollutants that cover different concentration scales and it is defined as the difference between modeled and observed mean concentrations normalized by dividing by the mean observed concentration for that period. The r coefficient is useful to measure the linear relationship between two variables. The *MFB* is a measure of mean relative bias and indicates systematic errors (Borrego et al., 2008).

Finally, bivariate polar plots were built considering observations and RF results, using the openair library of the R programming language (Carslaw and Ropkins, 2012; R Core Team, 2019).

3 Results and discussion

3.1 Evaluation of the results of the Random Forest model

In general, modelled CO, NO, NO_2, NO_x and PM_{10} concentrations in both sites were in good agreement with the corresponding observations (see Table 4 and Figs. 3 and 4). For the CNEA site, *NMB* showed a bias $< 10\%$ for all the pollutants and *MFB* of the daily concentrations was between -5.8% (NO) and 12% (CO). For PC, *NMB* was between -1.2% (PM_{10}) and 8.6% (NO). Our results showed that the model tended to slightly overpredict the concentrations of all the pollutants except for NO in CNEA and NO_2 and PM_{10} in PC. *MFB* of the daily concentrations was between 1.1% (PM_{10}) and 9.2% (NO). Calculations of diurnal cycles utilizing RF outcomes reproduced adequately the bimodal behavior during BLD. The agreement was best for NO_2, where the Pearson correlation coefficient (r_{dc}) was 0.89 for both sites while all other r_{dc} were above 0.74. As shown in Figs. 5 and 6, the RF model adequately reproduced the BLD daily concentrations for all the emitted pollutants.

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Table 4. Summary of the evaluation statistics used in Random Forest for the testing period (BLD). *NMB* and *MFB* are presented in percentage for the daily mean concentrations (d) and Pearson Correlation coefficient for diurnal cycle (dc).

		$\left(\frac{\bar{M}}{\bar{O}}\right)_{BLD}$	$NMB_d[\%]$	r_{dc}	$MFB_d[\%]$
PC	CO	1.04	4.0	0.81	6
	NO_x	1.03	3.3	0.77	5.5
	NO	1.08	8.6	0.74	9.2
	NO_2	1.00	-0.4	0.89	3.2
	PM_{10}	0.98	-1.2	0.78	1.1
CNEA	CO	1.10	10	0.76	12
	NO_x	1.01	1.2	0.76	3.3
	NO	0.98	-1.3	0.64	-5.8
	NO_2	1.03	3.4	0.89	6.3
	SO_2	1.04	3.8	0.65	4.3
	O_3	1.08	7.1	0.80	7.3

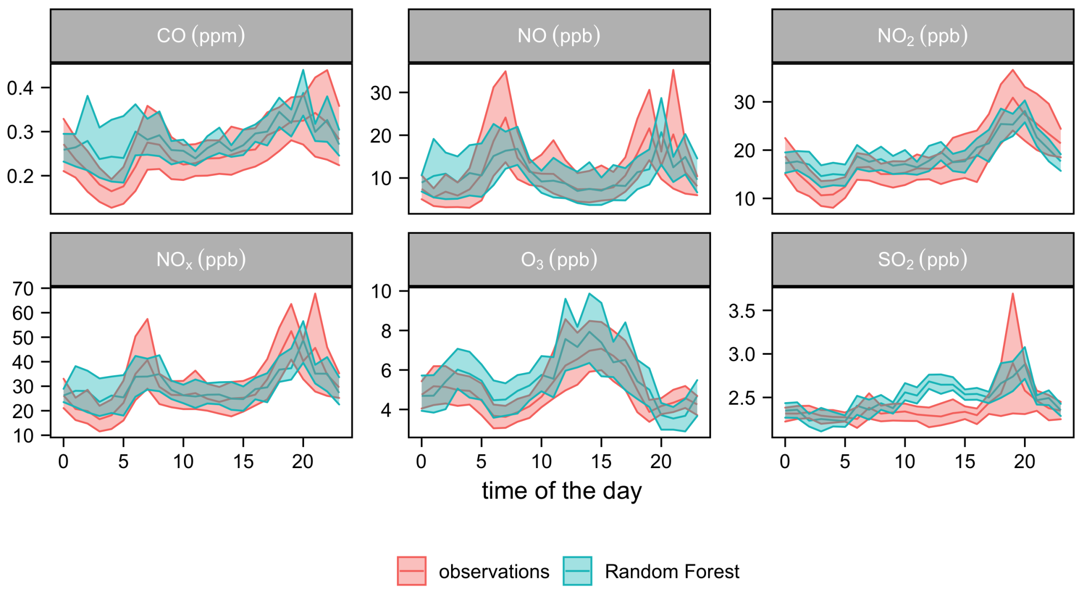

Figure 3. Diurnal cycles for the testing period (from 1 March 2020 to 15 March 2020) for CNEA. The line represents the average diurnal cycle and the shaded area represents the standard deviation.

Results for O_3 were also satisfactory, particularly considering its secondary nature with complex dynamics depending on multiple factors such as radiation energies, VOC and NO_x concentrations and their ratio (Seinfeld and Pandis, 1998). Model

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

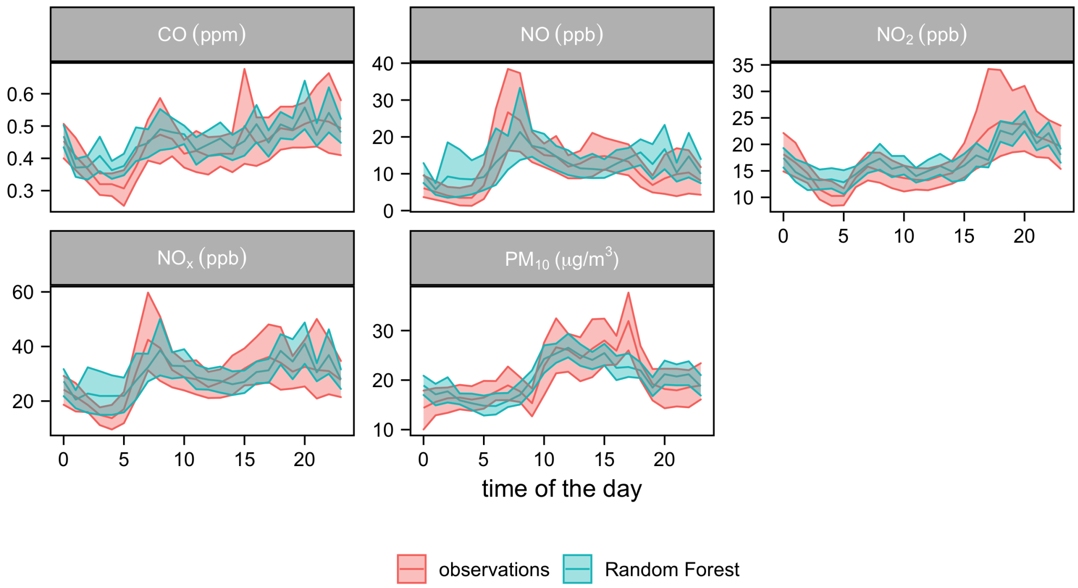

Figure 4. Diurnal cycles for the testing period (from 1 March 2020 to 15 March 2020) for PC. The line represents the average diurnal cycle and the shaded area represents the standard deviation.

performance indicators were $NMB = 7.1\%$ and $r_{dc} = 0.80$. Other processes involved in O_3 chemistry (like the ratios O_3/VOCs and O_3/NO_x) in the MABA were analyzed, as a further way to test the RF model performance. The ratio O_3–CO was used as a proxy for VOCs, because direct VOCs observations were unavailable in the MABA and traffic-borne VOCs are intimately linked to CO (Bon et al., 2011; Cazorla et al., 2020). Overall, 100% of O_3–CO, O_3–NO_2 and O_3–NO_x and 73% of O_3–NO daily ratios from RF were within a factor of 2 of those resulting from the observations (see Figure S4). The Pearson correlation coefficients (r) between observed and estimated O_3–CO and O_3–NO_x daily ratios were found to be 0.79 and 0.89 respectively. In this context, this model was suitable to reproduce not only the levels of primary contaminants in the two analyzed sites, but also the formation of O_3 at the CNEA site. On the other hand, higher discrepancies were found for SO_2 ($r_{dc} = 0.65$) although $NMB < 4\%$. The diurnal cycle of SO_2, which has been identified to be highly linked to diesel trucks (D'Angiola et al., 2010), had a sharp peak between 18:00 and 20:00 that could not be entirely captured by the model. Concentrations from 12:00 to 17:00 were also overestimated. Bivariate polar plots showed a similar pattern for the RF results and the BLD observations for all pollutants (Figures S6 and S7).

3.2 Quantifying and analyzing the changes in concentrations during the lockdown periods

In this discussion we compare the concentrations that were measured during the LD and the PLD phases with the corresponding BAU concentrations estimated by the RF model and/or the observations during MAM2019. The corresponding

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

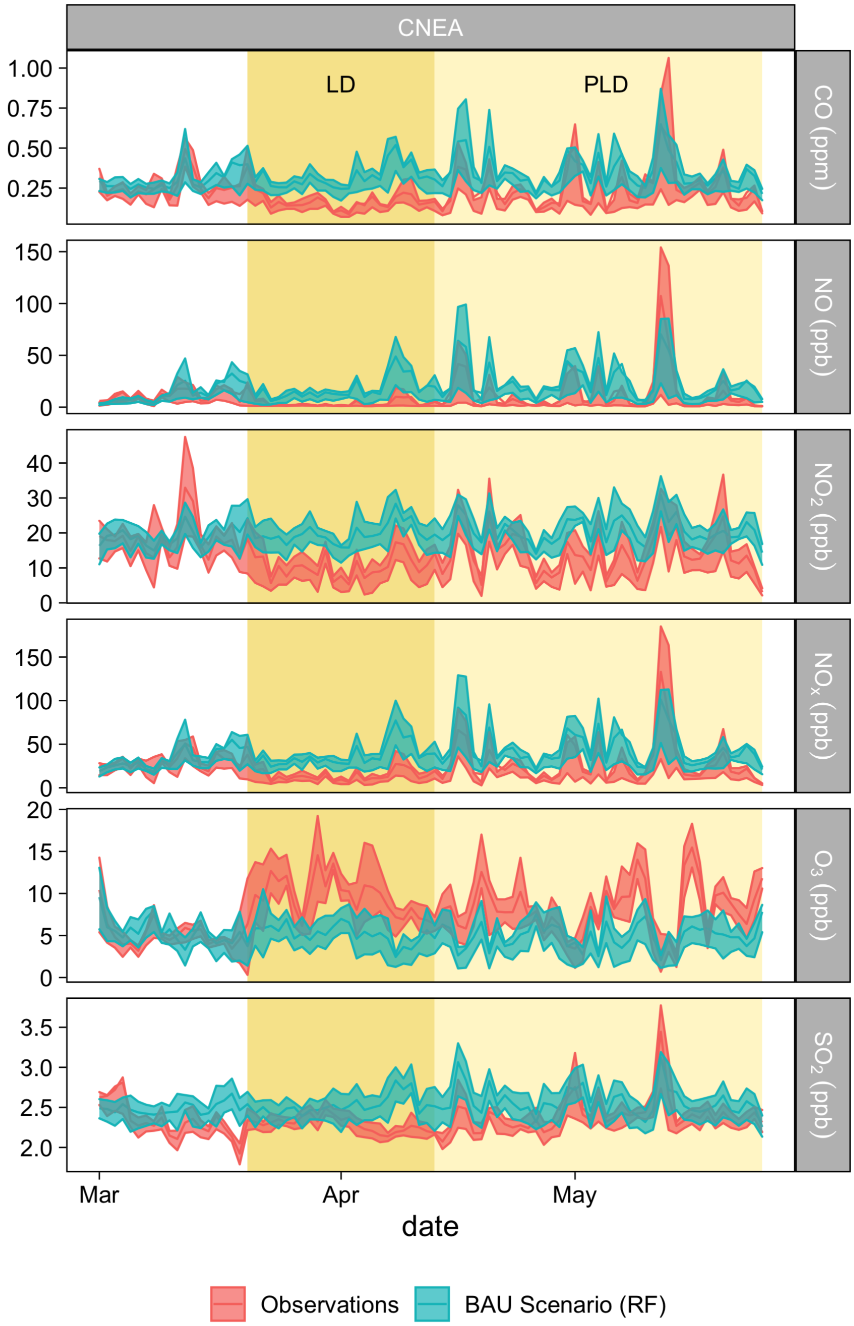

Figure 5. Average daily concentrations for CNEA site. The line represents the 24 h average concentration and the shaded area represents the daily levels between the 25 and 75 percentile.

Open Access Earth System Science Data Discussions

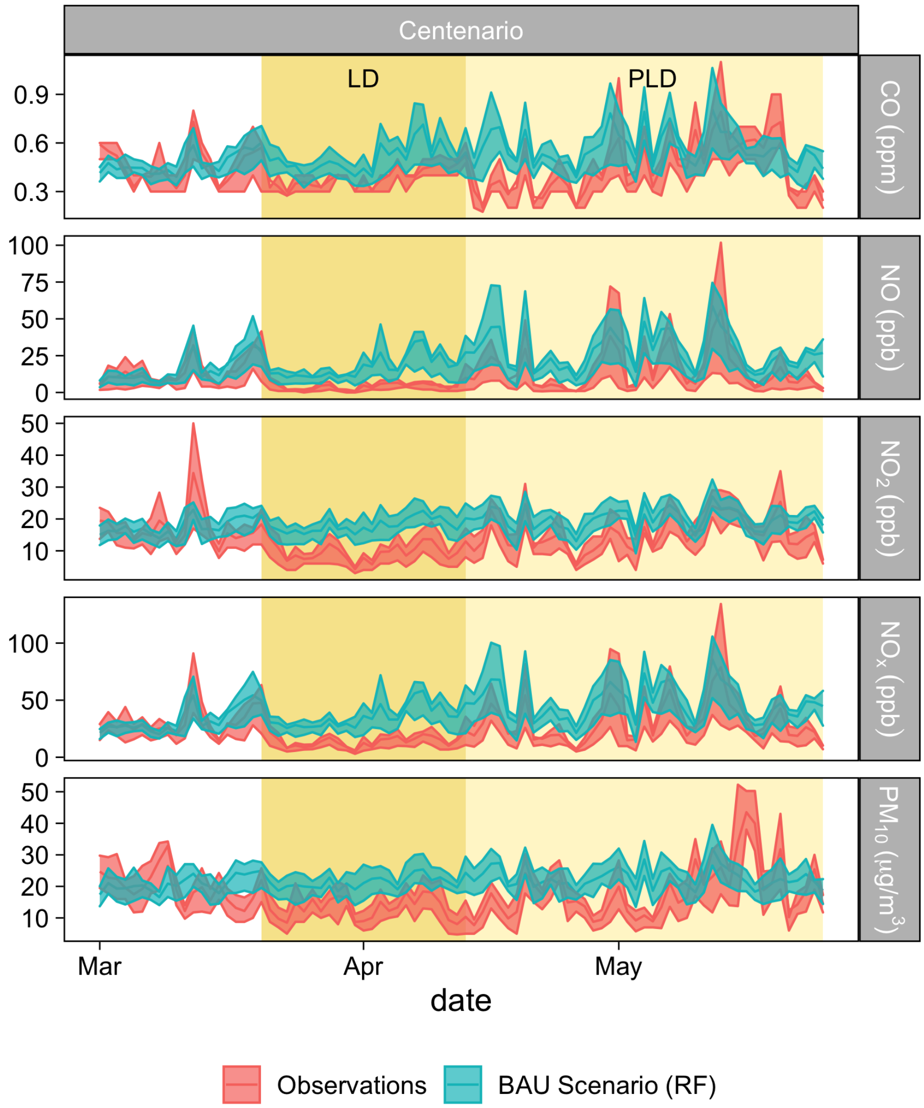

Figure 6. Average daily concentrations for PC site. The line represents the 24 h average concentration and the shaded area represents the daily levels between the 25 and 75 percentile.

percent relative changes (RC_{RF} or $RC_{obs2019}$) were estimated on the basis of Eqs. 4 and 5, respectively.

$$RC_{RF}[\%] = \frac{\text{Period}_{obs2020} - \text{Period}_{RF}}{\text{Period}_{RF}} \times 100 \tag{4}$$

$$RC_{obs2019}[\%] = \frac{\text{Period}_{obs2020} - \text{MAM2019}_{obs}}{\text{MAM2019}_{obs}} \times 100 \tag{5}$$

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

Table 5. Summary of the concentrations and relative changes for the LD and the PLD for PC and CNEA sites compared with Random Forest (RF) and March-April-May 2019 (MAM2019) Observations. In every case, the RC were calculated considering the mean value for each period.

	BLD		MAM2019	LD				PLD			
			Conc.	Conc.		RC [%]		Conc.		RC [%]	
	Bias	NMB_d [%]	obs	obs	RF	RF	MAM2019	obs	RF	RF	MAM2019
PC											
CO (ppm)	0.02	4.0	0.54	0.39	0.50	-22	-28	0.45	0.57	-21	-17
NO_x (ppb)	1.12	3.3	41.7	15.0	35.9	-58	-64	30.7	47.4	-35	-26
NO (ppb)	1.09	8.6	22.9	5.2	17.8	-71	-77	15.2	27.3	-44	-34
NO_2 (ppb)	0.04	-0.4	18.8	9.8	18.1	-46	-48	15.5	20.1	-23	-18
PM_{10} (µgm^{-3})	-0.33	-1.2	22.0	13.6	21.3	-36	-38	18.4	22.8	-19	-16
CNEA											
CO (ppm)	0.03	10.0	0.35	0.17	0.32	-47	-51	0.25	0.35	-28	-28
NO_x (ppb)	0.46	1.2	38.8	14.9	37.6	-60	-62	29.0	45.1	-36	-25
NO (ppb)	-0.15	-1.3	21.4	4.3	17.7	-75	-80	14.6	24.2	-40	-32
NO_2 (ppb)	0.61	3.4	17.4	10.6	19.9	-47	-39	14.4	20.9	-31	-17
SO_2 (ppb)	0.10	3.8	2.8	2.3	2.5	-9	-19	2.4	2.6	-7	-16
O_3 (ppb)	0.38	7.1	6.8	9.6	5.1	87	40	8.0	4.9	65	17

where $\text{Period}_{obs2020}$ corresponds to the air quality observations during the LD or the PLD, the MAM2019_{obs} represents the air quality observations from March to May 2019 and Period_{RF} are the RF estimates for the LD or the PLD.

Being both monitoring sites highly influenced by vehicular emissions, the reduction of $\sim 80\%$ in traffic that was registered during the LD period led to a significant air quality improvement of primary pollutants (Figures 7 and 8). At CNEA, located in the suburbs, RC_{RF} were -60%, -47% and -9% for NO_x, CO and SO_2 respectively (Table 5). On the other hand, observed O_3 levels were 87% and 65% higher in comparison with the estimations for the LD and PLD respectively. At PC, located in a residential-commercial area where activities during the LD period were more intense than in the suburbs, RC_{RF} were -58% (NO_x), -22% (CO) and -36% (PM_{10}). Table 5 also shows an increment of primary pollutants for the PLD period compared to the LD, reflected in smaller relative differences. This is consistent with the increment in traffic flow.

Figures 5 and 6 allow visualizing the differences in daily concentrations between observations and RF estimates for the three considered periods (BLD, LD and PLD). For CNEA, CO and NO_x observations and predictions for the BLD period showed *NMB* <10%. Noticeably, most of the changes were observed right from the day after lockdown. Pollutant levels were almost fully recovered by the last week of the PLD period.

In what follows, the results are presented by species, highlighting the most relevant relative changes in concentrations and their relationship with wind direction and speed, using bivariate polar plots (Figures 9 and 10). Bivariate polar plots can also be helpful to distinguish potential sources that impact the monitoring sites.

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

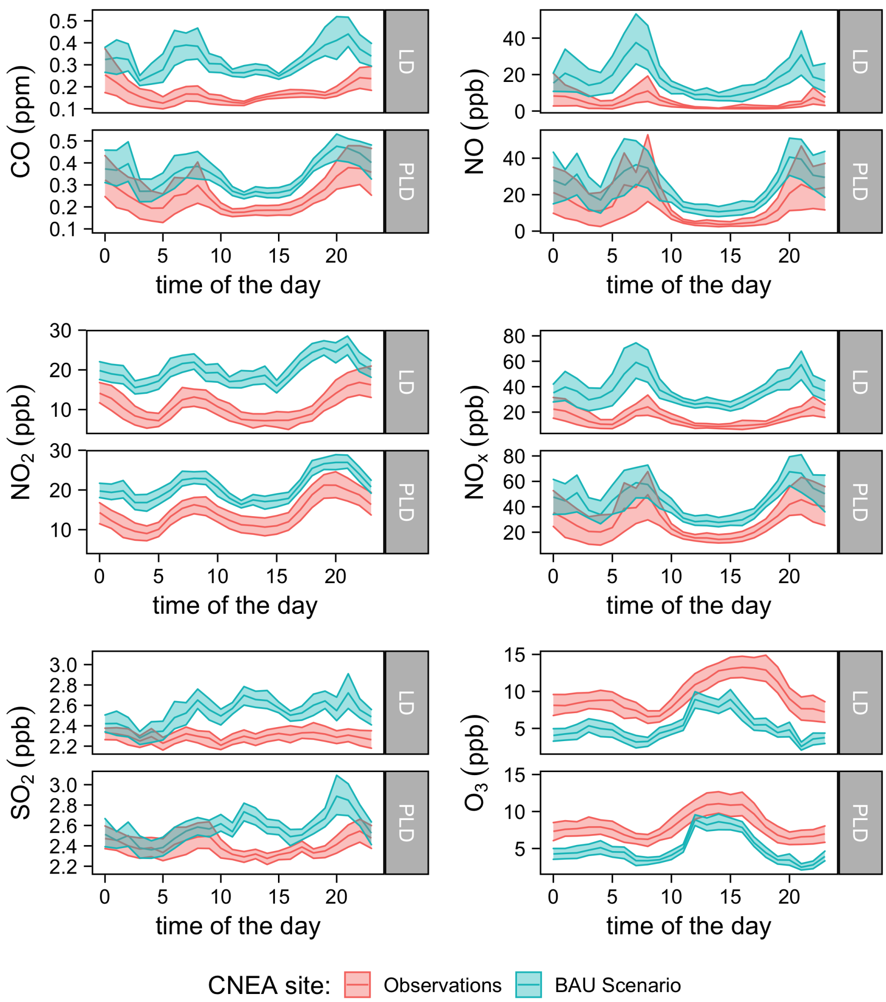

Figure 7. Mean diurnal cycle for the different pollutants for the LD (from 20 March 2020 to 13 April 2020) and the PLD (13 April 2020 to 25 May 2020) for CNEA site.

3.2.1 Carbon monoxide

The combination of results of $RC_{obs2019}$ and RC_{RF} and the *NMB* detected in the model for CO (Table 5), that is 10% for CNEA and 4% for PC, reveals the importance to assess the decrease in concentrations during COVID-19 restrictions using the RF model in order to consider the meteorological differences between 2019 and 2020. For PC, during the LD phase, $RC_{obs2019}$ was -28% while RC_{RF} was -22%. Considering that the model overestimates CO concentrations (*NMB* of 4%), RC_{RF} would be as small as -19%. Therefore, according to the RF model, the CO reductions (the opposite of RC_{RF}) during the LD period may have been of only 19% while the simple comparison between observations indicated that a larger reduction (28%) would have occurred. Nevertheless in the PLD, estimates of relative changes were similar: 21% (RC_{RF}) and 17% ($RC_{obs2019}$). For CNEA,

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

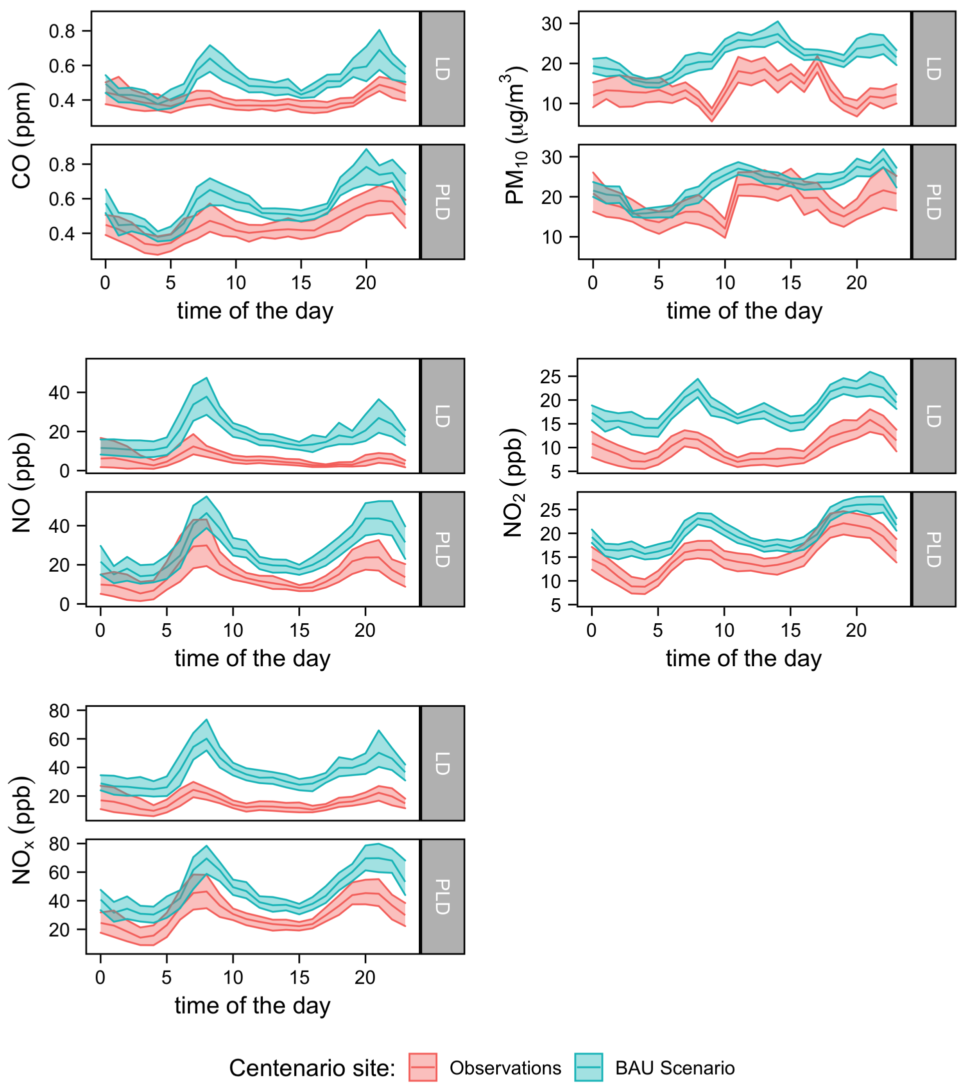

Figure 8. Mean diurnal cycle for the different pollutants for lockdown (20-03-20 to 13-04-20) and partial lockdown (13-04-20 to 25-05-20) for the site PC.

during the LD phase, $RC_{obs2019}$ was -47% while RC_{RF} was -51%. Taking the bias into account, differences might be as small as -42%. Instead, during the PLD, although raw numbers were the same (28%) for both relative changes, the positive bias in RF estimations makes that RC_{RF} might have been as small as 21%.

As expected, in CNEA, the partial recovery of traffic during the PLD was reflected in a smaller reduction in the CO concentrations: 47% (LD) versus 28% (PLD). However, a similar reduction was not observed in PC: 22% (LD) versus 21% (PLD). We do not have a plausible explanation for this relatively sustained CO level during the LD and the PLD.

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Open Access Earth System Science Data Discussions

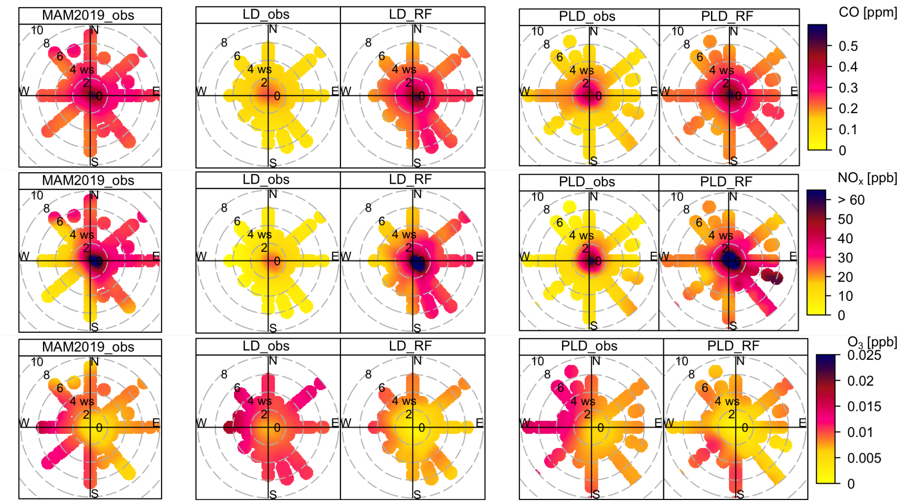

Figure 9. Bivariate polar plot for CNEA of hourly means for observations during MAM2019 and lockdown periods versus the BAU scenario estimated with RF model. The radial axis represents wind speed, the angular axis represents wind direction, and the color scale represents pollutant concentrations.

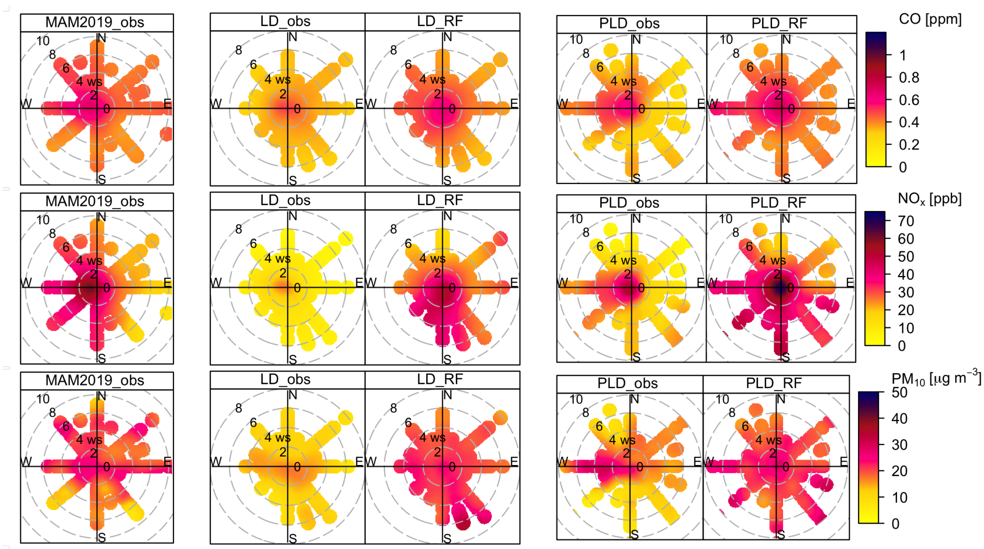

Figure 10. Bivariate polar plot for PC of hourly means for observations during MAM2019 and lockdown periods versus the BAU scenario estimated with the RF model. The radial axis represents wind speed, the angular axis represents wind direction, and the color scale represents pollutant concentrations.

Observed CO had lower concentration values and flatter diurnal patterns than our simulations of a BAU scenario (Figs. 7 and 8). Relative changes appeared to be significantly lower for PC than CNEA (-47% vs. -22%), but the difference in absolute terms is only 0.04 ppm (0.11 and 0.15 ppm for PC and CNEA respectively). This is linked to a general decrease in mobile emissions, reflected in the fact that monitored hourly concentrations in CNEA were between 22% and 62% lower than the RF estimates. The highest differences corresponded to the morning rush hour, when RF estimates peaked while observed concentrations were largely undifferentiated from the general flat pattern. In PC, these relative decreases were in the order 36% (morning rush hour) > 22% (afternoon) > 0.4% (night). This decrease in the CO concentrations is indicative of a decrease in traffic flow of gasoline vehicles, which are known to emit relatively higher levels of CO than diesel vehicles. This reduction far surpasses any bias detected in RF simulations, particularly during rush hours, where RF showed close to no bias (Figs. 3 and 4).

As shown in Figure 9, for the CNEA site during MAM2019, concentrations were similar for all wind directions and speeds (up to 8 m/s). The largest relative changes between the 2020 observations and the RF simulations were when winds were coming from the E and SE (both for the LD and the PLD). These were probably due to a reduction in traffic on the highway (see Section 2.4.1), which according to Diaz Resquin et al. (2018), is one of the principal sources of fuel combustion emissions.

An equivalent analysis for PC (Figure 10) yielded similar results during MAM2019, although concentrations seemed to be largest when winds were from the W. However, relative changes during the LD and the PLD did not seem to have a clear dominant wind direction. During the PLD, sources from the W reappeared.

3.2.2 Nitrogen oxides

The drastic reduction of vehicular emissions impacted positively in the NO_x levels. Both sites presented relative changes of around -60% for the LD and an absolute difference of around -20 ppb for the NO_x concentrations, when comparing RF estimations with MAM2020$_{obs}$. Considering that the NO_x biases are positive, a similar analysis to that discussed for the CO relative changes could also be done. However, the *NMB* of this pollutant is in the range between 1.2% and 3% and, for this reason, both relative changes remain similar. Discrepancies of $\sim$ 10% in the relative changes were found in the PLD period. Concentrations were consistently smaller during the day (see Figs. 7 and 8).

In both sites, the relative change of NO_x was larger than that of CO. Arguably, this indicates that the power plants did not contribute in any major way to the observed differences. This is probably due to a reduced circulation of diesel vehicles, which are the major NO_x emitters (D'Angiola et al., 2010; Ghaffarpasand et al., 2020).

Fig. 9 shows the bivariate polar plots of the NO_x concentrations at the CNEA site. The bivariate polar plot in MAM2019 provides evidence for two main contributing sources. One source was due to air masses from E-SE directions at low wind speeds and the second source was associated with higher wind speeds from N-NW direction. The source to the E-SE could be dominated by ground-level road traffic emissions that are closer to the site because high concentrations under low wind speeds are indicative of surface emissions released with little or no buoyancy (Uria-Tellaetxe and Carslaw, 2014). Also, the wind direction where this source was dominant corresponds to the highway previously described in Section Observational Data - CNEA. The source to the N-NW was associated with high concentrations at high wind speeds, which is indicative of emissions at a greater distance. It is plausible to attribute these NO_x levels to the main access avenue that connects the city with the suburbs and is located in this direction, due to the presence of heavy-duty diesel vehicles and buses and the number of flowing traffic stops. During the LD and the PLD, the highest RC_{RF} were present when winds were coming from the highway. This serves as further evidence that the observed effects were mainly due to changes in traffic, and not to the changes in residential emission patterns due to lifestyle changes during the lockdown.

In the case of PC, as shown in Figure 10, during MAM2019, the main sources seemed to be located to the W and SW of the station. These two directions entailed the largest changes due to restrictions during the LD period. During the PLD period, in a similar manner than CO, the sources to the W were partially restored (although concentrations from the SW remained low).

3.2.3 Ozone

By contrast to the other pollutants considered, the O_3 increased when compared to a no-restrictions scenario. Its relative changes estimated using RF were 87% and 65% during the LD and the PLD periods respectively while the ones estimated through direct comparison with $MAM2019_{obs}$ were 40% and 17%. Moreover, if the positive bias is considered, the differences between RC_{RF} and $RC_{obs2019}$ could be even larger.

Recent studies of the lockdown effects on atmospheric composition have also reported large O_3 increases at urban sites and indicated the need of analyzing changes in precursor emissions and meteorological parameters in light of their role in the nonlinear response in the O_3 concentrations (Ordóñez et al., 2020; Tobías et al., 2020; Nakada Kondo and Urban, 2020; Shi and Brasseur, 2020). Hence, consideration of the joint effects of the changes on precursors and meteorology are of great value to understand the differences between the relative changes estimated using RF concentrations and $MAM2019_{obs}$. Based on Figures 5 and 7, we provide plausible explanations for these discrepancies.

It is well-known that decreasing NO_x levels in a VOC-limited regime tend to increase O_3. It is most likely that the lower concentrations of freshly emitted NO registered during LD and PLD in CNEA provoked a decline in the local scavenging of O_3, leading to higher O_3 concentrations, particularly in the morning (Tobías et al., 2020; Nakada and Urban, 2020). Even though NO is the pollutant that had the highest relative decrease during the LD and the PLD, its reduction is not enough to explain the overall relative increase in O_3, and therefore NO_2 might have played a role as well. Lower NO_2 levels could have also resulted in more OH to initiate O_3 production because the inhibition of termination reaction favors faster O_3 accumulation (Seguel et al., 2012).

With respect to the role of aerosols, it is worth noting that a significant decrease in PM_{10} was registered in PC. This likely implied consequent reduction not only in the mass concentrations of $PM_{2.5}$ and PM_1 but especially in the number concentration of fine and ultrafine particles (Arkouli et al., 2010; Gelman Constantin et al., 2021). A similar situation most likely occurred in CNEA. This could have led to greater photolysis due to the decrease in emissions of fine particles as a consequence of the vehicular restrictions imposed during the lockdowns, which in turn could have led to higher O_3 concentrations (Wang et al., 2019).

Lastly, meteorological factors might be relevant. The effects of meteorology can be rather complex since the O_3 precursor concentrations and reaction rates are affected in multiple ways (Wang et al., 2017). Therefore, they are not easy to analyze individually. In our simulations, although only winds and the variable daynight were included directly, temperature and relative humidity do affect O_3 precursors. Solar radiation, which is highly relevant for O_3 chemistry, is also linked to the variable

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

daynight. In this particular case, during the LD, elevated O_3 concentrations occurred on days with high temperatures and low winds, which favor the photochemical production of O_3 and the accumulation of ozone and its precursors.

As expected, the bivariate polar plots (Figure 9) show that O_3 behaved opposite to NO_x, having the largest increases when winds came from the E and SE during the LD and also when they came from the E and NW during the PLD.

From these results, we can also derive that the area where the CNEA site is located behaves as a region with a VOC-limited chemical regime, because the reduction in NO_x emissions caused an increase in ozone concentrations (Blanchard and Fairley, 2001; Heuss et al., 2003; Yarwood et al., 2003; Blanchard and Tanenbaum, 2006). We identified a similar behavior of increasing O_3 concentrations under decreasing NO_x levels when analyzing the 2019 data for weekends. This is related to the denominated weekend effect in a VOC-limited regime (Koo et al., 2012).

3.2.4 Sulphur dioxide

During the LD, the SO_2 seemed to be slightly lower (about 9%) than the RF simulations and major differences between model and observations during the lockdown periods were during the afternoon and night (Figure 7). However, it should be noted that during the BLD the model overestimated SO_2 during the afternoon by about 10%. These values are quite different to direct comparison with MAM2019$_{obs}$, which exhibited 19% lower concentrations during the same period. During the PLD, the SO_2 concentrations were 7% and 16% lower than RF and MAM2019$_{obs}$ respectively.

A potential reason for observing smaller differences between BAU estimates and observations would be that the main source of this pollutant are the vehicle emissions of heavy-duty diesel trucks, that are mainly associated with essential activities, which were the least affected by lockdown restrictions. Therefore, further research would be needed before drawing conclusions about this pollutant.

3.2.5 Particulate matter 10 μm

During the LD phase, the RC of PM_{10} levels were similar: -36% (RF) and -38% (obs2019). Also, in this case, the negative bias, -1.2%, would make this difference even smaller. During the PLD, RC_{RF} was -19% and $RC_{obs2019}$, -16%. BLD levels were recovered about eight weeks after LD's inception.

When winds are taken into account (Figure 10), we observe a general reduction from all directions during the LD. Two sources account for this: (i) the anthropogenic PM_{10} emissions close to the monitoring site that were mostly from vehicle diesel combustion and soot resuspension and (ii) natural sources, such as dust emissions, from the nearest large open area. In a similar fashion to CO and NO_x, sources from the W were reestablished during the PLD.

3.3 Vehicle emission reduction strategies and air pollution in the MABA

Strategies for controlling pollution from vehicular emissions in the MABA must take into account the relative reductions of NO_x and VOCs to avoid an unintended increment in O_3 concentrations. The atmosphere in the MABA is usually cleaned up during the night, due to a flat topography and the city's wind dynamics. Therefore, criteria pollutants rarely surpass air quality

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

norms. Even though no specific policies to reduce them have been implemented, greenhouse emission policies that are in place and affect traffic may have a major impact. These include (i) technological advances in diesel buses, that should reduce NO_x and PM_{10}, without a major impact in VOCs and (ii) an increase of the fraction of electric cars, which should reduce NO_x and VOC concentrations. Thus, if NO_x emissions decrease like they did during the COVID lockdown, this will likely result in an important increment in tropospheric O_3 in the MABA if no additional measures regarding VOCs emissions are included. In fact, under the VOC-limited regime identified for the MABA, control of VOCs emission would be more efficient to reduce local peaks in O_3.

This highlights the importance of having comprehensive air quality policies rather than focusing on reductions in individual pollutants.

4 Code and data availability

Hourly concentrations of CO, NO, NO_2, SO_2 and O_3 in CNEA, are available in .csv format at https://data.mendeley.com/datasets/h9y4hb8sf8/1 (Diaz Resquin et al., 2021). We also provide an introductory R notebook with some baseline simulations. For PC regulatory averages are publicly available and can be accessed through their website (https://data.buenosaires.gob.ar/dataset/calidad-aire). Nevertheless hourly data is not regularly reported, but can be requested to the Environmental Protection Agency of Buenos Aires City. To enable a machine learning quick start to reproduce the baseline experiments, we also added to the dataset the meteorological data used to run the simulations. It is publicly available at the website of the National Weather Service (https://www.smn.gob.ar/descarga-de-datos).

5 Summary and conclusions

The RF model was trained with a set of 1-year air pollutant concentrations determined in two monitoring sites of the metropolitan area of Buenos Aires. The performance of the model used in a predictive mode was tested on the basis of observations registered before the outbreak of the COVID-19 pandemic. Observations in the two first phases of the lockdown measures imposed were compared against observations in 2019 and the business-as-usual RF simulated concentrations. In addition, this study provided information on O_3 and SO_2 concentrations that is still scarce and fragmentary for the large urban conglomerate. The main conclusions are listed below:

(i) The resulting set of explanatory variables for the different pollutants in each site provides evidence of the need for careful identification during the training period. Although ideally the best explanatory variables could be identified by trial and error by non-experienced users of RF models, it is advisable to count with expert judgment for a meaningful and relatively fast selection;

(ii) The RF model was able to reproduce air quality observations at two monitoring stations in the MABA when it was tested for a 15-day period previous to the outbreak of the COVID-19 pandemic. This approach allowed predicting pollutant daily mean values with a mean bias of less than 11% by using data of air quality, emissions and meteorology

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

and analyzing the effect of wind direction and speed in pollutant concentration, which is useful when characterizing pollution sources;

(iii) The atmospheric concentrations of CO, NO_x and PM_{10} decreased and O_3 increased in comparison with the same period of the previous year and the RF estimations. In the case of CO, NO_x and PM_{10}, the difference between the two methodologies was less than 11%, but considering the model bias of each pollutant, these differences could be larger in all cases except for PM_{10} during the LD. However, in the case of O_3, which has a complex chemistry and nonlinear dependence with its precursors and with meteorology, it was larger than 40%. The relative changes in pollutant concentrations attributable to lockdown are closely linked to the reduction in traffic. On one hand, this could be noted in the changes in the observed and simulated diurnal patterns. On the other hand, the effect of the circulation of vehicles could also be observed when analyzing the bivariate polar plots made possible by the use of the modelling technique. This allowed us to locate likely emission sources, mostly from high traffic directions.

(iv) The main advantage of using RF estimations instead of comparing with the same period of the previous year lies in taking into account changes in the meteorology that might influence pollutant concentrations, with only a marginal increase in the workload and computational cost. This feature of the model can be used to evaluate the effects of air pollution control policies and measures while taking into account changes in meteorology. For a successful application, it is essential to use a set of data in which emissions are similar to those that are expected to be simulated. Taking into account that the MABA has thus far only six long-term air quality monitoring stations, we believe that this methodology could be used by local authorities, both for forecasting and evaluating regulatory measures. Relations between different pollutants can also be readily included, which is of particular interest for those that have very complex chemistry, such as O_3. The observational input data needed for future RF simulations can be readily updated. Most available RF models are user friendly, rather straightforward to implement and do not require large computational capacity. The methodology is amenable to be adapted to different time periods and sites and implemented by the technical staff of regulatory agencies. Expert advice may be needed during the selection of the predictive variables and model optimization;

(v) In this work we provide the first year-long in situ observational dataset on tropospheric O_3 and SO_2, outside of an industrial area for the MABA in the last decade. We also provide co-located concentrations of CO, NO and NO_2;

(vi) According to our measurements, the MABA seems to be in a VOC-limited regime. If VOC emissions are not carefully regulated, a NO_x reduction would imply an increase in the tropospheric O_3. Knowing how the concentrations of O_3 in the troposphere respond to reducing the emissions of its precursors is relevant when planning appropriate strategies to reduce CO, NMVOCs and NO_x emissions. Even though this classification is limited due to the fact that we only have single point measurements, this could be a useful starting point for a more thorough characterization of the ozone regime in this urban area.

https://doi.org/10.5194/essd-2021-318
Preprint. Discussion started: 13 October 2021

Author contributions. Díaz Resquín, Melisa: Conceptualization, Methodology, Validation, Data curation, Data Analysis, Formal analysis, Writing - original draft, review & editing. Lichtig, Pablo: Data Analysis, Formal analysis, Writing - original draft, review & editing. Alessandrello, Diego: Data Acquisition, Data curation, Writing - original draft. De Oto, Marcelo: Data Acquisition, Writing - original draft, review & editing. Castesana, Paula: Data Analysis, Writing - review. Rössler, Cristina: Data Analysis, Writing - original draft. Gómez, Darío: Funding acquisition, Supervision, Formal analysis, Writing - original draft, review & editing. Dawidowski, Laura: Funding acquisition, Conceptualization, Supervision, Formal analysis, Writing - original draft, review & editing.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. We want acknowledge the participation of the entire group of Atmospheric Chemistry of the National Atomic Energy Commission of Argentina (CNEA) to hold the campaign even during Lockdown. This study was carried out in part with the aid of grants PICT-O 2016-4802 (Agencia Nacional de Promoción Científica y Tecnológica, Argentina) and PICT 2016-3590 (Fondo para la Investigación Científica y Tecnológica). The research work that leads to the results has also been funded by the EU Horizon 2020 Marie Skłodowska-Curie project PAPILA (GA 777544, MSCA action for research and innovation staff exchange). The authors wish to thank the Environmental Protection Agency of Buenos Aires (APRA) and the National Weather Service of Argentina (SMN) for sharing air quality and meteorological data for this study.

References

Act 1356: Preservación del recurso aire y prevención y control de la contaminación atmosférica., Available at https://www.buenosaires.gob.ar/sites/gcaba/files/documents/ley_1356.pdf (2021/09/07), 2004.

Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., Kamath, C., Kansal, M., Lange, A., Mandayam, C., Oplinger, A., Pluntke, C., Roessler, T., Schlosberg, A., Shekel, T., Vispute, S., Vu, M., Wellenius, G., Williams, B., and Wilson, R. J.: Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.1), 2020.

Anapolsky, S.: ¿ cómo nos movemos en el AMBA? Conclusiones de la evidencia empírica y alternativas post-covid, Universidad de San Martín. ISSN: 2469-1631 Serie: Documentos de Trabajo del IT. Available at https://www.unsam.edu.ar/institutos/transporte/publicaciones/Documento%2018%20Comonos%20movemos%20en%20el%20AMBA%20-%20Anapolsky.pdfl (2021/09/07), 2020.

Arkouli, M., Ulke, A. G., Endlicher, W., Baumbach, G., Schultz, E., Vogt, U., Müller, M., Dawidowski, L., Faggi, A., Wolf-Benning, U., and Scheffknecht, G.: Distribution and temporal behavior of particulate matter over the urban area of Buenos Aires, Atmos. Pollut. Res., 1, 1–8, https://doi.org/https://doi.org/10.5094/APR.2010.001, 2010.

Barros, V., Clarke, R., and Dias, P. S.: Climate change in the La Plata basin, Publication of the Inter-American Institute for Global Change Research (IAI), São José dos Campos, Brazil, 2006.

Blanchard, C. and Tanenbaum, S.: Weekday/Weekend differences in ambient air pollutant concentrations in atlanta and the southeastern United States., J Air Waste Manag. Assoc., 56, 271–284, https://doi.org/doi: 10.1080/10473289.2006.10464455., 2006.

Blanchard, C. L. and Fairley, D.: Spatial mapping of VOC and NO_x-limitation of ozone formation in central California, Atmos. Environ., 35, 3861–3873, https://doi.org/https://doi.org/10.1016/S1352-2310(01)00153-4, 2001.

Bon, D. M., Ulbrich, I. M., de Gouw, J. A., Warneke, C., Kuster, W. C., Alexander, M. L., Baker, A., Beyersdorf, A. J., Blake, D., Fall, R., Jimenez, J. L., Herndon, S. C., Huey, L. G., Knighton, W. B., Ortega, J., Springston, S., and Vargas, O.: Measurements of volatile organic compounds at a suburban ground site (T1) in Mexico City during the MILAGRO 2006 campaign: measurement comparison, emission ratios, and source attribution, Atmos. Chem. Phys., 11, 2399–2421, https://doi.org/10.5194/acp-11-2399-2011, 2011.

Borrego, C., Monteiro, A., Ferreira, J., Miranda, A. I., Costa, A. M., Carvalho, A. C., and Lopes, M.: Procedures for estimation of modelling uncertainty in air quality assessment, Environ. Int., 34, 613–620, https://doi.org/https://doi.org/10.1016/j.envint.2007.12.005, 2008.

Carslaw, D. C. and Ropkins, K.: openair — An R package for air quality data analysis, Environ. Model. Softw., 27-28, 52–61, https://doi.org/https://doi.org/10.1016/j.envsoft.2011.09.008, 2012.

Castesana, P., Diaz Resquin, M., Huneeus, N., Puliafito, E., Darras, S., Gómez, D., Granier, C., Osses Alvarado, M., Rojas, N., and Dawidowski, L.: PAPILA dataset: a regional emission inventory of reactive gases for South America based on the combination of local and global information, Earth Syst. Sci. Data Discuss., 2021, 1–34, https://doi.org/10.5194/essd-2021-208, 2021.

Cazorla, M., Herrera, E., Palomeque, E., and Saud, N.: What the COVID-19 lockdown revealed about photochemistry and ozone production in Quito, Ecuador, Atmos. Pollut. Res., https://doi.org/10.1016/j.apr.2020.08.028, 2020.

D'Angiola, A., Dawidowski, L. E., Gómez, D. R., and Osses, M.: On-road traffic emissions in a megacity, Atmos. Environ., 44, 483–493, https://doi.org/10.1016/j.atmosenv.2009.11.004, 2010.

Decree 1074/18: Available at https://www.opds.gba.gov.ar/sites/default/files/Decreto%201074%2018.pdf (2021/09/07), 2018.

Decree 297/2020: Available at http://servicios.infoleg.gob.ar/infolegInternet/anexos/335000-339999/335741/norma.htm (2021/09/07), 2020.

Diaz Resquin, M., Santágata, D., Gallardo, L., Gómez, D., Rössler, C., and Dawidowski, L.: Local and remote black carbon sources in the Metropolitan Area of Buenos Aires, Atmos. Environ., 182, 105–114, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.03.018, 2018.

Diaz Resquin, M. C., Alessandrello, D., De Oto, M., Lichtig, P., Bajano, H., Ponso, A., Bajano, F., Dawidowski, L., and Gómez, D.: AQ-CNEA-CAC Air quality dataset (2019-2020): "A machine learning approach to address air quality changes during the COVID-19 lockdown in Buenos Aires, Argentina", v1, Mendeley Data [data set], https://doi.org/10.17632/h9y4hb8sf8.1, 2021.

Faridi, S., Yousefian, F., Janjani, H., Niazi, S., Azimi, F., Naddafi, K., and Hassanvand, M. S.: The effect of COVID-19 pandemic on human mobility and ambient air quality around the world: A systematic review, Urban Clim., 38, 100 888, https://doi.org/https://doi.org/10.1016/j.uclim.2021.100888, 2021.

Feng, R., jun Zheng, H., Gao, H., ran Zhang, A., Huang, C., xi Zhang, J., Luo, K., and ren Fan, J.: Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China, J. Clean. Prod., 231, 1005–1015, https://doi.org/10.1016/j.jclepro.2019.05.319, 2019.

Freitas, S. R., Longo, K. M., Alonso, M. F., Pirre, M., Marecal, V., Grell, G., Stockler, R., Mello, R. F., and Gacita, M. S.: PREP–CHEM–SRC–1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models, Geosci. Model Dev., 4, 419 – 433, 2011.

Gaubert, B., Bouarar, I., Doumbia, T., Liu, Y., Stavrakou, T., Deroubaix, A., Darras, S., Elguindi, N., Granier, C., Lacey, F., Müller, J. F., Shi, X., Tilmes, S., Wang, T., and Brasseur, G. P.: Global Changes in Secondary Atmospheric Pollutants During the 2020 COVID-19 Pandemic, J. Geophys. Res.: Atmospheres, 126, 1–22, https://doi.org/10.1029/2020JD034213, 2021.

Gelman Constantin, J., Londonio, A., Bajano, H., Smichowski, P., and Gómez, D.: Plasma-based technique applied to the determination of 21 elements in ten size fractions of atmospheric aerosols, Microchem. J., 160, https://doi.org/10.1016/j.microc.2020.105736, 2021.

Ghaffarpasand, O., Beddows, D. C., Ropkins, K., and Pope, F. D.: Real-world assessment of vehicle air pollutant emissions subset by vehicle type, fuel and EURO class: New findings from the recent UK EDAR field campaigns, and implications for emissions restricted zones, Sci. Total Environ., 734, 139 416, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.139416, 2020.

Grange, S. K., Lee, J. D., Drysdale, W. S., Lewis, A. C., Hueglin, C., Emmenegger, L., and Carslaw, D. C.: COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban areas, Atmos. Chem. Phys., 21, 4169–4185, https://doi.org/10.5194/acp-21-4169-2021, 2021.

Heuss, J. M., Kahlbaum, D. F., and Wolff, G. T.: Weekday/Weekend Ozone Differences: What Can We Learn from Them?, J Air Waste Manag. Assoc., 53, 772–788, https://doi.org/10.1080/10473289.2003.10466227, 2003.

IGN: Mapas base de Argentina Bicontinental y Argentina Parte Continental Americana [data set], Available at https://www.ign.gob.ar/NuestrasActividades/InformacionGeoespacial/CapasSIG (2021/09/07).

Jiang, N. and Riley, M. L.: Exploring the utility of the random forest method for forecasting ozone pollution in SYDNEY, J. Environ. Protect. Sustainable develop, 1, 245–254, 2015.

Koo, B., Jung, J., Pollack, A. K., Lindhjem, C., Jimenez, M., and Yarwood, G.: Impact of meteorology and anthropogenic emissions on the local and regional ozone weekend effect in Midwestern US, Atmos. Environ., 57, 13–21, https://doi.org/https://doi.org/10.1016/j.atmosenv.2012.04.043, 2012.

Kroll, J. H., Heald, C. L., Cappa, C. D., Farmer, D. K., Fry, J. L., Murphy, J. G., and Steiner, A. L.: The complex chemical effects of COVID-19 shutdowns on air quality, Nat. Chem., 12, 777–779, https://doi.org/10.1038/s41557-020-0535-z, 2020.

Li, K., Jacob, D. J., Liao, H., Zhu, J., Shah, V., Shen, L., Bates, K. H., Zhang, Q., and Zhai, S.: A two-pollutant strategy for improving ozone and particulate air quality in China, Nat. Geosci., 12, 906–910, https://doi.org/10.1038/s41561-019-0464-x, 2019.

Liu, Y., Wang, T., Stavrakou, T., Elguindi, N., Doumbia, T., Granier, C., Bouarar, I., Gaubert, B., and Brasseur, G. P.: Diverse response of surface ozone to COVID-19 lockdown in China, Sci. Total Environ., 789, 147 739, https://doi.org/10.1016/j.scitotenv.2021.147739, 2021.

Masih, A.: Machine learning algorithms in air quality modeling, Glob. J. Environ. Sci. Manag., 5, https://doi.org/10.22034/gjesm.2019.04.0, 2019.

Muhammad, S., Long, X., and Salman, M.: COVID-19 pandemic and environmental pollution: A blessing in disguise?, Sci. Total Environ., 728, 138 820, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138820, 2020.

Nakada, L. Y. K. and Urban, R. C.: COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil, Sci. Total Environ., 730, https://doi.org/10.1016/j.scitotenv.2020.139087, 2020.

Nakada Kondo, L. Y. and Urban, R. C.: COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil, Sci. Total Environ., 730, 139 087, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.139087, 2020.

Ordóñez, C., Garrido-Perez, J. M., and García-Herrera, R.: Early spring near-surface ozone in Europe during the COVID-19 shutdown: Meteorological effects outweigh emission changes, Sci. Total Environ., 747, 141 322, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141322, 2020.

Pineda Rojas, A. L., Borge, R., Mazzeo, N. A., Saurral, R. I., Matarazzo, B. N., Cordero, J. M., and Kropff, E.: High PM10 concentrations in the city of Buenos Aires and their relationship with meteorological conditions, Atmos. Environ., 241, 117 773, https://doi.org/https://doi.org/10.1016/j.atmosenv.2020.117773, 2020.

Puliafito, S. E., Allende, D. G., Castesana, P. S., and Ruggeri, M. F.: High-resolution atmospheric emission inventory of the argentine energy sector . Comparison with edgar global emission database, Heliyon, p. e00489, https://doi.org/10.1016/j.heliyon.2017.e00489, 2017.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2019.

Rahman, M. M., Paul, K. C., Hossain, M. A., Ali, G. G. M. N., Rahman, M. S., and Thill, J.-C.: Machine Learning on the COVID-19 Pandemic, Human Mobility and Air Quality: A Review, IEEE Access, 9, 72 420–72 450, https://doi.org/10.1109/ACCESS.2021.3079121, 2021.

Reich, S., Magallanes, J., Dawidowski, L., Gómez, D., Grošelj, N., and Zupan, J.: An Analysis of Secondary Pollutants in Buenos Aires City, Environmental Monitoring and Assessment, 119, 441–457, https://doi.org/10.1007/s10661-005-9035-2, 2006.

Seguel, R. J., Morales S., R. G., and Leiva G., M. A.: Ozone weekend effect in Santiago, Chile, Environ. Pollut., 162, 72–79, https://doi.org/10.1016/j.envpol.2011.10.019, 2012.

Seinfeld, J. and Pandis, S.: Atmospheric Chemistry & Physics: From Air Pollution to Climate Change, Wiley, 1998.

Shi, X. and Brasseur, G. P.: The Response in Air Quality to the Reduction of Chinese Economic Activities During the COVID-19 Outbreak, Geophys. Res. Lett., 47, 1–8, https://doi.org/10.1029/2020GL088070, 2020.

Šimić, I., Lovrić, M., Godec, R., Kröll, M., and Bešlić, I.: Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon, Environ. Pollut., 263, https://doi.org/10.1016/j.envpol.2020.114587, 2020.

Srivastava, A.: COVID-19 and air pollution and meteorology-an intricate relationship: A review, Chemosphere, 263, 128 297, https://doi.org/https://doi.org/10.1016/j.chemosphere.2020.128297, 2021.

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., Kloog, I., Davoli, M., Michelozzi, P., and Bellander, T.: A random forest approach to estimate daily particulate matter, nitrogen dioxide, and ozone at fine spatial resolution in Sweden, Atmosphere, 11, https://doi.org/10.3390/atmos11030239, 2020.

Tobías, A., Carnerero, C., Reche, C., Massagué, J., Via, M., Minguillón, M. C., Alastuey, A., and Querol, X.: Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-CoV-2 epidemic, Sci. Total Environ., 726, 138 540, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138540, 2020.

Uria-Tellaetxe, I. and Carslaw, D. C.: Conditional bivariate probability function for source identification, Environ. Model. Softw., 59, 1–9, https://doi.org/https://doi.org/10.1016/j.envsoft.2014.05.002, 2014.

Velders, G. J., Willers, S. M., Wesseling, J., van den Elshout, S., van der Swaluw, E., Mooibroek, D., and van Ratingen, S.: Improvements in air quality in the Netherlands during the corona lockdown based on observations and model simulations, Atmos. Environ., 247, 118 158, https://doi.org/10.1016/j.atmosenv.2020.118158, 2021.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y. F., Li, L., and Zhang, L.: Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects, Sci. Total Environ., 575, 1582–1596, https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017.

Wang, W., Li, X., Shao, M., Hu, M., Zeng, L., Wu, Y., and Tan, T.: The impact of aerosols on photolysis frequencies and ozone production in Beijing during the 4-year period 2012–2015, Atmos. Chem. Phys., 19, 9413–9429, https://doi.org/10.5194/acp-19-9413-2019, 2019.

Yarwood, G., Stoeckenius, T. E., Heiken, J. G., and Dunker, A. M.: Modeling Weekday/Weekend Ozone Differences in the Los Angeles Region for 1997, J Air Waste Manag. Assoc., 53, 864–875, https://doi.org/10.1080/10473289.2003.10466232, 2003.

Yu, R., Yang, Y., Yang, L., Han, G., and Move, O. A.: RAQ–A random forest approach for predicting air quality in urban sensing systems, Sensors (Switzerland), 16, https://doi.org/10.3390/s16010086, 2016.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M. L., Zhang, M., and Di, B.: Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment, Environ. Pollut., 233, 464–473, https://doi.org/10.1016/j.envpol.2017.10.029, 2018.