



Estimating changes in air pollutant levels due to COVID-19 lockdown measures based on a business-as-usual prediction scenario using data mining models: A case-study for urban traffic sites in Spain

Jaime González-Pardo^a, Sandra Ceballos-Santos^a, Rodrigo Manzanas^b, Miguel Santibáñez^{c,d}, Ignacio Fernández-Olmo^{a,*}

^a Department of Chemical and Biomolecular Engineering, Universidad de Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain

^b Santander Meteorology Group, Dpto. De Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander 39005, Spain

^c Department of Nursing, Global Health Research Group, Universidad de Cantabria, Avda. Valdecilla s/n, 39008 Santander, Cantabria, Spain

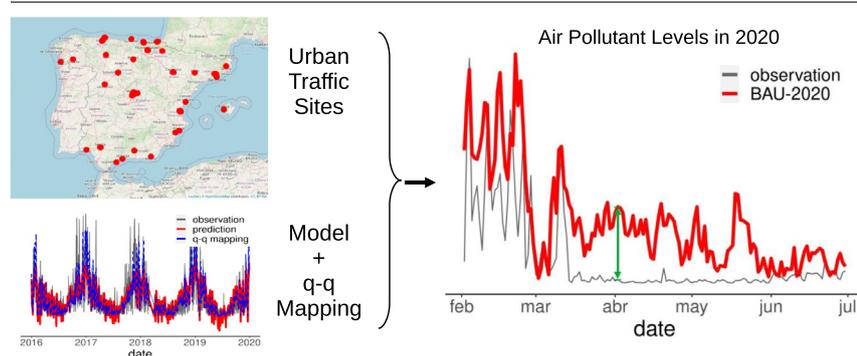
^d Research Nursing Group, IDIVAL, Calle Cardenal Herrera Oria s/n, 39011 Santander, Cantabria, Spain



HIGHLIGHTS

- COVID-19 lockdown restrictions have derived into a reduction pattern of NO_x in 2020.
- MLR, RF and KNN are proposed to take into account meteorological variability.
- Q-Q Mapping post-correction improves model performance and extreme event prediction.
- Smaller changes in air quality were observed from modelled BAU scenario.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 13 December 2021

Received in revised form 31 January 2022

Accepted 6 February 2022

Available online 10 February 2022

Editor: Pavlos Kassomenos

Keywords:

Air quality

COVID-19

Multiple Linear Regression

Random Forest

K-Nearest Neighbors

q-q mapping

ABSTRACT

In response to the COVID-19 pandemic, governments declared severe restrictions throughout 2020, presenting an unprecedented scenario of reduced anthropogenic emissions of air pollutants derived mainly from traffic sources. To analyze the effect of these restrictions derived from COVID-19 pandemic on air quality levels, relative changes in NO, NO₂, O₃, PM10 and PM2.5 concentrations were calculated at urban traffic sites in the most populated Spanish cities over different periods with distinct restrictions in 2020. In addition to the changes calculated with respect to the observed air pollutant levels of previous years (2013–2019), relative changes were also calculated using predicted pollutant levels for the different periods over 2020 on a business-as-usual scenario using Multiple Linear Regression (MLR) models with meteorological and seasonal predictors. MLR models were selected among different data mining techniques (MLR, Random Forest (RF), K-Nearest Neighbors (KNN)), based on their higher performance and accuracy obtained from a leave-one-year-out cross-validation scheme using 2013–2019 data. A q-q mapping post-correction was also applied in all cases in order to improve the reliability of the predictions to reproduce the observed distributions and extreme events. This approach allows us to estimate the relative changes in the studied air pollutants only due to COVID-19 restrictions. The results obtained from this approach show a decreasing pattern for NO_x, with the largest reduction in the lockdown period above –50%, whereas the increase observed for O₃ contrasts with the NO_x patterns

* Corresponding author.

E-mail addresses: jaime.diez@unican.es (J. González-Pardo), sandra.ceballos@unican.es (S. Ceballos-Santos), rodrigo.manzanas@unican.es (R. Manzanas), miguel.santibanez@unican.es (M. Santibáñez), ignacio.fernandez@unican.es (I. Fernández-Olmo).

with a maximum increase of 23.9%. The slight reduction in PM10 (−4.1%) and PM2.5 levels (−2.3%) during lockdown indicates a lower relationship with traffic sources. The developed methodology represents a simple but robust framework for exploratory analysis and intervention detection in air quality studies.

1. Introduction

The global spread of the COVID-19 pandemic in early March 2020 has become the largest health crisis in recent years, comparing it to the Spanish flu of the early 20th century (Munnoli et al., 2020; Sanchez-Lorenzo et al., 2021). In response to the high mortality caused by the SARS-CoV-2 virus and its rapid spread, governments around the world declared severe control measures and restrictions. Some of the restrictions implemented included mobility limitations or social distancing regulations. Many countries had to reinforce lockdown measures to full lockdown for several weeks due to the emergency situation in which they found themselves, with the closure of non-essential jobs and the reduction of human mobility to essential activities, such as food supply. In addition, remote work was implemented in many companies to minimize the economic crisis, keeping it beyond the lockdown period. The restrictions derived from the COVID-19 pandemic have presented an unprecedented scenario of the reduction of anthropogenic emissions of air pollutants derived mainly from the reduction in traffic due to mobility restrictions.

The unforeseen decrease in emissions of air pollutants associated with traffic activity due to the COVID-19 pandemic and mobility restrictions for several months throughout 2020 has presented a unique opportunity to study the impact of a large-scale intervention in human activity on air quality. Many scientific studies appeared even a few months after lockdown periods in which changes in pollutant levels were analyzed and quantified. However, quantifying the effect of COVID-19 pandemic restrictions is not straightforward, starting with the selection of the base case used to derive changes in concentrations, which should include as many external factors that would affect air quality levels as possible to isolate the COVID-19 effect. The most widely used methodology to quantify the effect of changes in pollutant concentrations is the comparison with a reference measurement period (Gkatzelis et al., 2021). This has been done using two main approaches depending on the measurement period used: (i) Some studies have compared the concentrations observed during the lockdown with those observed in the previous period, quantifying the changes through the differences between the week before and after the lockdown (Stratoulias and Nuthammachot, 2020; Tobías et al., 2020; Martorell-Marugán et al., 2021). (ii) Another reference measurement period used was the observations of previous years in the same period as the lockdown, obtaining the differences on the same day between different years (Chauhan and Singh, 2020; Mendez-Espinosa et al., 2020; Al-Abadleh et al., 2021). These methodologies allow to identify and quantify empirical relative changes with a very simple analysis. However, both approaches have some disadvantages in relating these differences to the effects of the COVID-19 pandemic. The first approach compares different weeks of the same year, without taking into account the effects of seasonality. Differences in pollutant concentrations between seasons have been widely studied (Sillanpää et al., 2006; Pérez et al., 2008; Rogula-Kozłowska et al., 2012) showing for example differences in PM2.5 concentrations of 12 $\mu\text{g}/\text{m}^3$ in summer and 60 $\mu\text{g}/\text{m}^3$ in winter in Kraków, Poland (Samek et al., 2018). The second approach corrects for these uncertainties that arise due to seasonality by comparing similar time periods between years. Observations from 2020 are compared to measurements of pollutants levels from 2019, and often several previous years, over the same time period. Seasonal effects for similar periods in different years should not differ, reducing the uncertainties due to this. However, some uncertainties may still arise due to meteorological variability, including exceptional events.

Air pollution levels can be strongly affected by meteorological factors. In addition, emissions from some anthropogenic sources can increase or decrease because of the changes in their activity due to meteorological conditions: e.g. increase in road traffic on rainy days or domestic heating in cold

seasons (Li et al., 2021). On the other hand, meteorological variables such as wind or precipitation can affect the immission of some pollutants by increasing their dilution, transport or dispersion (Heinsohn and Kabel, 1998). Several studies have analyzed the importance of meteorological effects on air quality in different scenarios, showing some differences when they were taken into account (Carslaw and Taylor, 2009; Barmpadimos et al., 2011; Henneman et al., 2015). Recent studies have applied several approaches to reduce the effects of meteorology on the interpretation of air quality changes due to the COVID-19 pandemic by using statistical and data mining techniques. In particular, two main approaches have been used to quantify the effects of meteorology on the analysis of pollutant changes. First, instead of comparing 2020 observations with a base case such as the methodology described above, some authors used meteorological normalisation to deseasonalise and deweather air pollutant reference time series (Grange and Carslaw, 2019; Ropkins and Tate, 2021). This approach consists of reducing the variability in the air quality time series by training a model with independent variables that can explain the variability and eliminate its influence. The result is a time series with invariant meteorological conditions that can then be exposed to further exploratory analysis. Typically, the models used to explain variability are based on machine learning techniques that provide the partial influence of each weather variable, using the entire time series to train the model. Then, one of the methodologies described above should be used in addition to meteorological normalisation to quantify the related changes with respect to a base case in order to remove uncertainty derived from meteorological conditions.

Another approach used to account for the influence of meteorology is to predict pollutant concentrations during the period of intervention (i.e. the period when COVID-19 restrictions took place) using statistical models developed from training data sets of observed air quality levels and meteorological values from previous years. Such predictive models have been used in previous studies to isolate either short- or long-term interventions on air quality levels (Carslaw et al., 2012; Bennett et al., 2013; Vu et al., 2019). This approach is applied in this work to quantify the relative changes in air quality levels due to COVID-19 pandemic-related restrictions. First, developed models can predict pollutant concentrations assuming a business-as-usual (BAU) scenario. The BAU scenario uses meteorological data from 2020 and accounts for seasonality differences, disregarding the effects of the COVID-19 pandemic. Therefore, the 2020 prediction can be used as a base case for comparison with actual observed concentrations, providing a true measure of the relative changes derived from the restrictions implemented during the COVID-19 pandemic (Fu et al., 2021; Kowalski et al., 2021).

Many statistical and data mining techniques can be used to analyze the meteorological influence on air quality, with linear regression and Classification and Regression Trees (CART) being the most widely used, both of which allow to carry out the two approaches described above. Multiple Linear Regression (MLR) methods have been widely applied to predict and model atmospheric pollutant concentrations (Zhong et al., 2018; Wong et al., 2021). The predictions resulting from these models can then be used as BAU scenario for the second approach. An advantage of MLR is that it is possible to interpret the regressed coefficients (β) as the partial influence of each meteorological variable on air quality levels, which provides another way to study the effect of a variable on pollutant concentrations. These β coefficients not only allow to perform the first approach described above, but the influence of a certain event or condition can be obtained by aggregating it as a dependent variable. In this way, Briz-Redón et al. (2021) show how lockdown restrictions derived from the COVID-19 pandemic have affected air quality through the partial influence of a lockdown variable. Briz-Redón et al. (2021) compare the accuracy of MLR when lockdown is accounted for as a binary variable in the analytical

expression. However, lockdown restrictions have evolved throughout 2020 differently between cities, making it difficult to characterize them in a single dependent variable. On the other hand, CART methods generally provide a more accurate prediction being able to model non-linear relationships but having a simple and useful interpretation of the relationship between variables. Furthermore, ensemble techniques based on CART (e.g. random forests) are commonly used in prediction and classification studies, whereby a BAU scenario predicted by meteorological data can be obtained to assess the COVID-19 effect (Grange et al., 2021; Jephcote et al., 2021; Lovrić et al., 2021). These methods can be also used to perform meteorological normalisation (Grange et al., 2018; Shi et al., 2021).

However, unambiguous quantification of the impact of COVID-19 on air quality may not be possible due to the difficulty of knowing whether changes were due to mobility restrictions or other factors that were not taken into account. This uncertainty can be reduced by an appropriate selection of air quality stations, e.g. choosing those closest to the emission sources that are likely to be most influenced by changes in the emission patterns, such as urban traffic sites.

The aim of this work was to estimate the relative changes in air quality levels only due to COVID-19 lockdown measures; to calculate the relative changes for the different periods representing distinct restriction measures over 2020, observed daily concentrations of NO, NO₂, O₃, PM10 and PM2.5 at urban traffic stations in Spanish cities with more than 100,000 inhabitants were compared with predicted daily concentrations based on a BAU scenario using MLR models with meteorological and seasonal predictors. MLR models were selected among different statistical models (MLR, Random Forest (RF), K-Nearest Neighbors (KNN)), based on their higher performance and accuracy obtained from a training and cross-validation procedure of a relatively large data series (2013–2019).

2. Data and methods

The data sources used in this work are publicly accessible in a persistent data repository (González-Pardo et al., 2021b, doi:<https://doi.org/10.5281/zenodo.5642868>). On the other hand, the software and scripts developed in this work were carried out using R (R Core Team, 2021) and Python (Van Rossum, 1995); and are available at the GitHub repository: González-Pardo et al., 2021a, <https://github.com/Jaimedgp/AirQualityCOVID>.

This study was focused on urban traffic sites from the most populated Spanish cities (with more than 100,000 inhabitants). Traffic emission should have been strongly affected by the COVID-19 closure restrictions, being even more noticeable in larger cities with higher traffic under normal conditions. Moreover, ground-level measurements provided by these air quality monitoring sites should be more sensitive to changes in emission source changes and are more relevant to human health.

The entire year 2020 has been divided into five periods according to the evolution of the COVID-19 restrictions in Spain: the pre-lockdown period, from January 1, to March 13; the lockdown period, from March 14, the date when the Spanish government declared the total lockdown, to May 1; the de-escalation period, from May 2, to June 20, when each city reduced mobility restrictions according to their COVID-19 impact, following the government's de-escalation plan; the "normality" period, from June 21, when all Spanish communities already passed all de-escalation phases, with the Spanish government abolishing all mobility restrictions and ending the state of alarm declared in March, to October 24; and the second lockdown, from October 25, when the Spanish government had to declare a new lockdown, less restrictive than the first one, allowing each Region to make changes to the mobility restrictions. The first lockdown was the only period where the same measures applied to all the Spanish Regions.

Single-site models were developed for each pollutant in each air quality station. To build the prediction models, a key data set with pollutant concentrations, as dependent variable, and meteorological data, as independent variable, was prepared for each air quality station.

2.1. Data used in this study

2.1.1. Air quality data

Concentrations of nitrogen monoxide (NO), nitrogen dioxide (NO₂), ozone (O₃), particulate matter of less than 10 µm (PM10) and particulate matter of less than 2.5 µm (PM2.5) have been downloaded from the European Environment Agency (EEA) (See Table S.1). Road traffic is the main source of NO and NO₂ (Seinfeld and Pandis, 2016), being also strongly related to PM2.5 levels. Since many traffic sites did not have PM2.5 data, PM10 concentrations were also included in the study, because PM2.5 is a fraction of PM10. Although O₃ is a secondary pollutant, it was also included due to its interaction with NO_x in the photolytic cycle (Heinsohn and Kabel, 1998).

Daily pollutants concentrations time series from 2013 to 2020 have been obtained from EEA using the *saqgetr* package for R (Grange, 2019). Despite the reliability of the source, some negative values of concentrations, with no physical meaning, were shown being necessary to preprocess the data removing those values. After the preprocess, only air quality data, by pollutant, were retained when there were observations available for more than 3 years and at least the 80% of daily data between March 2020 and June 2020. These months include the entire time-period of lockdown and de-escalation phases in Spain. After excluding some pollutants' time series, 210 site-pollutant pairs from 60 air quality monitoring sites were selected for the study (Table S.2), covering 34 Spanish cities (see Fig. 1) with a total population of 13,481,269 inhabitants (Instituto Nacional de Estadística, 2020).

Time series from 2013 until 2019 have been used to model development, reserving 2020 observations for the analysis of changes in air pollutant levels. For this purpose, some outliers have been removed from the training time series, defined as values outside 5 times the interquartile range. These outliers are not closely related to meteorological events, and would be difficult to model by the selected predictors.

2.1.2. Meteorological data

Daily meteorological data (Table S.3) have been obtained from the nearest location to the selected air quality stations with at least 80% of available records between 2013 and 2020. A good selection of predictors with high explanatory power of air quality can increase the accuracy and performance of the model. Hence, meteorological variables with high influence in pollutants levels were included in the study (Carslaw and Taylor, 2009).

In particular, daily temperature (maximum, mean and minimum; °C), precipitation (mm) and surface pressure (maximum and minimum; hPa) were downloaded from the OpenData platform of the Agencia Estatal de Meteorología (AEMET) through its Application Programming Interface (API) using the *pyaemet* python framework developed for this purpose (González-Pardo, 2021). The measured temperature variability due to seasonal effect will account for the annual cycle of observed pollutant levels in their time series. On the other hand, precipitation will affect PM10 and PM2.5 deposition rates leading to lower levels in ambient air. The surface wind speed (m/s) and direction (in degrees, being 90° for East) were included due to their influence in pollutants transport that will affect the local measurements. Both were obtained from the National Oceanic Atmospheric Administration (NOAA) using the *worldmet* R package (Carslaw, 2020). In addition, daily solar radiation (W/m²) and relative humidity (%) were downloaded from the ERA5-Land reanalysis data set (Muñoz-Sabater, 2019). Solar radiation has been included because of the influence of photochemistry on ozone formation from primary air pollutants.

2.2. Model development

A wide range of statistical and data mining techniques have been used in many different studies for air quality models development, depending in many cases on the study objectives for both future predictions and meteorological normalisation. In this study, three different techniques have been tested to obtain the best model to predict the 2013–2019 data set of the

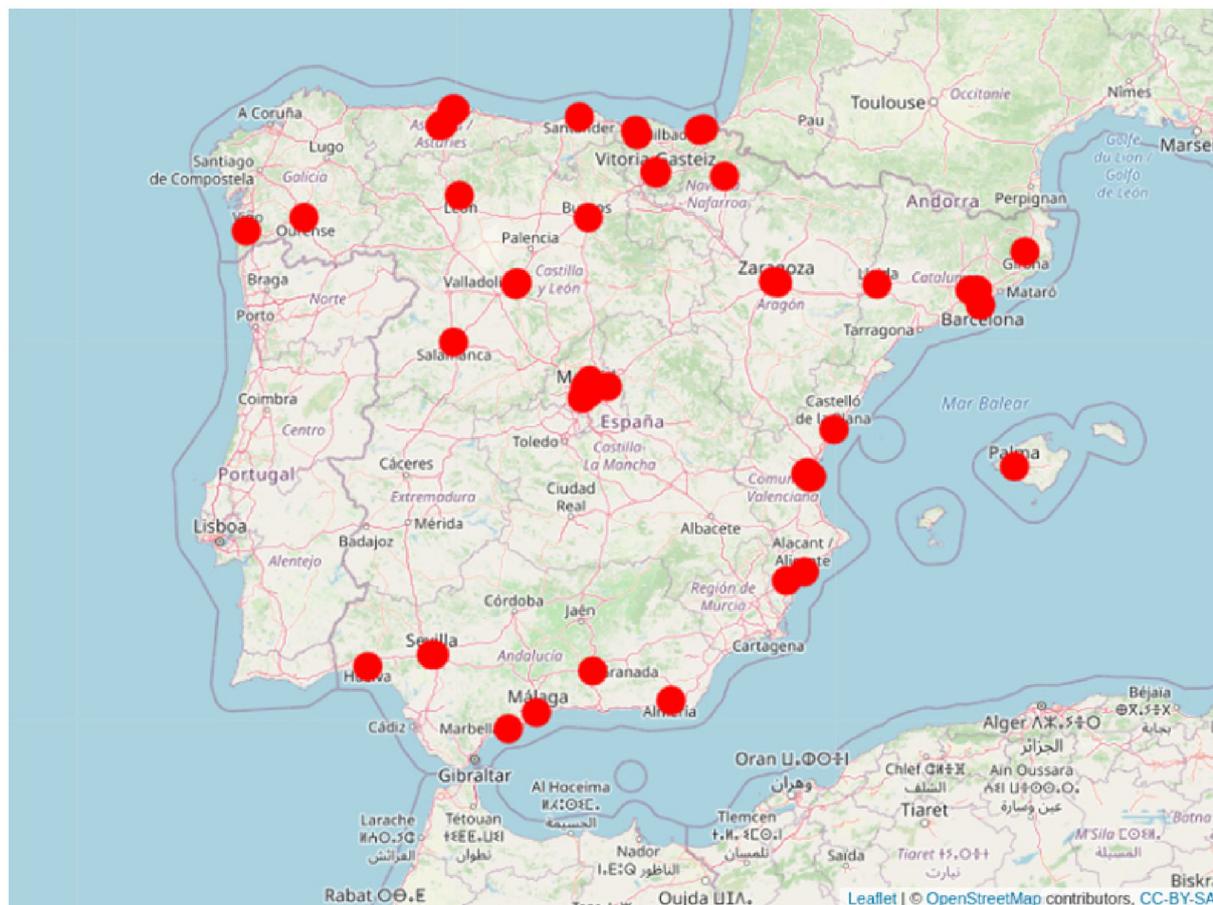


Fig. 1. Map showing the urban traffic sites selected for this study.

studied air pollutants, namely MLR, RF and KNN. The best model obtained will be used to estimate 2020 air quality levels under a BAU scenario, and calculate the concentration changes in the selected air pollutants in traffic sites.

MLR is a simple statistical technique that allows for modelling the linear relationship that occurs between one dependent, predictand variable (the concentration of each studied pollutant in this work) and a number of independent, predictor variables (meteorological data here) through a set of coefficients β (James et al., 2013). Since both the independent and dependent variables are known, these coefficients can be adjusted by minimizing the residual sum of squares. In order to assess the effect that persistent meteorological conditions may have on the concentration of the different pollutants, we have analyzed the potential added value of incorporating predictor information from up to 3 days prior to the target day being predicted in our MLR models, which were fitted using the *lm* R function.

Linear regression models are often used as a good benchmark for more modern data mining techniques such as CART, which are commonly used in forecasting and classification studies (Petetin et al., 2020a; Fabregat et al., 2021). Tree-based methods model the independent variable by segmenting and stratifying the space of the dependent variables, thus being able to account for non-linearities in the predictor-predictand link. This is done by means of certain rules imposed on each node that divide the predictor space into two branches, doing it iteratively until a terminal node (or leave) is obtained. However, trees are very sensitive to the training data partition used to build the tree and can easily tend to overfit if no constraints are implemented, such as pruning, which can worsen the predictive performance of the trees. New ensemble techniques have been developed to improve CART by aggregating many trees following a bagging approach -all trees are grown in parallel without any pruning constraints-, which allows to increase the predictive power of individual trees while reducing

the risk of overfitting the training data. However, if there is a very strong predictor, most or all of the bagged trees will use it in the top split, thus being all models highly correlated. To decorrelate the trees, Ho (1995) proposed the RF technique, where only a subset of predictors (m_{try}) out of the total available ones (p) are used at each split; $m_{try} \approx p/3$ is typically used in regression problems. These predictors are randomly chosen by means of bootstrapping. In this study, the RF technique using the *randomForest* R package (Liaw and Wiener, 2002) was applied to predict air quality levels from meteorological data. A sensitivity analysis to the number of trees in the forest has been performed for 10, 20, 30, 50, 70, 100 and 200 trees in order to obtain the best model configuration for our target problem.

Another data mining technique tested in this work was KNN, which has been already used in previous studies to predict drought anomalies (Jiang et al., 2021). The KNN algorithm estimates the target predictand variable using the observed values corresponding to the K closest predictor configuration in the training data set. In particular, in this study, the air quality levels on a given day are estimated based on the observations recorded on the K days with the most similar meteorological conditions. This algorithm is very sensitive to the value of K used. For lower K, the model will fit better in the training set tending to overfit. As the K value grows, the generalisation power of the model increases, improving the accuracy on the test data set but reducing the variance of the predictions. Hence, a sensitive analysis of the K parameter has been performed for 1 to 10 nearest neighbors in order to obtain the best KNN configuration for air quality prediction. A disadvantage of the KNN algorithm is that the closeness metric, defined by the Euclidean distance, is calculated equally for all predictors, regardless of their relative explanatory power. Therefore, the inclusion of less explanatory variables may result in a loss of precision.

In order to minimize the bias-variance trade-off (Von Luxburg and Schölkopf, 2011; Friedman et al., 2017), an empirical quantile-quantile

Table 1

The five periods into which the year 2020 has been divided according to the evolution of the COVID-19 pandemic restrictions in Spain.

Periods	Starting dates
Pre-lockdown	2020-01-01
Lockdown	2020-03-14
De-escalation	2020-05-01
Normality	2020-06-21
Second lockdown	2020-10-25

mapping (q-q mapping) post-correction has been applied to the three techniques described above. To do so, the R package *downscaleR* (Iturbide et al., 2018) was used, which allows for adjusting a set of n quantiles of the probability density function (PDF) of the prediction to the corresponding observed empirical PDF (Déqué, 2007). This technique has been widely used by the climate community to match the low-resolution outputs of global numerical models to the available observations (Marau, 2013; Manzanos et al., 2018).

To assess the predictive performance of the three techniques considered, a leave-one-year-out cross-validation scheme was applied for 2013–2019 period using the *caret* R package (Kuhn, 2021). Only 10 representative sites per pollutant were selected for this cross-validation experiment due to the high computational cost, focusing on those sites with more available data, located in high population cities and providing a good representation of spatial variability (Table S.4). The performance comparison was made based on the Root Mean Square Error (RMSE), the bias ratio (predicted mean divided by the observed one), the variance ratio (predicted variance divided by the observed one) and the Pearson correlation. Moreover, since some of the studied pollutants present a pronounced annual cycle, an unseasonalised correlation was also considered. Then, the overall best model found according to these performance metrics was employed to predict the 2020 daily air pollutant concentrations and analyze the changes in such levels due to the COVID-19 restrictions.

The quantification of the COVID-19 pandemic restrictions' effect was done by means of the daily percentage relative change (RC_i) for each pollutant between the daily air quality observations from 2020 and a daily reference concentration using the following equation:

$$RC_i[\%] = 100 \cdot \frac{C_{i,obs} - C_{i,ref}}{C_{i,ref}} \quad (1)$$

where for each pollutant $C_{i, obs}$ is the observed air quality level on day i in 2020, and $C_{i, ref}$ is the reference air quality level on the same day i . Two approaches have been studied for this reference concentration, resulting in two relative changes, RC and RC^* : (i) RC , using the average observations in 2013–2019 ($C_{i, ref} = C_{i, 2013-2019}$), and (ii) RC^* , using the 2020 prediction obtained by the best-fit model using meteorological data ($C_{i, ref} = C_{i,BAU-2020}$). The difference between RC and RC^* explains the influence of meteorological variability on air quality during 2020. Analyses of the changes in air quality levels by COVID-19 were performed for the five periods described in Table 1 according to the evolution of the restrictions, obtaining the average RC and RC^* for each pollutant and each period at each site. A workflow of the methodology used in this work is shown in Fig. S.1.

3. Results and discussion

3.1. Observed changes in air pollutant levels

Before accounting for meteorological variability, the observed relative change (RC) of each pollutant was obtained by comparing the 2020 observations with the average concentrations during the same periods of the previous years ($C_{i, ref} = C_{i, 2013-2019}$), according to Eq. (1). The average RC was then obtained for the five periods described in Table 1 for each studied city. For cities with several urban traffic sites, an average RC was calculated. Box plots of the observed RC from the 34 studied cities are shown in Fig. 2, and the average and standard deviation for all of these cities is presented in Table 2.

Although some variabilities between cities can be observed in Fig. 2, on average the analysis of the RC shows a decrease in NO and NO₂ concentrations during all periods affected by the pandemic, being higher for NO. The largest reduction for both pollutants was observed during the lockdown period (−60% and −55.1% respectively) while in the de-escalation and normality periods the 2020 observations increased back to values close to previous years. Finally, in the last period after the second lockdown, contaminant levels decreased again but with less reduction than in the first lockdown (−36% and −27.6% respectively). These abrupt changes in NO_x concentrations during the lockdown period were consistent with previous studies using the same methodology. Similar RC were observed in Madrid and Barcelona (Baldasano, 2020), Valencia (Donzelli et al., 2021) and in some southern European cities such as Nice, Rome, Valencia and Turin (Sicard et al., 2020).

In contrast, O₃ shows similar but opposite behaviour to that found for NO and NO₂, with the mean concentration increasing in lockdown period

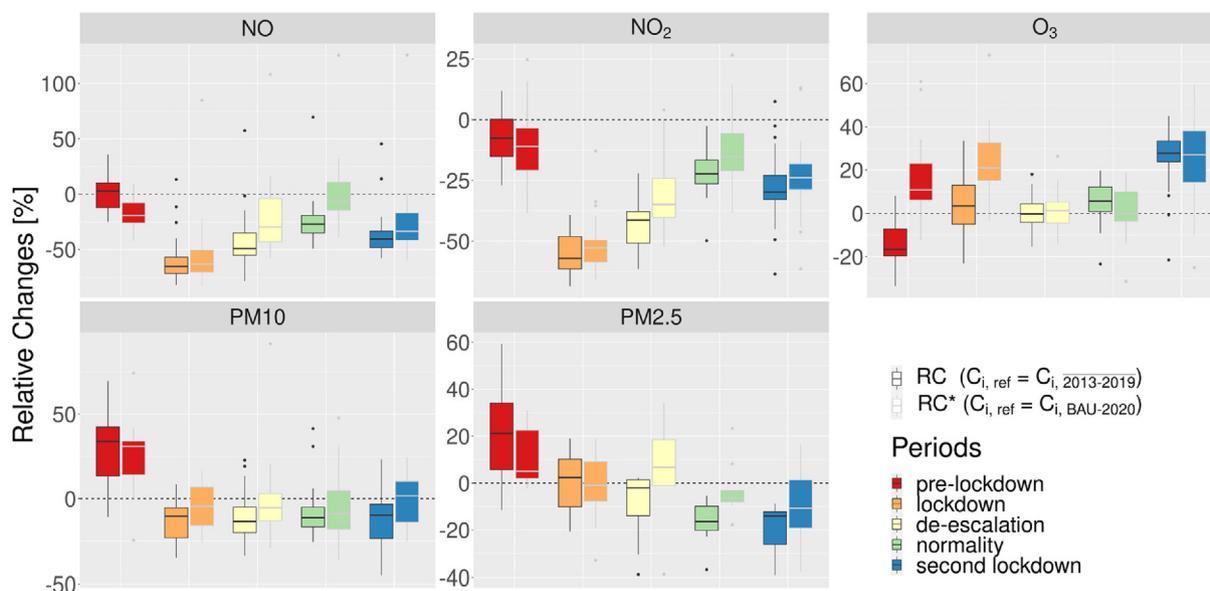


Fig. 2. Box plots of the relative change of air pollutants in 2020 with respect to 2013–2019 observations (RC, black border) and with respect to predicted values (RC*, grey border). The filled colour represents each studied period (Table 1); the line inside the boxes represents median values.

Table 2

Average relative percentage change RC [%] and standard deviations of NO, NO₂, O₃, PM10 and PM2.5 concentrations in 2020 respect of 2013–2019 values (C_{2013–2019}) at studied cities for each period.

Periods	NO [%]	NO ₂ [%]	O ₃ [%]	PM10 [%]	PM2.5 [%]
Pre-lockdown	0.2 ± 14.2	-7.8 ± 9.7	-13.8 ± 10.2	30 ± 22.6	21.2 ± 21.4
Lockdown	-60.1 ± 19.5	-55.1 ± 8.3	4.4 ± 12.9	-12.9 ± 12.7	0.2 ± 13.4
De-escalation	-43.2 ± 23.8	-42.5 ± 9.4	-0.7 ± 8.3	-10.8 ± 15.4	-9 ± 15.5
Normality	-23.9 ± 20.2	-21.5 ± 8.8	5.3 ± 9.3	-7.6 ± 16.3	-16.7 ± 9.6
Second lockdown	-36 ± 20.3	-27.6 ± 13.5	24.5 ± 14.8	-11.1 ± 17	-18.9 ± 10.3

and decreasing in de-escalation. However, for O₃ the largest increase was obtained in the second lockdown period, with a previous increase observed in normality period. This increase in O₃ concentrations opposite to the NO_x decrease was also obtained by some authors (Collivignarelli et al., 2020; Ropkins and Tate, 2021; Sicard et al., 2020) being in our study lower in magnitude. PM10 concentrations did not change significantly during the pandemic with averages ranging from -7.6% at normality to -12.9% at lockdown. However, these values represent a large decrease compared to the levels of the pre-lockdown period with an average RC of 30%. Similarly, PM2.5 presents a smooth decreasing trend throughout 2020 ranging from 21.2% in pre-lockdown to -18.9% in the second lockdown.

Since in urban traffic sites NO and NO₂ are mainly emitted by road traffic sources, the reduction in their concentration observed in 2020 can be explained by the mobility restrictions implemented during the lockdowns. In the first months of 2020, when no restrictions were implemented, the observed concentrations of NO and NO₂ were similar to the previous years. The reductions in NO_x were greater in lockdown and second lockdown periods as the mobility restrictions and enclosure were declared. During the first lockdown, more dramatic restrictions were implemented with total lockdown and the closure of non-essential jobs resulting in the greatest reductions in NO_x concentrations. The impact of traffic restrictions during COVID-19 lockdown on NO₂ levels in three European urban megacities (London, Milan and Paris) has been recently studied by Collivignarelli et al. (2021). These authors reported higher reductions in NO₂ levels in London and Paris (between -65.7 and -80.8%) and smaller ones in Milan (-8.6 to -42.4%) with respect to the average of Spanish traffic sites (-55%), although a large variability was found at the studied Spanish cities, from a minimum of -12.9% in Burgos to a maximum of -65.8% in Palma de Mallorca. Ceballos-Santos et al. (2021) also analyzed the relative change in NO₂ concentration during lockdown at the urban traffic sites of a small Spanish region (Cantabria), finding an average reduction of -56.1%. They also found a change in the daily behaviour of NO and NO₂ levels, drastically reducing the peaks corresponding to the rush hours, around 8 a.m. and 8 p.m., when arrivals and departures in work environments, educational centres, and shopping centres occur.

In the periods of de-escalation and normality, mobility restrictions were gradually reduced until they were abolished. The return of human mobility led to an increase in road traffic, which explains the smaller reductions observed in both periods. Since NO is more predominant than NO₂ in vehicle emissions, the greater reduction observed for NO suggests that the reductions obtained in 2020 were mainly due to the traffic restrictions implemented due to the COVID-19 pandemic.

On the other hand, the near-mirror image of O₃ and NO behaviour can be explained by the role of NO in the O₃ titration process (NO + O₃ → NO₂ + O₂). Therefore, decreasing NO emissions and concentration could lead to a reduction of O₃ consumption, which would result in an increase of its concentrations (Briz-Redón et al., 2021; Ropkins and Tate, 2021). However, the pattern of RC observed was less clear compared to NO and NO₂, being lower in magnitude. Since O₃ is a secondary pollutant produced mainly in the presence of NO_x, volatile organic compounds (VOCs) and solar radiation, this discrepancy can be attributed to changes in its precursors (Martorell-Marugán et al., 2021; Sicard et al., 2020; Siciliano et al., 2020; Adam et al., 2021). While similar reductions in NO_x emissions have occurred worldwide during the pandemic, different decreases in VOCs emissions have resulted in different VOCs/NO_x ratios, leading to different rates of O₃ production. Furthermore, since meteorological variability was

not taken into account in this section, the O₃ changes observed in Fig. 2 may also be due to a change in solar radiation during these periods with respect to the same periods of the reference years (Adams, 2020; Tobías et al., 2020).

The increase observed in pre-lockdown for PM10 concentrations in a period supposedly similar to the values of previous years and the abrupt change observed in the next period can be explained by some exceptional events that occurred in this period, likely the intrusion of Saharan dust.

However, although the RC obtained are in line with the expected effect of mobility restrictions on air quality, the comparison of the 2020 observations with the 7-year baseline values does not take into account any factors that may influence the change in air quality behaviour compared to previous years. Therefore, to isolate the COVID-19 lockdown effect, meteorological variability was taken into account by modelling air quality using the historical relationship between pollutant concentrations and daily meteorological values.

3.2. Predictive model selection: cross-validated results

The performance of MLR, RF and KNN was assessed under a cross-validated framework (see Section 2.2) to select the best model configuration to predict 2020 air quality data. The cross-validated results obtained for MLR with regards to the sensitivity to meteorological persistence are shown in Fig. S.2. All pollutants show similar bias ratios close to 1 (the ideal value) regardless meteorological persistence is considered or not. The other four performance metrics obtained show an improvement when meteorological data from previous days are introduced, especially when the previous day ($t = 0, -1$) is taken into account, with lesser improvements obtained for the two previous days ($t = 0, -1, -2$) and 3 days before ($t = 0, -1, -2, -3$). It is worth noting the poorer results obtained for the variance ratio with values below 0.5 (except for O₃), which are far from the ideal value of 1. The best results, except for RMSE, were obtained for O₃ with variance ratio of 0.63 and correlation of 0.76. The good performance of MLR for O₃ in bias and variance can be explained by its almost normal distribution, while the good correlation may be due to its strong annual cyclical nature that can be described by the meteorological variables used. On the other hand, the worst correlations and variance results were obtained for PM10 and PM2.5 This may suggest that these pollutants are the least related to the meteorological variables selected as predictors. In addition, unlike O₃, PM10 and PM2.5 do not have a strong annual cycle, and their time series have a more irregular profile, being more difficult to model using MLR. Furthermore, compared to NO_x and O₃, a limited number of urban traffic sites had measurements of PM10, and mainly PM2.5, reducing the size of the training data set.

For RF, the cross-validated results obtained with regards to the sensitivity to the number of trees considered (ntree) are shown in Fig. S.3. As for MLR, the bias ratio is around 1 for all pollutants and ntree values, with the highest value for NO. Also, the variance ratios are lower than expected with values around 0.5, obtaining the best result of variance and correlations for O₃ and the worst for PM10. Fig. S.3 shows an improvement in both correlations and RMSE values as the number of trees increases. However, this improvement results in a lower and worse variance ratio, which can be explained by the bias-variance trade-off mentioned in Section 2.2.

Finally, the cross-validated results obtained for KNN in relation to the choice of K are shown in Fig. S.4. Unlike MLR and RF, KNN was able to achieve variance ratios close to 1 for all pollutants when only the nearest

neighbor was used ($K = 1$). However, this is at the cost of delivering low correlations and high RMSE values. Nevertheless, as K increases, both correlations and RMSE improve significantly, while obtaining lower variance ratios.

Figs. S.2 to S.4 show the worst results for the variance ratio for the three data mining techniques studied, indicating the difficulty for modelling the distribution of air quality observations. These weak results can be explained by the poor capacity of the models to predict extreme pollution events such as the high NO and NO₂ concentrations observed during winter periods. Therefore, a q-q mapping post-correction was applied in all cases in order to improve the reliability of the predictions to reproduce the observed distributions. An example of the q-q mapping correction of the probability density function for the MLR at the Escuelas Aguirre site, Madrid, is shown in Fig. 3.

The cross-validated results obtained for the three studied techniques after the q-q mapping correction are shown in Figs. S.5–S.7. As expected, the variance ratios show a significant improvement, with values close to 1 for MLR and KNN, and around 0.7 for RF. The other metrics have similar results to those obtained without q-q mapping, with a slight improvement in correlations and a higher RMSE for some models. Based on these results, the best model configuration was defined for each technique. In particular, results are obtained from the MLR with 3 days of meteorological persistence ($t = 0, -1, -2, -3$), the RF with 100 trees and KNN with $K = 10$; the performance metrics obtained for these models are shown in Fig. 4, comparing “raw” (i.e. prior to q-q mapping) predictions and after q-q mapping post-correction. Although the “raw” predictions present nearly negligible biases with respect to observations regardless of the data mining technique used, and the correlations and the RMSE attained are also good, the variance of these “raw” predictions is in general low. The q-q mapping technique is able to improve the predictions' variance so that it becomes similar to that of the observations. This ensures that our q-q mapping is not acting as a predictive technique (the real predictive capacity comes from the set of predictors used, which is skillfully exploited by the data mining techniques proposed) but just as a post-correction, which improves the representation of high-order moments (including not only the variance, but also the extremes) in the predictions.

The comparison of the three techniques shows both for correlation and for RMSE the worst results for the KNN model. However, the variance ratio

obtained for KNN remains close to 1, with the RF being the lowest with values around 0.7. Overall, the best trade-off for all the studied metrics once applied the q-q mapping procedure is obtained for MLR model with 3 days of meteorological persistence, and it was therefore selected for the 2020 prediction as the BAU scenario to analyze the lockdown effect due to the COVID-19 pandemic. The better performance obtained for the MLR model with respect to other statistical techniques such as RF was also observed by Venter et al. (2020) using similar predictors. Fig. 5 shows an example of the observed and predicted time series (2013–2019) at the Escuelas Aguirre site (Madrid) for the 5 studied pollutants.

3.3. Comparison between observed and estimated relative changes in air pollutant levels

Following the results from the previous section, daily concentrations for 2020 were predicted from meteorological variables using the MLR technique with q-q mapping post-correction. These predictions were used as BAU scenario to compare with observed concentrations in 2020, using Eq. (1) ($C_{i,ref} = C_{i,BAU-2020}$) to calculate relative changes when meteorological variability was taken into account (RC^*).

The estimated RC^* obtained from the predicted values are compared with the observed RC in Fig. 2. Observed and estimated relative changes show a similar pattern during the periods affected by the pandemic for the studied pollutants, mainly for NO and NO₂. The lockdown period shows the largest reduction for both pollutants, followed in the next two periods by an increase back to values close to BAU and ending with a second reduction in the last period. However, the reduction observed in these periods using the 2020 predictions is smaller in magnitude than that obtained with previous years' values averaging from -60.1% for NO and -55.1% for NO₂ in the lockdown period using $C_{2013-2019}$ as reference to -54.7% and -51.3% respectively (see Table 3). These new values were similar to those obtained in previous studies with machine learning techniques (Petetin et al., 2020a, 2020b). Furthermore, the differences in relative changes between methodologies were consistent with the results obtained by Venter et al. (2020) and Shi et al. (2021). These differences are larger for the summer periods of de-escalation (21.8% for NO and 11.2% for NO₂) and normality period (23.8% for NO and 8.8% for NO₂), being also larger in magnitude for NO in all periods. In fact, this methodology allows

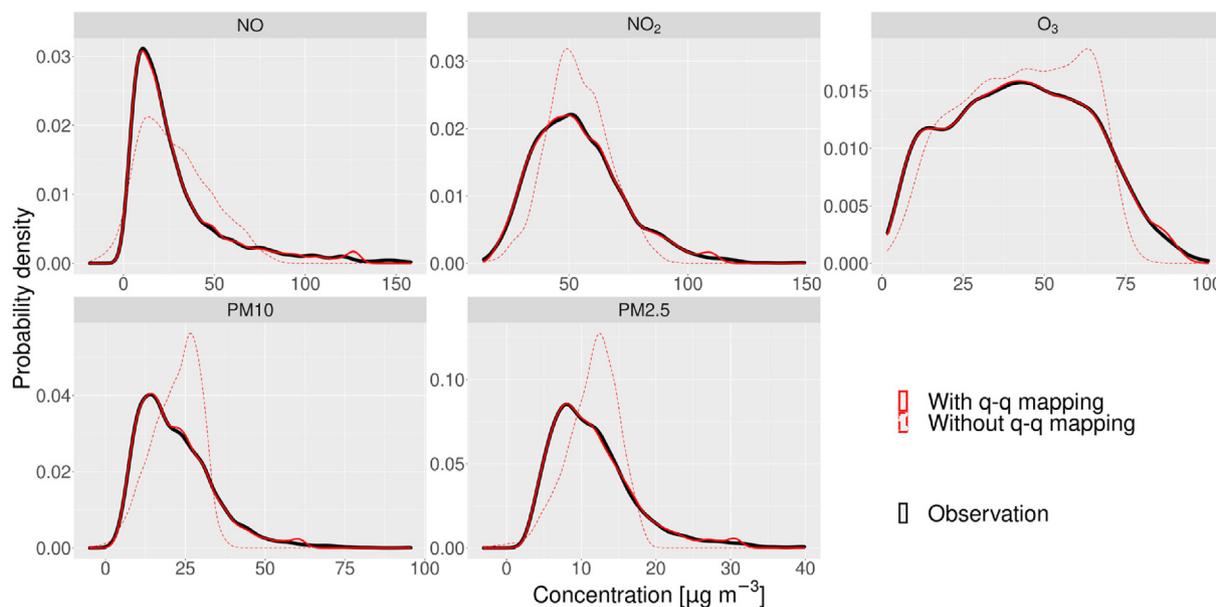


Fig. 3. The probability density function observed (black solid) and predicted without (red dashed) and with (red solid) q-q mapping post-correction for Escuelas Aguirre (Madrid) site using multiple linear regression technique with meteorological data of the 3 previous days. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

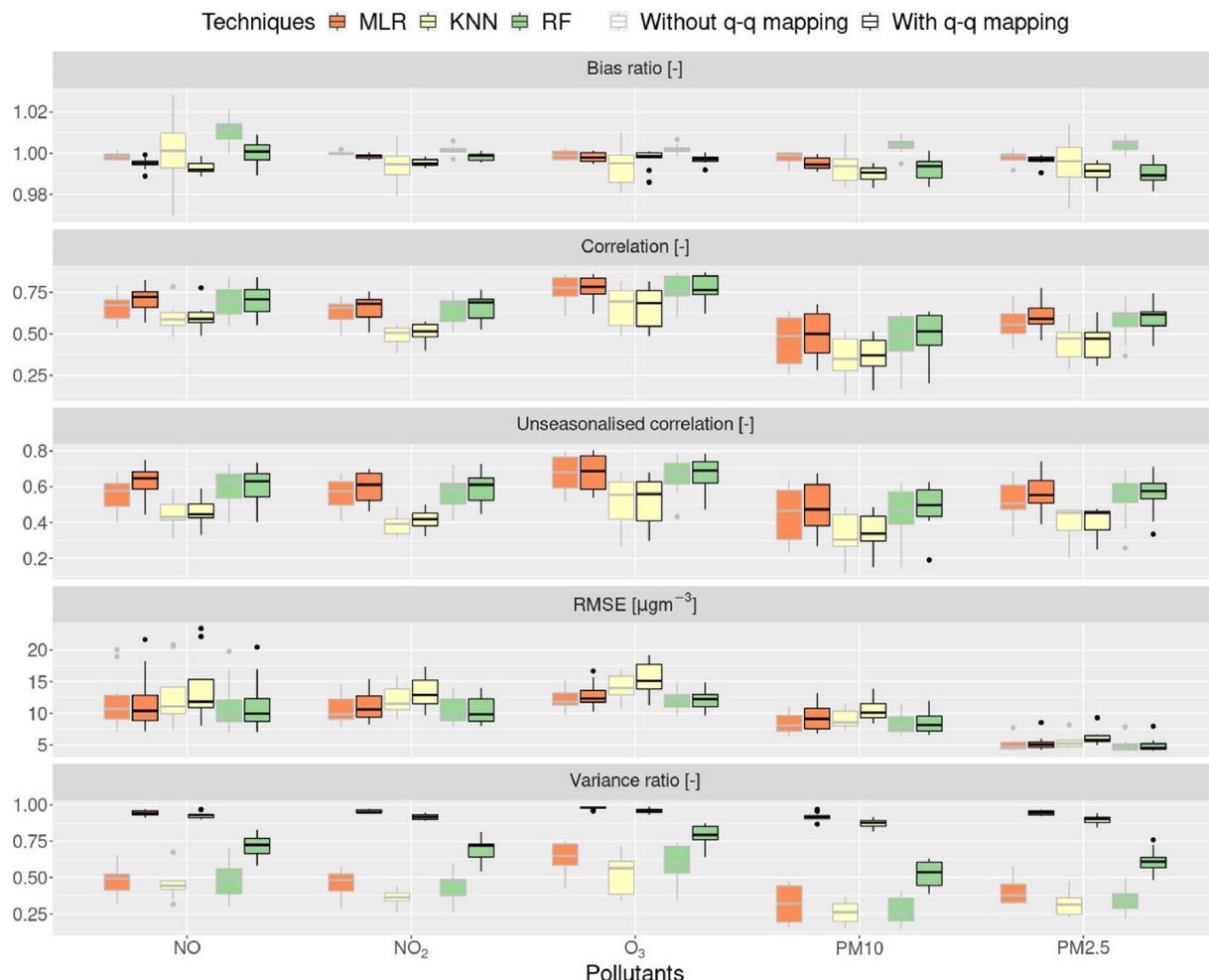


Fig. 4. Box plots comparing the cross-validated results obtained for the best model configuration found for each of the three data mining techniques considered without (grey-bordered) and with q-q mapping post-correction (black-bordered). Colours correspond to the Multiple Linear Regression model with 3 days of meteorological persistence (orange), the K-Nearest Neighbor model with $K = 10$ (yellow) and the Random Forest model with 100 trees (green). Each box plot contains the results obtained for the 10 sites shown in Table S.4 for each pollutant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to see NO concentrations return to BAU values for the normality period (-0.1% for the relative change). Since these reductions between RC and RC* in pandemic periods for NO and NO₂ were obtained considering

meteorological variability, it is reasonable to relate them to anomalous meteorological conditions during 2020 that may have reduced their observed concentrations.

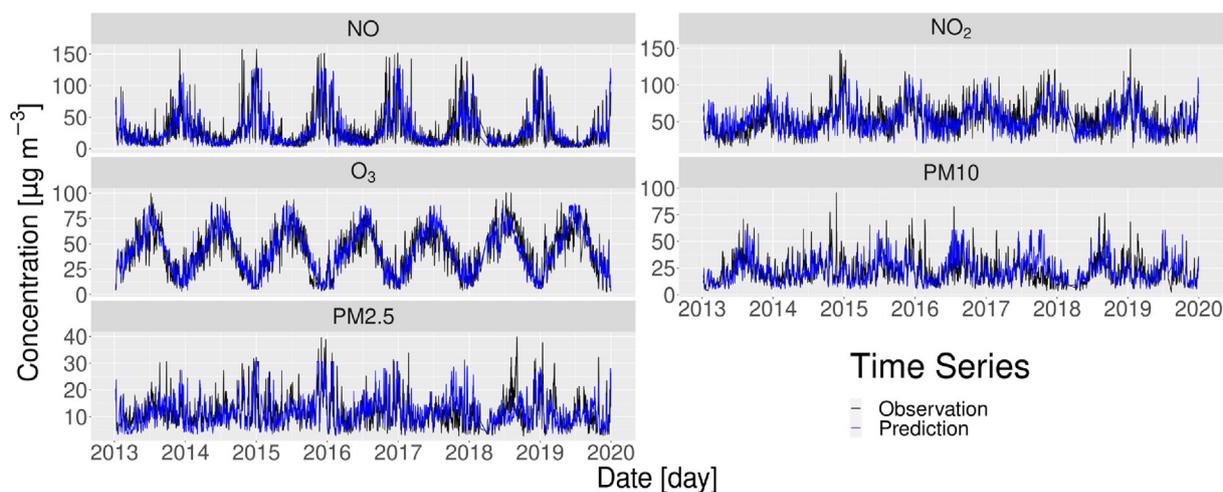


Fig. 5. Comparison between observed time series (black) of pollutant concentrations at the Escuelas Aguirre traffic monitoring site (Madrid) and those predicted by a multiple linear regression model with 3 days of meteorological persistence with q-q mapping (blue). Daily data from January 2013 to December 2019 are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Average relative percentage change RC* [%] and standard deviations of NO, NO₂, O₃, PM10 and PM2.5 concentrations in 2020 respect to predicted values C_{BAU-2020} at studied sites for each period.

Periods	NO [%]	NO ₂ [%]	O ₃ [%]	PM10 [%]	PM2.5 [%]
Pre-lockdown	-17.2 ± 13.1	-10.7 ± 13.6	15.3 ± 18.3	23.5 ± 20.4	10.5 ± 12.2
Lockdown	-54.7 ± 28.9	-51.3 ± 10.3	23.9 ± 15.5	-4.1 ± 12.8	-2.3 ± 16.2
De-escalation	-21.4 ± 31.5	-31.3 ± 13.5	1.0 ± 9.9	-1.2 ± 24.1	6.3 ± 20.2
Normality	-0.1 ± 28.6	-12.7 ± 13.7	1.1 ± 10.6	-5.5 ± 20.2	-2.7 ± 11.9
Second lockdown	-25.8 ± 32	-22.9 ± 14.6	23.5 ± 20.7	-0.4 ± 15.6	-10.1 ± 18.1

During the pre-lockdown period no lockdown restrictions were implemented. Therefore, BAU predicted air quality levels for an ideal model are expected to be equal to observed values (relative changes should move to zero). However, Fig. 2 and Table 3 show average RC* of -17.2, -10.7, 15.3, 23.5 and 10.5% for NO, NO₂, O₃, PM10 and PM2.5, respectively. The negative RC* observed for NO_x during this period can be considered as model overprediction, most likely caused by the lack of predictors with a high impact on these pollutant concentrations. Thus, the highest predicted concentrations of NO_x in this period can be explained by the lack of emissions rates from road traffic as predictors. These emissions rates may have been lower in the early 2020 because of the progressive implementation of urban mobility plans, such as Low Emission Zones (LEZ), with respect to the first years of the reference period (2013–2019). In addition, although total lockdown was declared on 14 March 2020 in Spain, the COVID-19 health crisis started much earlier with the first case in Spain in January 2020. While the months after the declaration of lockdown are undoubtedly more affected by the mobility restrictions derived from the COVID-19 pandemic, the pre-lockdown period cannot be isolated from the effects of COVID-19, with a decline in economic and social activity in Spain and especially in large cities such as Madrid or Barcelona. These decreases in NO_x concentrations prior to lockdown restrictions were not observed in Table 2 for RC, although they could be related to a decrease in traffic mobility. However, the estimated RC* relative changes obtained from the predicted values were able to distinguish these discrepancies that might have been masked by meteorological variability in the RC results.

With respect to the other pollutants in the pre-lockdown period, the model underpredicts their concentration levels (positive relative changes in Table 3). In the case of O₃, the RC* obtained in this period shift to positive values instead of the negative ones obtained for RC in Section 3.1, which is consistent with the negative values obtained for NO_x in the same period and the close mirror behaviour of both according to the NO_x titration cycle (Grange et al., 2021; Venter et al., 2020). This increase of RC* with respect to RC for O₃ is also observed for the lockdown period with an average RC* of 23.9%, 19.5% higher than that observed in Section 3.1 and consistent with the average 20% increase observed by Betancourt-Odio et al. (2021) in Madrid. This suggests that meteorological conditions were not favourable for O₃ production during these months. However, the difference between RC* for the pre-lockdown and lockdown periods was smaller than that obtained for RC (8.6% and 16.6% respectively) as was the case for NO_x.

Regarding PM, the underpredicted values during pre-lockdown can be explained by the lack of some predictors, such as intrusion of Saharan dust (Salvador Martínez, 2020) into the Spanish territory, which increases the PM10 concentration, or atmospheric stability-related variables, which may strongly affect PM10 and PM2.5 levels. Since the model was able to better correct this underprediction for PM2.5 than for PM10, the inclusion of Saharan dust intrusion as a predictor in future PM10 models is recommended.

The discrepancies obtained between observed and estimated values during the pre-lockdown were also observed by other scientific studies (Grange et al., 2021; Sicard et al., 2020). Grange et al. (2021) suggest to use this period as a validation phase to calculate the model offset in order to subtract this from the predictions, obtaining smaller reductions for periods affected by constraints.

Still, the relative changes in air pollutant concentrations shown in Fig. 2 and Table 3 based on estimated levels should be attributed more to COVID-

19 restrictions during the pandemic than those calculated from observed concentrations, because 2020 BAU predicted concentrations were obtained from the same meteorological conditions under which the pollutants were measured. However, more accurate models, including additional, more explicative predictors might improve the results of this approach.

4. Conclusions

Different data mining techniques (MLR, RF, KNN) developed to predict air pollutant concentrations (NO, NO₂, O₃, PM10 and PM2.5) at Spanish urban traffic sites, using meteorological variables as predictors, were compared from a leave-one-year-out cross-validation using 2013–2019 data. A quantile-quantile mapping post-correction was also applied, improving models' performance and the prediction of extreme pollution events. The results from this cross-validated experiment showed the best overall performance for all metrics studied for the MLR, as well as better results when the meteorological data of the previous 3 days were introduced as predictors. Using MLR models trained from historical data (2013–2019), we predicted the business-as-usual air quality levels in 2020 to calculate the relative changes in air pollutant concentrations only due to the COVID-19 restrictions.

The main outcomes revealed a pattern of decrease for NO and NO₂ in which abrupt reductions appear in the periods affected by the enclosure, with the largest reduction in the first lockdown (-54.7% for NO and -51.3% for NO₂) when more dramatic restrictions were implemented. These results obtained for NO_x are consistent with previous studies and support the hypothesis of relating its decrease to strong traffic reductions occurring due to mobility restrictions derived from the COVID-19 pandemic. The slight reduction in PM10 (-4.1%) and PM2.5 levels (-2.3%) obtained during the lockdown period might indicate a minor relationship with traffic sources. On the other hand, an increase in O₃ concentrations has been obtained with a near-mirror behaviour with NO_x pattern. This increase in O₃ concentrations has been also observed by previous studies and is consistent with changes in the emissions of its precursors.

The methodology implemented in this work was able to reveal differences between relative changes, masked by meteorological variability. Thus, the predicted BAU scenario allows to reduce the abrupt changes obtained for the observed relative change RC and to obtain reductions in NO_x concentrations prior to the lockdown restrictions, probably due to the implementations of urban mobility schemes or due to an early decrease in mobility due to COVID-19. Therefore, this methodology represents a simple but robust framework for exploratory analysis and intervention detection in air quality studies; in particular, it helps to interpret the air quality improvement associated with mobility restrictions resulting from the COVID-19 pandemic and the imperative need to implement urban mobility schemes such as LEZs or electrification of the vehicle fleet to decrease NO_x concentrations. However, results have also revealed that air quality interventions, such as those observed in 2020, can lead to an increase in ozone levels, which is responsible for respiratory infections. Therefore, the implementation of traffic restrictions should be followed by interventions on other O₃ precursors, such as VOCs.

Funding

This research was developed in the framework of the project "Contaminación atmosférica y COVID-19: ¿Qué podemos aprender de

- doi.org/10.1016/j.envpol.2020.116011 URL: <https://www.sciencedirect.com/science/article/pii/S0269749120367002>.
- Jiang, Z., Rashid, M.M., Johnson, F., Sharma, A., 2021. A wavelet-based tool to modulate variance in predictors: an application to predicting drought anomalies. *Environ. Modell. Softw.* 135, 104907. <https://doi.org/10.1016/j.envsoft.2020.104907>.
- Kowalski, P.A., Szwargryk, M., Kiełpńska, J., Konior, A., Kusy, M., 2021. Numerical analysis of factors, pace and intensity of the corona virus (COVID-19) epidemic in Poland. *Ecol. Inform.* 63, 101284. <https://doi.org/10.1016/j.ecoinf.2021.101284> URL: <https://www.sciencedirect.com/science/article/pii/S1574954121000753>.
- Kuhn, M., 2021. caret: Classification and Regression Training. URL: <https://CRAN.R-project.org/package=caret> R package version 6.0-88.
- Li, F., Gu, J., Xin, J., Schnelle-Kreis, J., Wang, Y., Liu, Z., Shen, R., Michalke, B., Abbaszade, G., Zimmermann, R., 2021. Characteristics of chemical profile, sources and PAH toxicity of PM_{2.5} in Beijing in autumn-winter transition season with regard to domestic heating, pollution control measures and meteorology. *Chemosphere* 276, 130143. <https://doi.org/10.1016/j.chemosphere.2021.130143> URL: <https://www.sciencedirect.com/science/article/pii/S0045653521006123>.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. *R News* 2, pp. 18–22 URL: <https://CRAN.R-project.org/doc/Rnews>.
- Lovrić, M., Pavlović, K., Vuković, M., Grange, S.K., Haberl, M., Kern, R., 2021. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. *Environ. Pollut.* 274, 115900. <https://doi.org/10.1016/j.envpol.2020.115900> URL: <https://www.sciencedirect.com/science/article/pii/S0269749120365891>.
- Manzanas, R., Lucero, A., Weisheimer, A., Gutiérrez, J.M., 2018. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? *Clim. Dyn.* 50, 1161–1176. <https://doi.org/10.1007/s00382-017-3668-z>.
- Maraun, D., 2013. Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *URLJ. Clim.* 26, 2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1> https://www.inet.es/pob_xls/pobmun.zip.
- Martorell-Marugán, J., Villatoro-García, J.A., García-Moreno, A., López-Domínguez, R., Requena, F., Merelo, J.J., Lacasaña, M., de Dios Luna, J., Díaz-Mochón, J.J., Lorente, J.A., Carmona-Sáez, P., 2021. Datac: a visual analytics platform to explore climate and air quality indicators associated with the covid-19 pandemic in Spain. *Sci. Total Environ.* 750, 141424. <https://doi.org/10.1016/j.scitotenv.2020.141424> URL: <https://www.sciencedirect.com/science/article/pii/S0048969720349536>.
- Mendez-Espinosa, J.F., Rojas, N.Y., Vargas, J., Pachón, J.E., Belalcázar, L.C., Ramírez, O., 2020. Air quality variations in northern South America during the COVID-19 lockdown. *Sci. Total Environ.* 749, 141621. <https://doi.org/10.1016/j.scitotenv.2020.141621> URL: <https://www.sciencedirect.com/science/article/pii/S0048969720351500>.
- Munnoli, P.M., Nabapure, S., Yeshavanth, G., 2020. Post-COVID-19 precautions based on lessons learned from past pandemics: a review. *J. Public Health* 1–9. <https://doi.org/10.1007/s10389-020-01371-3>.
- Muñoz-Sabater, J., 2019]. ERA5-Land Hourly Data From 1981 to Present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.e2161bac> (Accessed on 12-FEBRUARY-2021).
- Pérez, N., Pey, J., Querol, X., Alastuey, A., López, J., Viana, M., 2008. Partitioning of major and trace components in PM₁₀-PM_{2.5}-PM₁ at an urban site in Southern Europe. *Atmos. Environ.* 42, 1677–1691. <https://doi.org/10.1016/j.atmosenv.2007.11.034> URL: <https://www.sciencedirect.com/science/article/pii/S1352231007010874>.
- Petetin, H., Bowdalo, D., Soret, A., 2020a. Assessment of the Impact of the COVID-19 Lockdown on Air Pollution Over Spain Using Machine Learning. *Barcelona Supercomputing Center*, pp. 18–20 URL: <http://hdl.handle.net/2117/330993>.
- Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., Pérez García-Pando, C., 2020b. Meteorology-normalized impact of the covid-19 lockdown upon NO₂ pollution in Spain. *Atmos. Chem. Phys.* 20, 11119–11141. <https://doi.org/10.5194/acp-20-11119-2020> URL: <https://acp.copernicus.org/articles/20/11119/2020/>.
- R Core Team, 2021. R: A Language And Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL: <https://www.R-project.org/>.
- Rogula-Kozłowska, W., Klejnowski, K., Rogula-Kopiec, P., Mathews, B., Szopa, S., 2012. A study on the seasonal mass closure of ambient fine and coarse dusts in Zabrze, Poland. *Bull. Environ. Contam. Toxicol.* 88, 722–729. <https://doi.org/10.1007/s00128-012-0533-y>.
- Ropkins, K., Tate, J.E., 2021. Early observations on the impact of the COVID-19 lockdown on air quality trends across the UK. *Sci. Total Environ.* 754, 142374. <https://doi.org/10.1016/j.scitotenv.2020.142374> URL: <https://www.sciencedirect.com/science/article/pii/S0048969720359039>.
- Salvador Martínez, P., 2020. Enero de 2004. Ministerio para la Transición Ecológica y el Reto Demográfico. URL: https://www.miteco.gob.es/images/es/episodios_actualizados_hasta_el29dediciembre_de_2020_tcm30-520542.pdf Datos propiedad de la Dirección General de Calidad y Evaluación Ambiental del Ministerio para la Transición Ecológica y el Reto Demográfico, suministrados en el marco del "Encargo del Ministerio para la Transición Ecológica a la Agencia Estatal Consejo Superior de Investigaciones Científicas para la detección de episodios naturales de aportes transfronterizos de partículas y otras fuentes de contaminación de material particulado, y de formación de ozono troposférico".
- Samek, L., Stegowski, Z., Styszko, K., Furman, L., Fiedor, J., 2018. Seasonal contribution of assessed sources to submicron and fine particulate matter in a central European urban area. *Environ. Pollut.* 241, 406–411. <https://doi.org/10.1016/j.envpol.2018.05.082> URL: <https://www.sciencedirect.com/science/article/pii/S0269749118312041>.
- Sanchez-Lorenzo, A., Vaquero-Martínez, J., Calbó, J., Wild, M., Santurtún, A., Lopez-Bustins, J., Vaquero, J., Folini, D., Antón, M., 2021. Did anomalous atmospheric circulation favor the spread of COVID-19 in Europe? *Environ. Res.* 194, 110626. <https://doi.org/10.1016/j.envres.2020.110626> URL: <https://www.sciencedirect.com/science/article/pii/S0013935120315231>.
- Seinfeld, J.H., Pandis, S.N., 2016. *Atmospheric Chemistry And Physics: From Air Pollution to Climate Change*. John Wiley & Sons.
- Shi, Z., Song, C., Liu, B., Lu, G., Xu, J., Van Vu, T., Elliott, R.J.R., Li, W., Bloss, W.J., Harrison, R.M., 2021. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. URL: <https://advances.sciencemag.org/content/7/3/eabd6696Sci>. *Adv.* 7. <https://doi.org/10.1126/sciadv.abd6696> arXiv: <https://advances.sciencemag.org/content/7/3/eabd6696.full.pdf>.
- Sicard, P., De Marco, A., Agathokleous, E., Feng, Z., Xu, X., Paoletti, E., Diéguez Rodríguez, J.J., Calatayud, V., 2020. Amplified ozone pollution in cities during the COVID-19 lockdown. *Sci. Total Environ.* 735, 139542. <https://doi.org/10.1016/j.scitotenv.2020.139542> URL: <https://www.sciencedirect.com/science/article/pii/S004896972033059X>.
- Siciliano, B., Dantas, G., da Silva, C.M., Arbilla, G., 2020. Increased ozone levels during the COVID-19 lockdown: analysis for the city of Rio de Janeiro, Brazil. *Sci. Total Environ.* 737, 139765. <https://doi.org/10.1016/j.scitotenv.2020.139765> URL: <https://www.sciencedirect.com/science/article/pii/S1574954121000753>.
- Sillanpää, M., Hillamo, R., Saarikoski, S., Frey, A., Pennanen, A., Makkonen, U., Spolnik, Z., Van Grieken, R., Braniš, M., Brunekreef, B., Chalbot, M.C., Kuhlbusch, T., Sunyer, J., Kerminen, V.M., Kulmala, M., Salonen, R.O., 2006. Chemical composition and mass closure of particulate matter at six urban sites in Europe. *Atmos. Environ.* 40, 212–223. <https://doi.org/10.1016/j.atmosenv.2006.01.063> URL: <https://www.sciencedirect.com/science/article/pii/S0045653521006123>.
- Stratoulis, D., Nuthammachot, N., 2020. Air quality development during the covid-19 pandemic over a medium-sized urban area in Thailand. *Sci. Total Environ.* 746, 141320. <https://doi.org/10.1016/j.scitotenv.2020.141320> URL: <https://www.sciencedirect.com/science/article/pii/S004896972034849X>.
- Tobías, A., Carnerero, C., Reche, C., Massagué, J., Via, M., Minguillón, M.C., Alastuey, A., Querol, X., 2020. Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-COV-2 epidemic. *Sci. Total Environ.* 726, 138540. <https://doi.org/10.1016/j.scitotenv.2020.138540> URL: <https://www.sciencedirect.com/science/article/pii/S1352231007010874>.
- Van Rossum, G., 1995. Python Tutorial. Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI), Amsterdam.
- Venter, Z.S., Anun, K., Chowdhury, S., Lelieveld, J., 2020. COVID-19 lockdowns cause global air pollution declines. *URLProc. Natl. Acad. Sci.* 117, 18984–18990. <https://doi.org/10.1073/pnas.2006853117> Datos propiedad de la Dirección General de Calidad y Evaluación Ambiental del Ministerio para la Transición Ecológica y el Reto Demográfico, suministrados en el marco del "Encargo del Ministerio para la Transición Ecológica a la Agencia Estatal Consejo Superior de Investigaciones Científicas para la detección de episodios naturales de aportes transfronterizos de partículas y otras fuentes de contaminación de material particulado, y de formación de ozono troposférico" arXiv: <https://www.pnas.org/content/117/32/18984.full.pdf>.
- Von Luxburg, U., Schölkopf, B., 2011. Statistical learning theory: models, concepts, and results. In: Gabbay, D.M., Hartmann, S., Woods, J. (Eds.), *Inductive Logic*. North-Holland, volume 10 of *Handbook of the History of Logic*, pp. 651–706 <https://doi.org/10.1016/B978-0-444-52936-7.50016-1> URL: <https://www.sciencedirect.com/science/article/pii/B9780444529367500161>.
- Vu, T.V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., Harrison, R.M., 2019. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmos. Chem. Phys.* 19, 11303–11314. <https://doi.org/10.5194/acp-19-11303-2019> URL: <https://acp.copernicus.org/articles/19/11303/2019/>.
- Wong, P.Y., Hsu, C.Y., Wu, J.Y., Teo, T.A., Huang, J.W., Guo, H.R., Su, H.J., Wu, C.D., Spengler, J.D., 2021. Incorporating land-use regression into machine learning algorithms in estimating the spatial-temporal variation of carbon monoxide in Taiwan. *Environ. Modell. Softw.* 139, 104996. <https://doi.org/10.1016/j.envsoft.2021.104996> URL: <https://www.sciencedirect.com/science/article/pii/S1364815221000396>.
- Zhong, Q., Ma, J., Shen, G., Shen, H., Zhu, X., Yun, X., Meng, W., Cheng, H., Liu, J., Li, B., Wang, X., Zeng, E.Y., Guan, D., Tao, S., 2018. Distinguishing emission-associated ambient air PM_{2.5} concentrations and meteorological factor-induced fluctuations. *URLEnviron. Sci. Technol.* 52, 10416–10425. <https://doi.org/10.1021/acs.est.8b02685> doi: 10.1021/acs.est.8b02685, arXiv: doi:10.1021/acs.est.8b02685. PMID: 30118598.