Assessing the COVID-19 Impact on Air Quality: A Machine Learning Approach

Article *in* Geophysical Research Letters · February 2021

DOI: 10.1029/2020GL091202

2 authors:

Yves Philippe Rybarczyk
Dalarna University
101 PUBLICATIONS **852** CITATIONS

SEE PROFILE

Rasa Zalakeviciute
Universidad de Las Américas
51 PUBLICATIONS **823** CITATIONS

SEE PROFILE

AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL091202

Special Section:
The COVID-19 pandemic: linking health, society and environment

Assessing the COVID-19 Impact on Air Quality: A Machine Learning Approach

Yves Rybarczyk[1,2] and **Rasa Zalakeviciute**[2,3]

[1]The Department of Data and Information Sciences, Dalarna University, Falun, Sweden, [2]SI2 Lab, Universidad de Las Americas, Quito, Ecuador, [3]Grupo de Biodiversidad Medio Ambiente y Salud (BIOMAS), Universidad de Las Americas, Quito, Ecuador

Abstract The worldwide research initiatives on Corona Virus disease 2019 lockdown effect on air quality agree on pollution reduction, but the most reliable method to pollution reduction quantification is still in debate. In this paper, machine learning models based on a Gradient Boosting Machine algorithm are built to assess the outbreak impact on air quality in Quito, Ecuador. First, the precision of the prediction was evaluated by cross-validation on the four years prelockdown, showing a high accuracy to estimate the real pollution levels. Then, the changes in pollution are quantified. During the full lockdown, air pollution decreased by −53 ± 2%, −45 ± 11%, −30 ± 13%, and −15 ± 9% for NO_2, SO_2, CO, and $PM_{2.5}$, respectively. The traffic-busy districts were the most impacted areas of the city. After the transition to the partial relaxation, the concentrations have nearly returned to the levels as before the pandemic. The quantification of pollution drop is supported by an assessment of the prediction confidence.

Key Points:

- A data driven modeling is applied to quantify the reduction of air pollution during the Corona Virus disease 2019 outbreak in Quito, Ecuador
- The accuracy of the models is high (mean PCC = 0.78), especially for predicting NO_2 (mean PCC = 0.87) and CO (mean PCC = 0.86)
- The average drop of pollution during the lockdown is: −53% for NO_2, −45% for SO_2, −30% for CO, and −15% for $PM_{2.5}$
- The industrial areas are less impacted by the quarantine than the traffic and residential districts
- The concentration of pollution tends to return to usual levels, as soon as the relaxed restriction is implemented

Correspondence to:
R. Zalakeviciute and Y. Rybarczyk, rasa.zalake@gmail.com; yry@du.se

Citation:
Rybarczyk, Y., & Zalakeviciute, R. (2021). Assessing the COVID-19 impact on air quality: A machine learning approach. *Geophysical Research Letters*, *48*, e2020GL091202. https://doi.org/10.1029/2020GL091202

Received 20 OCT 2020
Accepted 11 DEC 2020

1. Introduction

Throughout 2020, human population of the world faced a challenging new pandemic of Corona Virus disease 2019 (COVID-19), propagated by a severe acute respiratory syndrome corona virus 2 (SARS-CoV-2). Originating in China at the end of 2019, this complex-symptom disease exponentially spread to almost every part of the world (Worldometer, 2020). In order to slow down social interaction, the principal mode of infection, a range of regulatory measures has been implemented across the countries. Anthropogenic activities were limited or even prohibited, which in turn, positively affected the environmental quality, especially, due to the reduction in emissions by transportation and industries (Shi & Brasseur, 2020; Y. Wang et al., 2020). This was a unique global experiment, as the urban, regional and global air quality conditions have been persistently worsening due to a rapid increase in polluting sources by human population (Limb, 2016; UN., 2015; WHO., 2016). Most common atmospheric contaminants, such as particulate matter (PM, especially with aerodynamic diameters ≤2.5 μm, $PM_{2.5}$), nitrogen oxides (NO and NO_2), sulfur dioxide (SO_2), and carbon monoxide (CO), compromise respiratory and cardiovascular systems (EPA, 2018; Lelieveld et al., 2015; Pope et al., 2006). Moreover, a number of studies show that population exposed to a long-term elevated levels of air pollution have higher mortality rates due to COVID-19 (Coker et al., 2020; Geographic, 2020; Ogen, 2020; Wu et al., 2020).

Undoubtedly, nearly globally implemented social distancing and isolation measures have resulted in a significant reduction of air pollution (Bauwens et al., 2020; Ding et al., 2020; European Space Agency, 2020; NASA, 2020; Navinya et al., 2020; Nikkei Asian Review, 2020; Querol et al., 2007; Shi & Brasseur, 2020; P. Wang et al., 2020). These statistics of air quality improvements are commonly reported by comparing the restricted human activity periods with the "normal conditions," months or years before the sanitary emergency with no way to estimate the reliability of these results.

Ecuador is among the countries that implemented extreme quarantine measures. The country had a first reported case of COVID-19 on February 29, 2020, very quickly escalating in numbers, and in two weeks reporting the first deaths (Worldometer, 2020). In order to prevent health care system collapse, extremely strict national preventative rules were communicated on March 15, active from March 17, 2020. The full lockdown only permitted circulation for absolute necessities (e.g., medicine and food) once per week, based on the license plate number, 7:00–14:00 during the workdays (Zalakeviciute et al., 2020a). Gradually, however, more businesses could function every day anytime. While the infection numbers kept growing, the

relaxation to these rules was implemented starting June 2, 2020. Partial relaxation allowed private vehicle circulation every second day, excluding Sundays. This rule permitted all bus functioning (public and private lines) at a limited capacity for passengers. A further increase in permits to some offices were issued (shopping malls, governmental offices, private businesses, etc.).

High elevation capital Quito is characterized by the rainy and dry seasons, with a strong influence of El Niño phenomenon, affecting air pollution conditions, usually worsening during the rainier months (Zalakeviciute et al., 2018). In addition, due to technological advances and increasing strictness of traffic regulations, air quality tends to improve over time, regardless of the motorization trend (European Environment Agency, 2018, 2017; Zalakeviciute et al., 2018). As a result, the evaluation of urban pollution reduction during the sanitary quarantine based on pollution levels from previous years or months could introduce some uncertainties. To get a more accurate assessment of the effect of an event on air quality, a variable normalization technique can be applied (Grange et al., 2018; Grange & Carslaw, 2019). A meteorology-normalized approach was used to analyze the effect of the COVID-19 lockdown on NO_2 concentrations in Spain (Petetin et al., 2020) and in 20 North American cities (Goldberg et al., 2020).

This paper proposes an alternative approach, which consists of training a machine learning (ML) algorithm that will learn the effects of the climatic conditions and time on the atmospheric pollution. ML approach is preferred over chemical transport models, as the latter show a reduced performance in the complex terrain regions (Pani et al., 2020; Žabkar et al., 2015) and require an updated emissions inventory, which Quito lacks. Unlike the previous studies, NO_2, SO_2, CO, and $PM_{2.5}$ are analyzed, in the Ecuadorian capital. Since the lockdown affected essentially the human mobility, we investigate the differences in the reduction of pollution in different districts with varied sources of contamination (e.g., traffic vs. industry). One model for each city area and contaminant is built, in order to estimate the concentration of pollutants from meteorological and temporal features without the lockdown. The quantification of the concentration changes due to the reduced human activity, is obtained by comparing the value provided by the model to the real measurements.

2. Materials and Methods

2.1. Study Site

Ecuador, even though one of the smallest countries in South America, occupies a variety of biomes, including Pacific Coast region, Galapagos Islands Archipelago, Andes Cordillera and the Amazonian Rainforest. The capital Quito is set on a slender plateau, stretching on the side of the Pichincha volcano (elev. 4,800 meters above sea level (m.a.s.l.)), forming a part of the Andes Cordillera. Quito is the most elevated constitutional capital of the world (elev. 2,815 m. a.s.l.) (EMASEO., 2011). The quickly spreading and densifying metro area houses a population of 2.2 million people (INEC., 2011). As most of the cities in the developing countries, Quito reports a long-term air quality problem (Zalakeviciute et al., 2018).

The Secretariat of Environment of Metropolitan District of Quito manages an air quality network recording atmospheric pollutants and meteorological parameters, since 2004. Six sites with more complete data have been chosen for this study: urban traffic-busy site Belisario (elev. 2,835 m. a.s.l, coord. 78°29′24″W, 0°10′48″S), urban/industrial sites Camal (elev. 2,840 m. a.s.l, coord. 78°30′36″W, 0°15′00″S), and Carapungo (elev. 2,660 m. a.s.l, coord. 78°26′50″W, 0°5′54″S), suburban/residential traffic site Cotocollao (elev. 2,739 m. a.s.l, coord. 78°29′50″W, 0°6′28″S), suburban/agricultural site Guamani (elev. 3,066 m. a.s.l., coord. 78°33′5″W, 0°19′51″S), and suburban/industrial site Chillos (elev. 2,453 m. a.s.l, coord. 78°27′36″W, 0°18′00″S). More details on the study sites can be found in a previous study evaluating pollution reduction in Ecuador during the first month of COVID-19 quarantine (Zalakeviciute et al., 2020a).

2.2. Quito Air Quality Monitoring

The atmospheric pollution monitoring equipment were set up, following the procedures established by the Environmental Protection Agency of the United States (USEPA). CO concentrations were measured using ThermoFisher Scientific 48i (EPA No. RFCA-0981-054). SO_2 was measured using ThermoFisher Scientific 43i high level SO_2 analyzers (EPA No. EQSA-0486-060). NO_2 was measured using ThermoFisher Scientific

42i NOx analyzers. $PM_{2.5}$ was measured using Thermo Scientific FH62C14-DHS continuous ambient particulate monitors 5014i (EPA No. EQPM-0609-183).

To register meteorological parameters, complete automatic weather stations were collocated with the atmospheric pollution equipment at study sites. MetOne instruments were used to record wind speed and direction. Thies Clima were employed to measure relative humidity, temperature, and precipitation. Kipp & Zonen radiometers were used to measure solar radiation. Vaisala equipment were used to measure atmospheric pressure.

2.3. Machine Learning Modeling

ML models are trained to predict the pollutant concentrations from meteorological and time variables. Seven meteorological features are selected: relative humidity, precipitation, temperature, solar radiation, pressure, wind speed, and wind direction. The time attributes are: Julian day, week day, hour, and trend (date index). These temporal variables account for cyclical emission patterns, which are robust drivers of the concentration of several air pollutants (Henneman et al., 2015). Julian day, weekday, and hours inform respectively on the seasonal (wet vs. dry season), weekly (workday vs. weekend), and hourly (rush hours) occurrences of pollution peaks. Trend reflects the pollution reduction in Quito due to the implementation of fuel and traffic regulations (Zalakeviciute et al., 2018). The features to predict are the concentrations of NO_2, SO_2, CO, and $PM_{2.5}$ in the six different districts. We work on hourly data from January 1, 2016 to June 30, 2020. The data set is split in three parts. Data from January 1, 2016 to January 15, 2020 are used to train the model, based on a five-fold cross-validation. Then, two months before the full lockdown (January 16, to March 15, 2020) are used to test the prediction accuracy. Even though the COVID-19 quarantine measures were officially applied starting March 17, the city was showing the lockdown signs starting the March 15, the day of national regulation communication. The Root Mean Squared Error (RMSE) and the Pearson Coefficient of Correlation (PCC) are the metrics chosen to assess the model performance. To validate a model, the RMSE must be minimal, and the PCC the closest to 1. Finally, the remaining data consist of the values during the full (March 16 to June 1, 2020) and relaxed (June 2 to June 30, 2020) lockdown. The modification of the concentration of contaminants is quantified by subtracting the model predicted values from the actual values recorded during the quarantine.

Due to the complex nonlinear relationships between meteorology and air quality, we selected a nonparametric ML method. A Gradient Boosting Machine (GBM) was chosen, because it is a state-of-the-art decision tree-based ensemble method (Friedman, 2001). GBM builds an ensemble of shallow and weak successive trees with each tree learning and improving the previous one. As any boosting method, this algorithm adds sequentially new models to the ensemble. At each iteration, a new weak base-learner model is trained from the error of the entire learned ensemble. A model is considered weak if its error rate is just slightly better than random guessing. In boosting, a weak model is built at each iterative sequence, in order to improve the remaining errors. The generalization of the algorithm is a stagewise additive model of n individual regression trees Equation 1.

$$f(x) = \sum_{n=1}^{N} f^n(x) \tag{1}$$

GBM offers several advantages over the other machine learning algorithms. First, the data set does not need any specific preprocessing. Second, the algorithm handles missing values, and does not require data imputation. Third, since it is based on a decision tree method, feature selection is automatically embedded in the algorithm. Finally, it is possible to identify the importance of each variable in the resulting model.

3. Data

Four years and six months of atmospheric pollution and meteorological data were processed on the period prior and during the COVID-19 quarantine. For each contaminant, the concentration change is computed from the mean difference between ML-based business-as-usual and actual value measured during the outbreak. The H_2O-R package was used to implement the GBM algorithm in RStudio. We tuned the parameters of the library so the training (i.e., incrementation of the number of trees) stops after no improvement on

Table 1
Comparison of the predictive accuracy for the machine learning-based models (eight column) versus the four previous year average (ninth column).

District	Type	Pollutant	Machine learning model					Mean 2016-2019
			Number of trees	RMSE (train)	PCC (train)	RMSE (test)	PCC (test)	PCC
Belisario	Urban - traffic	NO2	8009	5.1	0.91	6.2	0.86	0.07
		SO2	7895	1.7	0.84	2.2	0.7	0.27
		CO	6644	0.1	0.93	0.2	0.88	0.09
		PM2.5	6585	5.2	0.85	6.5	0.75	0.22
Carapungo	Urban - industrial	NO2	6741	4.7	0.92	5.9	0.87	0.03
		SO2	6926	1.4	0.89	2.0	0.76	0.08
		CO	6803	0.1	0.92	0.2	0.85	0.29
		PM2.5	9852	7.2	0.83	9.1	0.72	0.04
Camal	Urban - industrial	NO2	8030	5.4	0.91	7.0	0.85	0.08
		SO2	10033	2.8	0.91	4.2	0.79	0.18
		CO	8222	0.2	0.93	0.2	0.86	0.13
		PM2.5	8725	9.7	0.76	12.1	0.59	0.08
Cotocollao	Suburban - traffic	NO2	12862	2.2	0.97	3.8	0.92	0.1
		SO2	13368	0.6	0.96	1.2	0.85	0.12
		CO	12253	0.1	0.97	0.2	0.88	0.06
		PM2.5	9291	5.6	0.88	9.5	0.61	0.18
Guamani	Suburban - agricultural	NO2	9483	4.3	0.94	5.8	0.88	0.23
		SO2	11099	1.1	0.85	1.7	0.6	0.25
		CO	9594	0.1	0.93	0.2	0.84	0.13
		PM2.5	9040	8.3	0.79	10.8	0.6	0.17
Chillos	Suburban - industrial	NO2	9202	4.5	0.9	5.9	0.82	0.15
		SO2	6494	4.7	0.84	6.5	0.66	0.07
		CO	7908	0.1	0.92	0.1	0.87	0.31
		PM2.5	8443	5.6	0.78	6.9	0.63	0.01

the cross-validated error is provided by 10 consecutive trees. The accuracy of the GBM models is compared to a benchmark, which consists of calculating the average concentration of each pollutant at each site from January 16 to June 30, during 2016–2019. This period is chosen, because the ML models are built on the same years. For illustrative purposes, hourly NO_2, SO_2, CO, and $PM_{2.5}$ concentrations were compiled as 24-h averages starting January 2020. Graphics were done using Microsoft Excel (MS Office) and Igor Pro (Wavemetrics) softwares. The data are available at https://data.mendeley.com/datasets/trs5j932s8/1 (Zalakeviciute, et al., 2020b).

4. Results and Discussion

4.1. Models and Performance

Twenty-four models (six sites × four contaminants) were built to predict the level of pollution without lockdown (business-as-usual). Table 1 indicates the number of trees used to implement this decision tree-based ensemble method. More than 6,000 trees were necessary to build the models. The smallest and highest number of trees were in Belisario (mean = 7,283) and Cotocollao (mean = 11,943), respectively.

The metrics used to evaluate the performance of the models (mean RMSE and mean PCC) show a high prediction accuracy. Also, the fact that the performance is systematically lower on the testing set (RMSE = 4.5; PCC = 0.78) than the training set (RMSE = 3.4; PCC = 0.89) suggests no overfitting. The best prediction on the testing set is obtained for NO_2 (RMSE = 5.8; PCC = 0.87) and CO (RMSE = 0.2; PCC = 0.86). The model accuracy for SO_2 is good (RMSE = 3; PCC = 0.73) and fair for $PM_{2.5}$ (RMSE = 9.1; PCC = 0.65). For these last two pollutants, there are more accuracy differences from one district to another (stddev PCC SO_2 = 0.08; stddev PCC $PM_{2.5}$ = 0.06) than for the best predicted gases (stddev PCC NO_2 = 0.03; stddev PCC CO = 0.01).

These results show that the reliability of the proposed ML-based models to predict the concentration of pollutants is higher for NO_2 and CO than SO_2 and $PM_{2.5}$. The prediction of the first three contaminants is better than $PM_{2.5}$. Those are contaminants strictly related to the fossil fuel combustion, and in an urban area they are directly resulting from human activity. On the contrary, $PM_{2.5}$ while also mostly emitted by mobile sources, it may come from other sources (Karagulian et al., 2015), settle on the surfaces and disperse again in the atmosphere due to wind ventilation.

On the other hand, a benchmark based on the averages of four previous years shows a very low accuracy. Its correlation with the values observed from mid-January 2020 to mid-March 2020 does not exceed 0.3 (PCC $NO_2 = 0.11$; PCC CO = 0.17; PCC $SO_2 = 0.16$; PCC $PM_{2.5} = 0.12$). This result suggests that applying a simple average of the concentration measured in 2016–2019, for the lockdown period, is not reliable to quantify the impact on air quality. This outcome highlights the benefit of building a model based on an ML approach, particularly because it considers actual meteorological conditions and temporal features.

It is to note that the model accuracy is also dependent on the characteristics of the metropolitan areas. The best prediction is obtained for the districts in which the source of pollution is traffic-based (PCC for Belisario = 0.8; men PCC for Cotocollao = 0.82). The performance is slightly worse for the industrial (PCC for Carapungo = 0.8; PCC for Camal = 0.77; PCC for Chillos = 0.75) and agricultural areas (PCC for Guamin = 0.73). Whether the district is urban (PCC = 0.79) or suburban (PCC = 0.77) does not influence significantly the quality of the prediction. This is because both areas are circulated by the same poor-quality fuel-powered buses and private vehicles. These results can be explained by the fact that the main source of pollution in Quito is traffic-based, especially peaking during rush hours (Zalakeviciute et al., 2018). These precise patterns can be learned by our data-driven models. Also, a possible distinction between inner-city and suburb is not verified, because the COVID-19 quarantine regulations are applied extremely strictly in all the metropolitan area.

4.2. Variable Importance

GBM algorithm allows us to understand how the features have been handled by the resulting models. The importance of the 11 independent variables for the six districts and four pollutants is represented in Figure 1. For a determined contaminant and site, this importance is expressed in percentage of each feature. If 15% is used as minimal threshold, the five more relevant factors to build the models are: trend, hour, wind speed (WS), relative humidity (RH), and temperature (Temp). The other six features do not exceed this cutoff whatever the district and the pollutant. The two temporal features, trend and hour, are always among the most important variables for the modeling of any contaminant. For the trend feature, it can be explained by the implementation of the successive regulations, which have progressively improved air quality in Quito over the last decade (Zalakeviciute et al., 2018). As previously mentioned, the daily pollution peaks occur during the morning and evening rush hours systematically making hour a reliable feature to predict the concentration of anthropogenic pollutants.

We note that the three most relevant meteorological features (WS, RH, and Temp) have, on average, a lower percentage of importance than the two temporal factors (Trend and Hour). WS is mostly used for the prediction of $PM_{2.5}$ and CO. The transport and re-suspension effect of wind on the traffic-related contamination can explain the importance of this factor in the models. Similarly, RH seems to influence mainly the concentration of these two pollutants. A previous study by the authors shows that wet conditions increase traffic-related pollution, because humidity reduces engine efficiency, especially in high-elevation cities (Zalakeviciute et al., 2018). Among the three relevant meteorological factors, temperature is the least important. Nevertheless, it is significantly used for the modeling of NO_2 in Carapungo and SO_2 in Camal, which are both urban-industrial areas. Temperature, which is positively correlated with WS and negatively correlated with RH, is known for causing thermal turbulences that contribute to dilute gas emissions.

In terms of districts, some unique percentages of importance emerge. For Cotocollao, trend constitutes half the features for predicting the NO_2 concentration. It means that the consecutive fuel and traffic regulations have largely impacted this area. On the contrary, the level of NO_2 and $PM_{2.5}$ in the suburban agricultural site of Guamani are more influenced by wind velocity, suggesting transport factor. Camal, as an urban-industrial

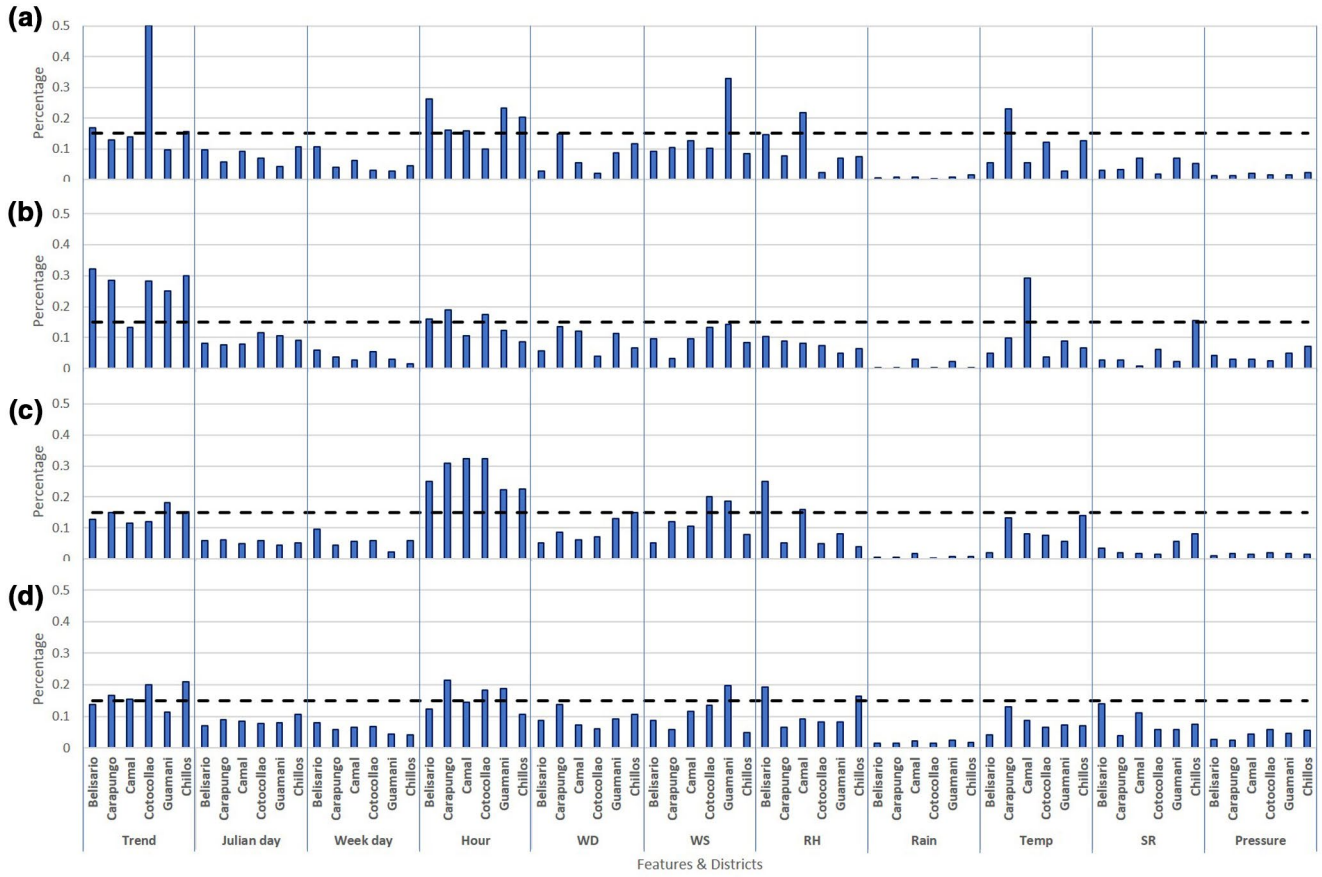

Figure 1. Variable importance for the modeling of NO_2 (panel (a)), SO_2 (panel (b)), CO (panel (c)), and $PM_{2.5}$ (panel (d)) for the six districts. The broken line is an arbitrary threshold (15% of importance) used to highlight the most relevant features for the prediction of each contaminant. WD, WS, RH, Temp and SR stand for wind direction, wind speed, relative humidity, temperature and solar radiation, respectively.

area, is the only district which exhibits concentration levels of SO_2 more affected by temperature than the trend feature, also suggesting increased temperature effect on pollution transport. Considering the dense road network in Belisario, it was expected to get RH with high percentage of importance for modeling $PM_{2.5}$ and CO in this urban area. For CO, directly linked to combustion sources, hour is systematically the most important feature independently of the district.

4.3. Effect of the Mobility Restrictions

Figure 2 represents the concentration of the four pollutants before (green-shaded area), during (red-shaded area) and after the full lockdown (yellow-shaded area), for two types of districts (traffic vs. industrial). A clear drop of concentrations is observed from the beginning of the lockdown (March 16, 2020), especially for NO_2, SO_2, and CO (Figure 2, panels a–f). The measurements are a bit noisier for $PM_{2.5}$, and the reduction is not so obvious in both, traffic (Belisario) and industrial (Carapungo) areas (Figure 2, panels g–h). For instance, high $PM_{2.5}$ concentrations can be noticed around 20 March. This peak seems to be caused by windy conditions, possibly re-suspending fine particulate matter into the atmosphere (Csavina et al., 2014). The contrasted and limited effect of the lockdown on the concentration of $PM_{2.5}$ is also observed in other cities around the world due to regional pollution events or an increase in private vehicles' use (Dhaka et al., 2020; Faridi et al., 2020). Another explanation is the fact that PM are not only emitted by vehicles, but also factories which continued to work almost normally during the quarantine. The relatively low reduction of $PM_{2.5}$ in the industrial districts of Quito (−20% for Camal, −19% for Carapungo, and +3% for Chillos) supports

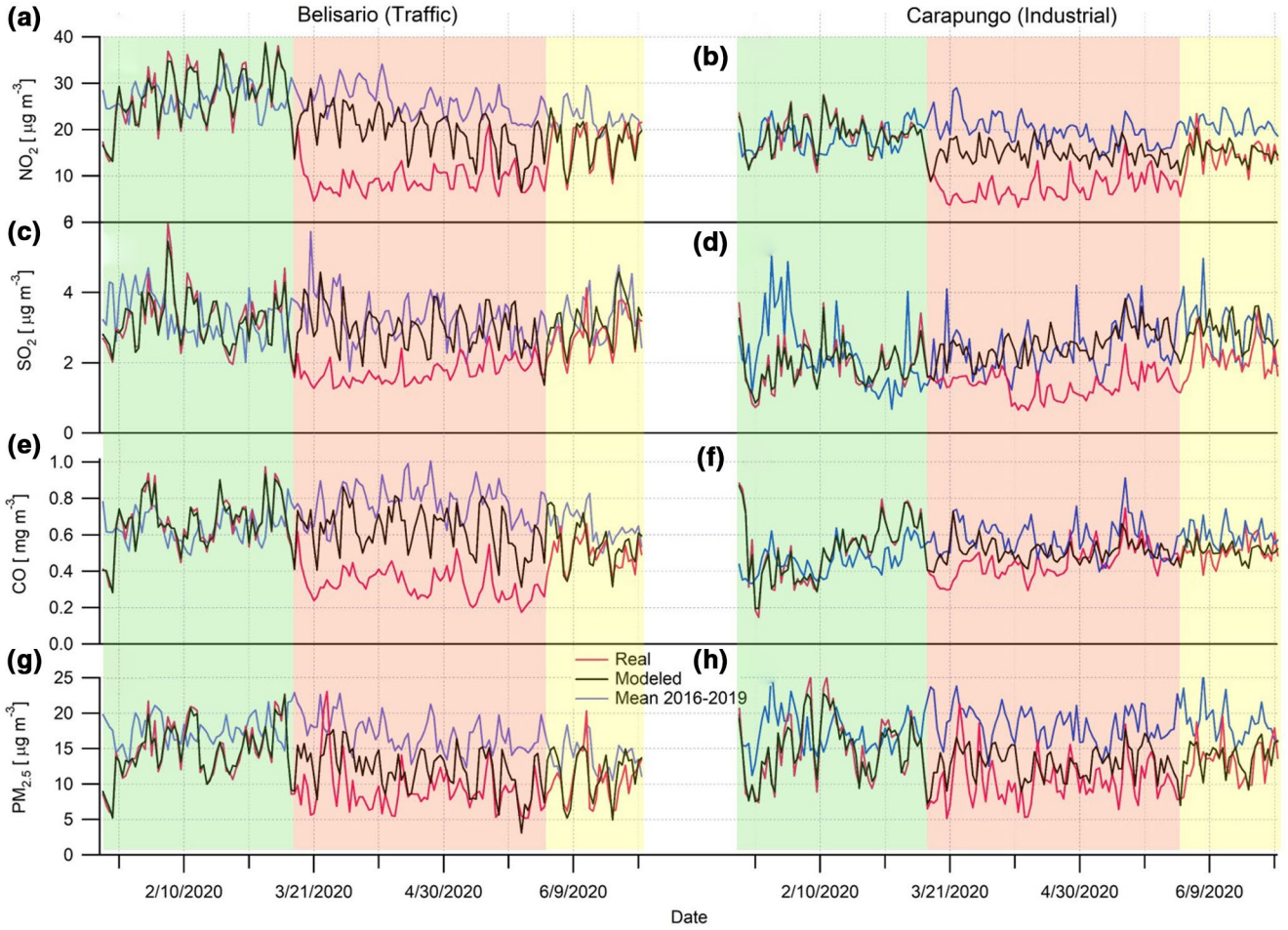

Figure 2. Real and estimated (ML modeled and four previous year mean) concentration of pollutants in traffic (Belisario) and industrial (Carapungo) areas during the prepandemic (green-shaded area), full-lockdown (red-shaded area), and partial relaxation (yellow-shaded area). For Belisario, panels (a), (c), (e) and (g) represent the concentrations of NO_2, SO_2, CO and $PM_{2.5}$, respectively. For Carapungo, NO_2, SO_2, CO and $PM_{2.5}$ are plotted in panels (b), (d), (f), and (h), respectively. ML: Machine Learning.

this hypothesis (Table 2). Once the full lockdown is over (June 1, 2020), the concentrations return progressively to close to the same levels as before the pandemic.

Moreover, Figure 2 shows the accuracy of the estimation of the pollution levels as if no lockdown was implemented. Before March 16, the overlap between predictive data, from the ML models, and actual data is almost perfect. It suggests that we can reasonably trust these predictive values as very close to the ones that would have been recorded if the human activities were usual. On the contrary, a reference based on the average of four previous years does not correlate at all with the real data. This visualization confirms that the proposed GBM-based models can provide an accurate benchmark to quantify the pollution drop during the lockdown. The estimation of pollution reduction provided by the ML approach is by far more reliable than a simple statistical averaging of the value recorded during the past years. The robustness of the GBM models is verified by the fact that actual and predictive data re-overlap progressively after the implementation of the partial relaxation measures.

Summarizing, during the full lockdown, NO_2 gets the highest drop, reducing by half its concentrations (overall mean: −53% ± 2%), independently of the district (Table 2a). SO_2 is also strongly reduced over the whole city (overall mean: −45% ± 11%), followed by CO (−30% ± 13%). CO decreases more in the traffic and residential areas (around 35%) than in the industrial areas (around 20%), which can be explained by the fact that CO is directly emitted from vehicles. As previously mentioned, the effect of the lockdown on $PM_{2.5}$

Table 2
Effect of the two levels of mobility restriction on the concentration of pollutants: (a) full-lockdown, and (b) partial relaxation.

a) Full lockdown

Type	Districts	$PM_{2.5}$	CO	NO_2	SO_2	Mean	SD
traffic	Belisario	-24%	-47%	-53%	-41%	-41%	±11%
traffic/residential	Cotocollao	-18%	-25%	-55%	-28%	-32%	±14%
residential	Guamani	-12%	-38%	-50%	-66%	-42%	±20%
industrial/residential	Camal	-20%	-40%	-54%	-45%	-40%	±12%
industrial	Carapungo	-19%	-11%	-50%	-46%	-32%	±17%
industrial	Chillos	3%	-16%	-54%	-46%	-28%	±23%

Mean ±SD traffic	-21%	±3%	-36%	±11%	-54%	±1%	-35%	±6%	-36%	±12%
Mean ±SD residential	-17%	±3%	-34%	±7%	-53%	±2%	-46%	±16%	-38%	±14%
Mean ±SD industrial	-12%	±11%	-22%	±13%	-53%	±2%	-46%	±1%	-33%	±17%
Mean ±SD overall	-15%	±9%	-30%	±13%	-53%	±2%	-45%	±11%	-36%	±15%

b) Partial relaxation

Type	Districts	$PM_{2.5}$	CO	NO_2	SO_2	Mean	SD
traffic	Belisario	-9%	-8%	-1%	-11%	-7%	±4%
traffic/residential	Cotocollao	-7%	-22%	-40%	-19%	-22%	±12%
residential	Guamani	11%	-10%	4%	-41%	-9%	±20%
industrial/residential	Camal	-4%	-21%	-2%	-13%	-10%	±8%
industrial	Carapungo	-4%	0%	-2%	-24%	-8%	±10%
industrial	Chillos	0%	-6%	-11%	-26%	-11%	±10%

Mean ±SD traffic	-8%	±1%	-15%	±7%	-21%	±19%	-15%	±4%	-15%	±4%
Mean ±SD residential	0%	±8%	-18%	±5%	-13%	±19%	-24%	±12%	-14%	±9%
Mean ±SD industrial	-3%	±2%	-9%	±9%	-5%	±4%	-21%	±6%	-9%	±7%
Mean ±SD overall	-2%	±7%	-11%	±8%	-9%	±15%	-22%	±10%	-11%	±7%

Note: Color scale from red to green highlights the percentage of pollution reduction. The greener is the cell, the higher is the pollution drop. SD stands for standard deviation.

is less important, but more significant for traffic-busy than industrial districts. For instance, the level of $PM_{2.5}$ kept constant in Chillos, but more reduced in traffic-busy areas, because this anthropogenic pollutant is mostly emitted by diesel vehicles. In the case of Chillos, it comes from thermoelectric power plant and factories, which maintained their activity during the pandemic.

When the restrictions on human mobility subsided slightly, the atmospheric pollution returned to almost normal levels (overall mean: −11% ± 7%, during the partial relaxation, against −36% ± 15%, during the full lockdown), especially for $PM_{2.5}$, CO, and NO_2 (Table 2b). Even if the reduction was not comparable with the lockdown period, SO_2 levels stayed lower than usual. There are also certain inconsistencies from one site to

another (e.g., −40% NO_2 in Cotocollao, against +4% in Guamani) and from one pollutant to another (e.g., +11% $PM_{2.5}$, against −41% SO_2 in Guamani). The consequences on air quality are a bit noisier during the partial relaxation due to a more stochastic human activity.

5. Conclusions

This study aimed to estimate the impact of COVID-19 on NO_2, CO, $PM_{2.5}$, and SO_2 in Quito, Ecuador. A GBM algorithm was used to build 24 models to predict the air quality as if COVID-19 quarantine never existed. Unlike other weather normalization studies focusing on a single pollutant, we addressed the prediction of more pollutants and investigated the importance of more variables. The temporal features were among the most important variables for the modeling of any contaminant, while meteorological parameters are pollutant- and site-dependent.

The accuracy of the models was evaluated through a cross-validation on several years before the lockdown, starting January 2016. The evaluation metrics show a high predictive performance for NO_2 and CO, a good one for SO_2 and fair for $PM_{2.5}$, varying for different districts in the city. The best model accuracy is obtained for the traffic-busy areas. This study suggests that previous years' average concentrations are not as reliable in estimating the effect of human activities on air quality and highlights the benefit of building a model based on an ML approach, considering actual meteorological conditions and temporal features.

Finally, the lockdown effect on air pollutants was analyzed for traffic and industrial districts. A clear reduction is observed for NO_2 (−53% ± 2%), SO_2 (−45% ± 11%), and CO (−30% ± 13%) concentrations during the full lockdown. The reduction in $PM_{2.5}$ concentrations display some variability depending on the type of monitoring site (traffic site: −21% ± 3%, industrial site: −12% ± 11%). Weather events, like increased wind speed, help explain the reduced effect of mobility regulations on $PM_{2.5}$. Unavoidably, implementing partial relaxation measures, the pollution concentrations progressively return to close to the same levels as before the pandemic.

Months into the pandemic, the world is overwhelmed with the implications to the economic development, public health, mobility, and environmental politics. However, this short break in a production of anthropogenic emissions showed a different reality with a global reduction of pollution. This unexpected scenario was an opportunity to assess the weight of the human activity (especially its mobility) on the atmospheric environment (Turner et al., 2020). The sanitary emergency highlighted the importance of strict traffic policies to improve urban air quality and reduce emission of greenhouse gases. In that sense, future work must be carried out on the impact of human mobility on the atmospheric pollution in the rapidly growing developing countries and an efficient transition to cleaner transportation. Although this study focuses on the COVID-19 pandemic, the reliability of our predictions supports the idea that an ML-based modeling is a robust method to quantify the impact of any event on air pollution.

Data Availability Statement

There are no real or perceived financial conflicts of interests for any author.The data supporting the conclusions can be obtained at Mendeley Data, V1, doi: 10.17632/trs5j932s8.1. (Zalakeviciute, et al., 2020b), "Air Quality in Quito during COVID-19 outbreak" at https://data.mendeley.com/datasets/trs5j932s8/1

Acknowledgments
Funding for this study is provided by the Universidad de Las Americas, Ecuador as a part of an internal research project AMB.RZ.20.02 and CEPRA-XV-2021-018 project funded by CEDIA (Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia.

References

Bauwens, M., Compernolle, S., Stavrakou, T., Müller, J.-F., van Gent, J., Eskes, H., et al. (2020). Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using TROPOMI and OMI Observations. *Geophysical Research Letters*, *47*(11), e2020GL087978. https://doi.org/10.1029/2020GL087978

Coker, E. S., Cavalli, L., Fabrizi, E., Guastella, G., Lippo, E., Parisi, M. L., et al. (2020). The Effects of Air Pollution on COVID-19 Related Mortality in Northern Italy. *Environmental and Resource Economics*, *76*, 611–634. https://doi.org/10.1007/s10640-020-00486-1

Csavina, J., Field, J., Félix, O., Corral-Avitia, A. Y., Sáez, A. E., & Betterton, E. A. (2014). Effect of wind speed and relative humidity on atmospheric dust concentrations in semi-arid climates. *The Science of the Total Environment*, *487*, 82–90. https://doi.org/10.1016/j.scitotenv.2014.03.138

Dhaka, S. K., Chetna, V., Panwar, V., Dimri, A. P., Singh, N., Patra, P. K., et al. (2020). PM2.5 diminution and haze events over Delhi during the COVID-19 lockdown period: An interplay between the baseline pollution and meteorology. *Nature Scientific Reports*, *10*(1), 13442. https://doi.org/10.1038/s41598-020-70179-8

Ding, J., van der, A. R. J., Eskes, H., Mijling, B., Stavrakou, T., van Geffen, J., & Veefkind, P. 2020. NOx emissions reduction and rebound in China due to the COVID-19 crisis. *Geophysical Research Letters*, e2020GL089912. https://doi.org/10.1029/2020GL089912

EMASEO (2011). *Municipio del distrito metropolitano de Quito: Plan de Desarrollo 2012–2022*, (1–166). Quito. http://www.emaseo.gob.ec/documentos/lotaip_2012/s/plan_de_desarrollo_2012_2014.pdf

EPA (2018). Health and environmental effects of particulate matter (PM) [WWW Document]. *Part. Matter Pollut*. URL. https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm

European Environment Agency (2017). *Air quality in Europe — 2017 report*, (1–80). https://doi.org/10.2800/777411

European Environment Agency (2018). *Air quality in Europe — 2018 report*, (1–88). https://doi.org/10.2800/850018

European Space Agency (2020). Coronavirus: nitrogen dioxide emissions drop over Italy. [WWW Document]. *Satell. images video*. URL. https://youtu.be/ARpxtAKsORw

Faridi, S., Yousefian, F., Niazi, S., Ghalhari, M. R., Hassanvand, M. S., & Naddafi, K. (2020). Impact of SARS-CoV-2 on Ambient Air Particulate Matter in Tehran. *Aerosol Air Qual. Res.*, *20*(8), 1805–1811. https://doi.org/10.4209/aaqr.2020.05.0225

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Geographic, N. (2020). Pollution made COVID-19 worse. Now lockdowns are clearing the air. [WWW Document]. *Sci. Coronavirus Cover*. Retrieved from. https://www.nationalgeographic.com/science/2020/04/pollution-made-the-pandemic-worse-but-lockdowns-clean-the-sky/

Goldberg, D. L., Anenberg, S. C., Griffin, D., McLinden, C. A., Lu, Z., & Streets, D. G. (2020). Disentangling the Impact of the COVID-19 Lockdowns on Urban NO2 From Natural Variability. *Geophysical Research Letters*, *47*(17), e2020GL089269. https://doi.org/10.1029/2020GL089269

Grange, S. K., & Carslaw, D. C. (2019). Using meteorological normalization to detect interventions in air quality time series. *The Science of the Total Environment*, *653*, 578–588. https://doi.org/10.1016/j.scitotenv.2018.10.344

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalization models for Swiss PM 10 trend analysis. *Atmospheric Chemistry and Physics Discussions*, *18*, 6223–6239. https://doi.org/10.5194/acp-18-6223-2018

Henneman, L. R. F., Holmes, H. A., Mulholland, J. A., & Russell, A. G. (2015). Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta. *Atmospheric Environment*, *119*, 201–210. https://doi.org/10.1016/j.atmosenv.2015.08.007

INEC (2011). Quito. *Poblacion, superficie (km2), densidad poblacional a nivel parroquial*. Retrieved from https://www.ecuadorencifras.gob.ec/search/Poblaci%C3%B3n,+superficie+(km2),+densidad+poblacional+a+nivel+parroquial/

Karagulian, F., Belis, C. A., Dora, C. F. C., Prüss-Ustün, A. M., Bonjour, S., Adair-Rohani, H., et al. (2015). Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric Environment*, *120*, 475–483. https://doi.org/10.1016/j.atmosenv.2015.08.087

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*, 367–371. https://doi.org/10.1038/nature15371

Limb, M. (2016). Half of wealthy and 98% of poorer cities breach air quality guidelines. *BMJ*, *353*. https://doi.org/10.1136/bmj.i2730

NASA (2020). Airborne Nitrogen Dioxide Plummets Over China. [WWW Document]. *Earth Obs*. URL. https://earthobservatory.nasa.gov/images/146362/airborne-nitrogen-dioxide-plummets-over-china

Navinya, C., Patidar, G., & Phuleria, H. C. (2020). Examining Effects of the COVID-19 National Lockdown on Ambient Air Quality across Urban India. *Aerosol Air Qual. Res.*, *20*(8), 1759–1771. https://doi.org/10.4209/aaqr.2020.05.0256

Nikkei Asian Review, 2020. *Blue skies return to China as coronavirus cuts coal consumption - Nikkei Asian* Review [WWW Document]. Spotlight. URL asia.nikkei.com/Spotlight/Coronavirus/Blue-skies-return-to-China-as-coronavirus-cuts-coal-consumption.

Ogen, Y. (2020). Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. *The Science of the Total Environment*, *726*, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605

Pani, S. K., Lin, N.-H., & Ravindra Babu, S. (2020). Association of COVID-19 pandemic with meteorological parameters over Singapore. *The Science of the Total Environment*, *740*, 140112. https://doi.org/10.1016/j.scitotenv.2020.140112

Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., et al. (2020). Meteorology-normalized impact of the COVID-19 lockdown upon NO2 pollution in Spain. *Atmospheric Chemistry and Physics*, *20*, 11119–11141. https://doi.org/10.5194/acp-20-11119-2020

Pope, C. A., Dockery, D. W., Pope, C. A., & Dockery, D. W. (2006). In: 2012. Health effects of fine particulate air pollution: Lines that connect health effects of fine particulate air pollution. *Journal of the Air & Waste Management Association*, *56*(6), Li 2247. https://doi.org/10.1080/10473289.2006.10464485

Querol, X., Viana, M., Alastuey, A., Amato, F., Moreno, T., Castillo, S., et al. (2007). Source origin of trace elements in PM from regional background. *Urban and Industrial Sites of Spain*, *41*(37), 7219–7231. https://doi.org/10.1016/j.atmosenv.2007.05.022

Shi, X., & Brasseur, G. P. (2020). The Response in Air Quality to the Reduction of Chinese Economic Activities During the COVID-19 Outbreak. *Geophysical Research Letters*, *47*(11), e2020GL088070. https://doi.org/10.1029/2020GL088070

Turner, A. J., Kim, J., Fitzmaurice, H., Newman, C., Worthington, K., Chan, K., et al. (2020). Observed Impacts of COVID-19 on Urban CO2 Emissions. *Geophysical Research Letters*, *47*(22), e2020GL090037. https://doi.org/10.1029/2020GL090037

UN. (2015). *The 2015 revision of world population prospects*. Retrieved from https://population.un.org/wpp/Publications/Files/WPP2015_DataBooklet.pdf

Wang, P., Chen, K., Zhu, S., Wang, P., & Zhang, H. (2020). Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. *Resources, Conservation and Recycling*, *158*, 104814. https://doi.org/10.1016/j.resconrec.2020.104814

Wang, Y., Yuan, Y., Wang, Q., Liu, C., Zhi, Q., & Cao, J. (2020). Changes in air quality related to the control of coronavirus in China: Implications for traffic and industrial emissions. *The Science of the Total Environment*, *731*(20), 139133. https://doi.org/10.1016/j.scitotenv.2020.139133

WHO. (2016). Air pollution levels rising in many of the world's poorest cities [WWW Document]. *WHO*. URL. http://www.who.int/mediacentre/news/releases/2016/air-pollution-rising/en/#.WrrLXYrQNwg.mendeley

Worldometer (2020). Coronavirus Cases [WWW Document]. COVID-19 CORONAVIRUS PANDEMIC. https://doi.org/10.1101/2020.01.23.20018549V2

Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., & Dominici, F. (2020). *Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis*, *6*(45), eabd4049. Retrieved from https://advances.sciencemag.org/content/6/45/eabd4049

Žabkar, R., Honzak, L., Skok, G., Forkel, R., Rakovec, J., Ceglar, A., et al. (2015). Evaluation of the high resolution WRF-Chem (v3.4.1) air quality forecast and its comparison with statistical ozone predictions. *Geoscientific Model Development*, *8*(7), 2119–2137. https://doi.org/10.5194/gmd-8-2119-2015

Zalakeviciute, R., López-Villada, J., & Rybarczyk, Y. (2018a). Contrasted effects of relative humidity and precipitation on urban PM2.5 pollution in high elevation urban areas. *Sustainability*, *10*(6), 2064. https://doi.org/10.3390/su10062064

Zalakeviciute, Rasa Rybarczyk, Y., & Diaz, V. (2020b). Assessing the COVID-19 impact on air quality in Quito Ecuador [WWW Document]. *Mendeley Data*. https://doi.org/10.17632/trs5j932s8.1

Zalakeviciute, R., Rybarczyk, Y., Lopez Villada, J., & Diaz Suarez, M. V. (2018b). Quantifying decade-long effects of fuel and traffic regulations on urban ambient PM2.5 pollution in a mid-size South American city. *Atmospheric Pollution Research*, *9*(1), 66–75. https://doi.org/10.1016/j.apr.2017.07.001

Zalakeviciute, R., Vasquez, R., Bayas, D., Buenano, A., Mejia, D., Zegarra, R., et al. (2020a). Drastic Improvements in Air Quality in Ecuador during the COVID-19 Outbreak. *Aerosol and Air Quality Research*, *20*(8), 1783–1792. https://doi.org/10.4209/aaqr.2020.05.0254