Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning

Mario Lovrić a, *, Kristina Pavlović a, Matej Vuković b, Stuart K. Grange c, d, Michael Haberle e, Roman Kern a

a Know-Center, Inffeldgasse 13/6, AT-8010, Graz, Austria
b Pro2Future, Inffeldgasse 25F, AT-8010, Graz, Austria
c Empa, Swiss Federal Laboratories for Materials Science and Technology, 8600, Dübendorf, Switzerland
d Wolfson Atmospheric Chemistry Laboratories, University of York, York, YO10 5DD, United Kingdom
e Institute of Highway Engineering and Transport Planning, Rechbauerstraße 12, AT-8010, Graz, Austria

ABSTRACT

During March 2020, most European countries implemented lockdowns to restrict the transmission of SARS-CoV-2, the virus which causes COVID-19 through their populations. These restrictions had positive impacts for air quality due to a dramatic reduction of economic activity and atmospheric emissions. In this work, a machine learning approach was designed and implemented to analyze local air quality improvements during the COVID-19 lockdown in Graz, Austria. The machine learning approach was used as a robust alternative to simple, historical measurement comparisons for various individual pollutants. Concentrations of NO2 (nitrogen dioxide), PM10 (particulate matter), O3 (ozone) and Ox (total oxidant) were selected from five measurement sites in Graz and were set as target variables for random forest regression models to predict their expected values during the city’s lockdown period. The true vs. expected difference is presented here as an indicator of true pollution during the lockdown. The machine learning models showed a high level of generalization for predicting the concentrations. Therefore, the approach was suitable for analyzing reductions in pollution concentrations. The analysis indicated that the city’s average concentration reductions for the lockdown period were: -36.9 to -41.6%, and -6.6 to -14.2% for NO2 and PM10, respectively. However, an increase of 11.6 to 33.8% for O3 was estimated. The reduction in pollutant concentration, especially NO2 can be explained by significant drops in traffic-flows during the lockdown period (-51.6 to -43.9%). The results presented give a real-world example of what pollutant concentration reductions can be achieved by reducing traffic-flows and other economic activities.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The COVID-19 pandemic has caused disastrous health and socio-economic crises across the globe (Alabdulmonem et al., 2020; McKee and Stuckler, 2020). Questions have been raised whether atmospheric pollution is a co-factor in disease development causing a higher lethality rate, especially in highly populated and polluted areas such as those in Italy (Conticini et al., 2020; Fattorini and Regoli, 2020). A study from China suggests there is a statistically confirmed relationship between air pollution by means of elevated concentrations of PM2.5, PM10, CO, NO2 and O3 and the COVID-19 infection rate (Zhu et al., 2020). Another study from Italy supports the insight by providing causal relationships between the COVID-19 spread and air quality (Delnevo et al., 2020). An interplay of air quality and the pandemic seems obvious.

On the other side, lockdowns have caused significant changes in air quality (Dutheil et al., 2020). A study on 44 Chinese cities (Bao and Zhang, 2020) showed a decrease in main air pollutants from 5.93 to 24.67% during the lockdown while megacities such as Sao Paulo showed even higher concentration drops (40–70%) for some pollutants (Krecl et al., 2020). A study on PM2.5 in capital cities showed concentration drops of 20–60% during the COVID-19 crisis (Rodríguez-Urrego and Rodríguez-Urrego, 2020). It is suggested...
that the pollution drop was mainly driven by a reduction in traffic (Kerimray et al., 2020) and industrial activities (Li et al., 2020). Even if lockdowns hinder economic growth and might cause various negative effects in the long term, drops in pollution concentrations may act as another factor which slows disease transmission in tandem with limiting human contact. Lockdowns in Europe were instituted gradually by means of governmental interventions (Desvars-Larrive et al., 2020). This massive intervention also poses a unique opportunity to study the change in various aspects of air quality, thus motivating our study.

We discuss and explore that for complete understanding of the true factors influencing pollutant concentrations, pure statistical tests or observational comparisons might be inadequate since weather conditions, particle persistence and seasonality affect concentrations by linear and non-linear processes (Simić et al., 2020). Furthermore, transport pathways and source distribution may act as another factor which slows disease transmission in the long term, drops in pollution concentrations caused by an industrial complex nearby. Graz North received primary road segment but also records higher pollutant concentrations due to an industrial complex nearby. Graz South is situated at a heavily frequented commuter-arterial, mean pollutant concentrations are lower than at Don Bosco. Graz South is situated at a secondary road segment but also records higher pollutant concentrations due to an industrial complex nearby. Graz North and West are classified as urban background sites and are located near minor roads with no specific emission contributors in immediate vicinity. A more detailed site description, photos of the sites and historical overview of the sites is given in Moser et al. (2019).

With the intention of understanding the potential effects of traffic, the traffic flow for the city of Graz was assessed. The traffic flow data were mainly measured with inductive loop detectors where the detectors measure the change in field when objects pass over them. Once a vehicle drives over a loop sensor, the loop field changes which allows the detection of the presence of an object (a vehicle). The "Traffic control and street lighting unit of the city of Graz" monitors and records the data at 1-min time frequency and provided data from January 2017 to May 2020 for two sites, namely Don Bosco and Ost.

To determine the start, end, and duration of the Austrian lockdown, we extracted these data from a dataset which contains a collection of governmental decisions and intervention measures for different countries regarding the COVID-19 lockdowns (Desvars-Larrive et al., 2020). The dataset consists of country codes, dates and measures countries took to control their COVID-19 epidemics.

The air quality data covers PM10 and NO2, from five sites (D, N, O, S, W) described in Fig. 1. O3 is measured only at N and S. Total oxidant (O3; NO2 + O3) was calculated and included in the analysis as an additional pollutant (Grange and Carslaw, 2019). Ox was included because it would indicate if the hypothesized changes in NO2 and O3 due to the lockdown measures were caused by a repartitioning of these two species, which has consequences for air quality management. An overview of the collected data is given in Table 1. Due to a very high correlation between PM2.5 and PM10 (up to 97% for Graz South), we did not take PM2.5 into account for this study. The variables in the accompanied dataset are abbreviated as follows: <site>,<pollutant>, i.e. s_NO2 would be NO2 measured on the Graz South measurement site. The other (predictive) variables were relative humidity, air pressure, air temperature and precipitation, wind direction and wind speed collected at the same sites as the air quality data if available, see Table 1. The collected measurement data are provided in a daily frequency.

2.2. Data processing

Some data points were excluded from the analysis due to the presence of outliers. Observations between 1st and 3rd January each year were excluded due to high PM10 concentrations caused by New Year firework shows. Additionally, PM10 observations between 26th and 30th March, 2020 were excluded because of

2.1. Data description

We collected environmental, pollution and weather data from publicly available sources provided by the Austrian government.1 In order to obtain a realistic picture of air quality during the lockdown, we analyzed the long term measurement data from January 2014 to May 2020 from five measurement sites in the Austria city of Graz (Süd (eng. South) - S, Nord (eng. North) - N, West (eng. West) - W, Don Bosco – D, Ost (eng. East) – O; Fig. 1). Graz is a medium-sized European city which has much in common in respect to size and layout to many other European urban areas. The latter two measurement sites are situated on arterial roads with high traffic volumes, especially during morning and evening rush hours. The most polluted measurement site of Graz is Don Bosco that struggles to meet the annual NO2 and PM10 regulatory limits of the EU-Council directive 96/62/EC. This is primarily because of the traffic related emissions, but also because of the emissions from a nearby steel-and iron-mill (Hinterhofer, 2014). Although Graz East is located at a heavily frequented commuter-arterial, mean pollutant concentrations are lower than at Don Bosco. Graz South is situated at a secondary road segment but also records higher pollutant concentrations due to an industrial complex nearby. Graz North and West are classified as urban background sites and are located near minor roads with no specific emission contributors in immediate vicinity. A more detailed site description, photos of the sites and historical overview of the sites is given in Moser et al. (2019).

With the intention of understanding the potential effects of traffic, the traffic flow for the city of Graz was assessed. The traffic flow data were mainly measured with inductive loop detectors where the detectors measure the change in field when objects pass over them. Once a vehicle drives over a loop sensor, the loop field changes which allows the detection of the presence of an object (a vehicle). The “Traffic control and street lighting unit of the city of Graz” monitors and records the data at 1-min time frequency and provided data from January 2017 to May 2020 for two sites, namely Don Bosco and Ost.

To determine the start, end, and duration of the Austrian lockdown, we extracted these data from a dataset which contains a collection of governmental decisions and intervention measures for different countries regarding the COVID-19 lockdowns (Desvars-Larrive et al., 2020). The dataset consists of country codes, dates and measures countries took to control their COVID-19 epidemics.

The air quality data covers PM10 and NO2, from five sites (D, N, O, S, W) described in Fig. 1. O3 is measured only at N and S. Total oxidant (O3; NO2 + O3) was calculated and included in the analysis as an additional pollutant (Grange and Carslaw, 2019). Ox was included because it would indicate if the hypothesized changes in NO2 and O3 due to the lockdown measures were caused by a repartitioning of these two species, which has consequences for air quality management. An overview of the collected data is given in Table 1. Due to a very high correlation between PM2.5 and PM10 (up to 97% for Graz South), we did not take PM2.5 into account for this study. The variables in the accompanied dataset are abbreviated as follows: <site>,<pollutant>, i.e. s_NO2 would be NO2 measured on the Graz South measurement site. The other (predictive) variables were relative humidity, air pressure, air temperature and precipitation, wind direction and wind speed collected at the same sites as the air quality data if available, see Table 1. The collected measurement data are provided in a daily frequency.

2.2. Data processing

Some data points were excluded from the analysis due to the presence of outliers. Observations between 1st and 3rd January each year were excluded due to high PM10 concentrations caused by New Year firework shows. Additionally, PM10 observations between 26th and 30th March, 2020 were excluded because of

1 https://www.umwelt.steiermark.at/cms/ziel/2060750/DE/.
abnormally high values driven by a Saharan dust event (Federal Office: MeteoSwiss, 2020; Hansen, n.d.; also Supplementary Fig. 1). Inspired by Grange et al. (2018a); Šimić et al. (2020), we created binary encoded temporal variables for season, month, weekday, and day of year. Variables with \( \geq 1000 \) missing values were removed, variables with rare missing values were imputed by backfilling (see missing value counts in Table 1). The processed data consists of 2324 days and 60 variables in total and is provided in table format within a persistent data repository (Lovrić et al., 2020b). The traffic data were aggregated to a daily frequency and stored as a time series for the two sites (O, D). The processed traffic data ranges from January 2017 to May 2020.

2.3. Exploratory analysis and machine learning methods

The data was analyzed by means of explorative analysis and regression models. Data exploration was conducted by means of swarm plots (showing the distribution of median concentrations in the lockdown time frame over the years) and principal component analysis (PCA). For a comprehensive discussion of PCA, readers are advised to consult Abdi and Williams (2010). We used PCA to investigate the existence of clusters among the measurement sites.
and pollutants as well as look for potential outliers.

The proposed investigation of the lockdown pollution level consists of training machine learning models for each of the pollutants’ concentration (as the target) and observing the difference of predicted to the true values, a concept presented in Lovrić et al. (2020a) and depicted in Schema 1. Our proposal follows the assumption that any increase in residuals can be mainly attributed to changes directly or indirectly associated with the lockdown.

The period of model training was between 3rd January 2014 and 31st December 2019. The data from 2020 was separated as an external validation set, VS (3rd January 2020–10th March 2020), a lockdown set, LD (10th March 2020 – 2nd May 2020) and a hard lockdown set, HLD (20th March 2020–14th April 2020) being the time frame of interest in this work. The trained models were then used to predict on the VS and the HLD. An overview of the dataset splits together with our expectation is given in Supplementary Table 1.

Machine learning was employed to model the expected concentrations (as if no lockdown happened). For machine learning, one needs a set of predictive variables and a target variable. The methodology used here closely follows the procedures described previously in Simić et al. (2020). The target variables in the models were the various pollutant concentrations (PM10, NO2, O3, Ox) from five sites in (D, N, O, S, W). The predictive variables (X) were weather and environmental conditions as well as temporal variables accompanied by their respective lag-values (values from the previous two days). These predictive variables allowed the machine learning model to capture seasonal behavior from activities such as traffic and weather and environmental conditions as well as temporal variables accompanied by their respective lag-values (values from the previous two days). These predictive variables allowed the machine learning model to capture seasonal behavior from activities such as industrial production and traffic flows and therefore, can be thought of as surrogate variables.

The machine learning algorithm used was Random Forest regression (RF) (Breiman, 2001) which has been utilized in a number of previous air pollution models and air quality data analysis studies (Grange et al., 2018a; Simić et al., 2020). Even though there is a plethora of machine learning algorithms available, it was not the authors’ aim to compare algorithms as this was done in previous research (Lovrić et al., 2020c; Simić et al., 2020). RF was chosen due to its good generalization as shown in previous research (Grange et al., 2018a; Grange and Carslaw, 2019) and due to its simplicity regarding ingestion of heterogeneous data. For training a (predictive) machine learning model one needs the aforementioned target (y) and a set of predictive variables (X). The difference between our past work (Simić et al., 2020) and this analysis is that in this work we did not include the lag-values of the respective (predicted) pollutant as well as no other pollutant concentrations as predictors in the respective models. The RF models were optimized by means of Bayesian optimization (Lovrić et al., 2020c) with 10x cross-validation (CV) and root mean square error (RMSE) as the cost function for model optimization. Model quality evaluated automatically based on the 10xCV RMSE. A feature selection step was introduced by means of permutation importance, previously explained in Lovrić et al. (2020c); Simić et al. (2020). The models were therefore selected by their best performance independent of how many features were ingested. Data processing and model training was conducted with Python. The functions, scripts and libraries are presented in prior work (Lovrić et al., 2020c; Simić et al., 2020).

2.4. Concept validation and method comparison

The obtained machine learning results, that is, the predicted (expected) pollutant concentrations were compared to the measured (true) values. Our machine learning approach was then evaluated against historical changes in the data (short and long term). To understand the potential causes, we evaluated the drop in pollutant concentration against the drop in traffic density. Furthermore, the modeled concentration reductions were aggregated to understand the city-wide average concentration reduction.

3. Results and discussion

3.1. Explorative analysis

For better understanding of the relationship between the pollutants, we conducted an exploratory data analysis phase to check the data quality and validate the relationships between the measurements. We opted for principal component analysis (PCA) due to its popularity and sensitivity regarding outliers. The PCA was conducted on the pollutants’ concentrations between 2014 and 2020 (Fig. 2). The first two principal components explain 91.0% variance and were selected for the explorative analysis. The PCA loadings plots show that the distinct measured pollutants group based on

Schema 1. Overview of the study methodology in order to detection changes in the relationships between the (dependent and independent) variables. Supplementary Table 1 conveys more details on the methodology.
their chemical composition, i.e. the same pollutants group together independent of the measurement site. Despite PM$_{10}$ have a wide range of sources, this pollutant was similar to NO$_2$, indicating that the PM$_{10}$ concentrations in Graz were primarily sourced from traffic processes during the analysis period. One can observe that in the two groups (PM$_{10}$ and NO$_2$), the northern site N has lower PC2 loadings. Furthermore, for NO$_2$ and PM$_{10}$, N is closer to W, which could be explained by both sites being less burdened by traffic emissions. On the contrary for the Don Bosco site (D), NO$_2$ and PM$_{10}$ were in proximity in the plot, which might be explained by a rather common source, most likely traffic, which is heavy at this site. Because of the common pattern of traffic-loaded sites having higher PC2 loads it can be inferred that the PC2 is more burdened by traffic emissions. O$_3$ showed distinctly different patterns because O$_3$ is generated by secondary processes and not directly emitted.

To understand the changes during the lockdown, we compared the HLD data during the time frame (20th March - 14th April 2020) to the same time frame during 2014–2019. The swarm plots of the pollutants’ concentrations are presented in Fig. 3. In the given time frames, particularly for NO$_2$, the concentrations were lower in 2020 as compared to the years 2014–2019. PM$_{10}$ and O$_x$ do not show clear patterns, whereas O$_3$ appears to have higher average concentration compared to data from 6 years before. Therefore, a general trend is present, traffic-sourced pollutants show concentration drops, while PM$_{10}$ shows a drop at Don Bosco, which is a traffic-burdened site.

### 3.2. Machine learning results

To gain understanding of the true changes in pollutant concentrations, we trained RF models with concentrations as target variables. The optimal/best models (results in Supplementary Table 2) obtained were re-trained on the complete training data sets between 2014 and 2019 and subsequently fitted to data from 2020 (VS, HLD, LD). The results of the RF modelling by means of the coefficient of determination ($R^2$), the root mean square error (RMSE) and the normalized RMSE (NRMSE%) (Simić et al., 2020) are given in Table 2. Time series plots (as 7-day moving averages) for the predictions (in 2020) and their respective true values for the six pollutants are shown in Fig. 4.

The results from Table 2 and Fig. 4 show that for the validation period (VS), the models’ prediction quality increases by means of %NRMSE (%NRMSE decreases) on average in following order PM$_{10}$ $>$ O$_3$ $>$ NO$_2$ $>$ O$_x$, while for the $R^2$ score the prediction quality increases as follows O$_x$ $<$ NO$_2$ $<$ PM$_{10}$ $<$ O$_3$. Additional processes such as long-range transport and secondary generation helps drives PM concentrations and these processes are not as relevant to the
other gaseous pollutants analyzed here (Grange et al., 2018b; Viana et al., 2008). This increased complication and additional processes are likely the reason for the decreased model performance for the PM10 pollutant. NO2 and O3 have reasonable prediction performance (validation set), i.e., O3 has R2 scores 0.84 (S) and 0.87 (N), while NO2 has R2 scores as high as 0.81 (W) and 0.76 (N) with lower scores for the traffic-loaded sites (S ≤ 0.45, O ≤ 0.50, D ≤ 0.55). A similar pattern appears with PM10 where the lower R2 scores are related to traffic-loaded sites (W ≤ 0.66, N ≤ 0.72) and S ≤ 0.71. These results overall suggest that the concept of using machine learning (ML) models can support understanding the true pollution.

For the PM10 and NO2 pollutants, the models show concentration reductions (Fig. 4 and Supplementary Fig. 2) where the observed concentrations were lower than those predicted by the RF models. O3 showed the opposite behavior where observed concentrations were higher than those predicted. The increases in O3 concentrations can be explained by a reduction of the NO−O3 titration cycle when NOx emissions (and concentrations) were low during the lockdown period.

When comparing the ratios of %NRMSE HLD to %NRMSE VS one can see the largest ratios with NO2 across the sites (2.77–3.98) meaning that they show the largest error in predicted (expected) vs true (measured) concentrations during the lockdown. The ratio is on average lower with O3 (0.41–0.56). The inverse results are due to a concentration rise instead of a concentration drop. Ox shows a less clear pattern; underprediction for North and overprediction for South (Supplementary Fig. 2). With PM10 the ML models suggest that the HLD and VS errors (0.95–1.36) do not deviate largely from each other, pointing to PM10 not being largely affected by the lockdown. Also, there is a period of unexpected high PM10 at the end of HLD which we attribute to lockdown fatigue. Regarding the sites, one can see the largest ratios across pollutants (%NRMSE VS to HLD) are at the East (O) and Don Bosco (D) sites which are more traffic-loaded than others.

### 3.3. Reduction in pollution/method comparison

With the aim of supporting the contribution of machine learning in understanding the pollution, we present a method comparison. The pollution during HLD (median 2020) was lower, while during VS it was higher: NO2 was lower (S 0.55, O 0.76, N 0.5), and PM10 was higher in HLD (S 0.95, O 0.72, N 0.72). For NO2, the HLD concentration was lower than the VS concentration for S (0.55 vs 0.61) and N (0.5 vs 0.76), while for PM10 the HLD concentration was higher than the VS concentration for S (0.95 vs 0.84) and O (0.72 vs 0.71). The concentration difference between HLD and VS for O3 was lower than that for NO2 and PM10, with O3 showing a concentration increase in HLD (S 0.84, O 0.87, N 0.71) and a decrease in VS (S 0.72, O 0.76, N 0.71). This indicates that the ML models can support understanding the true pollution.
compared to the respective medians of 2019 and 2014–2019 (i.e. a historical comparison) as well as the median of the predicted HLD 2020. The comparison is presented in Fig. 5.

The results show an overall good agreement for change in pollution of HLD 2020 referred to 2019 (short-term, in blue) and referred to 2014–2019 (long-term, in orange). Notable exceptions are: N_PM10 which shows an increase of 10.5% short-term and a 0.2% reduction long-term and W_NO2 showing a larger concentration drop in the short-term (30.6%) when compared to long-term (21.4%). One can also observe generally good agreement when comparing changes referred to 2019 and to the predictions of HLD 2020. Pollutants, for which the results disagree based on the short-term vs predicted comparison the most are: N_PM10, N_Ox, N_O3, S_O3, S_NO2 and W_NO2. For NO2 and O3, where good prediction results were achieved, one can assume that the true pollution change is lower than estimated by comparing it to the values from previous years. The agreement between the methods validates the machine learning approach as an alternative for analyzing the pollution drop as no ground truth data are available during the lockdown.

For the pollutants where five measurement sites are available, the median concentration drops were averaged across all measurement sites (D, N, O, S, W) to a “city average concentration drop”. When comparing the city average concentration drops for 2020 to the short-term (vs HLD, 2019), long-term (vs HLD, 2014–2019) and prediction (i.e. predicted HLD, 2020) one can see that for PM10 the ML models show a larger concentration drop whereas for NO2 we see a smaller concentration drop with the ML models (Table 3, top).

From the available traffic data, we calculated a reduction in median traffic during HLD 2020 against the same median time frame in 2019 and in 2017–2019 for the measurement sites D and O which are located close to roads (Fig. 1). The traffic measured at the detector loops at these measurement sites showed a reduction of 45.6% at O and 51.6% at D respectively (Table 3, bottom). This reduction of traffic can be correlated with the large reduction of NO2 at measurement site D and O (Fig. 5), as traffic was assumed to be the main sources for this pollutant (EEA, 2019). In contrast, the reduction of PM10 is not as pronounced. This demonstrates that traffic is one of many contributing sources of PM10 in Graz (Moser et al., 2019). Since industrial processes were also significantly curtailed in and around Graz during the lockdown period, it could be expected that PM10 would decrease more drastically. The lack of reduction outlines the more numerous and complicated processes which drive PM10 concentrations, and/or a high lag in the relationships involved in the PM10 related variables, i.e., the duration of the intervention was too short to lead to a significant drop in this particular pollutant.

Our results agree regarding a traffic-related drop in NO2 and an increase in O3 with a variety of studies conducted on air quality issues regarding the COVID-19 lockdown. A study from northern

| Reduction in NO2 | City average 2020 vs 2019 | City average 2020 vs 2014–2019 | City average 2020 vs predicted |
|------------------|---------------------------|-------------------------------|-------------------------------|
| Reduction in PM10 | -41.6%                    | -38.1%                        | -36.9%                        |
|                  | -6.6%                     | -11.0%                        | -14.2%                        |

| Median 2019 | Median 2017–2019 |
|-------------|------------------|
| Traffic reduction (D) | -51.6% | -43.9% |
| Traffic reduction (O)  | -45.6% | -45.6% |
China shows that a reduction in NO$_2$ and PM was most likely caused by a reduction in traffic and industrial activities (Wang et al., 2020). Furthermore, they also show an increase in O$_3$ consistent with our analysis as well. Another study from China where lockdown measures were introduced earlier than in Austria (1st January – 29th February 2020) (Shi and Brasseur, 2020) reveals a reduction of NO$_2$ up to 60% and a O$_3$ increase of 40–100%. A two-stage lockdown in India, which was introduced concurrently with the Austrian lockdown (Mahato et al., 2020), and data from this region shows a PM$_{10}$ concentration drop around 60% compared to the same period in 2019. These results are contrary to ours, since Graz did not experience such a PM$_{10}$ drop, revealed by both historical comparison and machine learning prediction. Reasons for this may be domestic heating which is difficult to evaluate since data on heating and stays in the city are difficult data to obtain. The same study observed consistent results in the drop of NO$_2$ (−52.6%). A study from the UK, which employed also a historical comparison shows a reduction of 48% in NO$_2$ concentration and an average increase in O$_3$ concentration of 11% across 126 urban sites, which is also consistent with our analysis. A maybe more relevant comparison to our study is given by Menut et al. (2020) since their study is conducted using a forecasting model to analyze COVID-19 caused pollution reduction Europe-wide. Their results show broad ranges in daily deltas for three pollutants, namely NO$_2$, PM$_{2.5}$ and O$_3$. For NO$_2$ at urban sites the reductions ranged −13% (Sweden) to −57.8% (Portugal); for PM$_{2.5}$ (urban sites) the range was from −23.5% (Portugal) to −4.6% (Poland). Ozone showed a variability (urban sites) from −2.7% (Ireland) to −17.6% (Belgium).

As the Air quality report in Europe (EEA, 2019) stated, there are still significant uncertainties regarding the meteorological effects on PM$_{10}$. Furthermore, our previous research (Simić et al., 2020) showed that among PM$_{10}$, NO$_x$, EC (elemental carbon) and OC (organic carbon); PM$_{2.5}$ was the most difficult to predict accompanied by the fact that there was a high auto-correlation for PM$_{10}$ (77%). In a broad analysis of PM$_{10}$ predictivity (Grange et al., 2018a) one can observe a huge variability for PM$_{10}$ model quality (R$^2$ 53.5–70.6). However, the predictive scores show that better models can be achieved at urban traffic sites, meaning traffic-driven PM$_{10}$ could be easier to predict with the given variables. Even though some research shows a decline in PM$_{10}$ during the lockdown we believe the fixed-effect and observational methods may not be enough to deliver a definite conclusion on the concentration reductions. It is a limitation of this study that long-range transport data and chemical speciation of the particulate matter is not available. More efforts must be put into chemical speciation of PM at the individual sites, especially measurement techniques which deliver “online” data.

4. Conclusions

In this work, we have explored the changes in air pollutant concentrations during the COVID-19 lockdown for the city of Graz, Austria. The exploration illuminated the relative influences of observed meteorological variables on a wide range of pollutants for an unprecedented historic event of human society. Besides using explorative methods, we employed random forest regression to analyze the differences between predicted (expected) and observed (true) pollution levels based on environmental data.

Our prediction models showed good generalization and performance for the analyzed pollutants indicating that the selection of independent variables (predictors) was sufficient to explain changes in pollutant concentrations. For PM$_{10}$ and NO$_x$, the predicted values were found to be above the measured concentrations during the lockdown. O$_3$ was under-predicted during the period of lockdown which can be explained with the inverse relationship with NO$_x$ concentrations. NO$_x$ emissions (and concentrations) were substantially reduced during the lockdown because of much lower traffic volumes across the city of Graz.

Although PM$_{10}$ concentrations were lower than expected during the lockdown period in Graz, the reduction was modest compared to NO$_2$. This demonstrates the more varied processes which control PM concentrations in urban atmospheres and potentially weaker model performance. Coincidentally, during the lockdown period, much of western Europe was affected by a Saharan dust event between March 26 and 30 (Federal Office: MeteoSwiss, 2020) and this confounded the PM climate in many European cities during this time. Other studies which have shown greater reductions in PM concentrations during the COVID-19 lockdowns were often conducted in megacities which present a very different urban environment when compared to Graz.

Our findings show that machine learning is a suitable tool to analyze pollution changes during events where atmospheric emissions change rapidly, such as the COVID-19 lockdown example presented. Although, the expected to true differences in pollutant concentrations based on machine learning models showed similar results with regard to the observational methods (historical comparison), it was an important technique to employ because it enabled for far more robust comparisons with the observed time series. Still, additional studies are needed with a wider scope (in terms of different geographical regions, possible additional influencing factors, as well as temporal analysis) to improve model generalization to obtain better estimates of event-based air pollution reductions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2020.115900

Authors’ contributions

Mario Lovrić – Concept, Machine learning, Writing. Kristina Pavlović – Literature research, Visualization, Data collection and curation. Matej Vuković – Machine learning, Visualization, Writing. Stuart K. Grange – Writing, Interpretation, Revision and editing. Michael Haberl – Data collection and curation, Writing, Interpretation. Roman Kern – Concept, Writing, Supervision, Revision and editing.

Funding

The Know-Center is funded within the Austrian COMET Program – Competence Centers for Excellent Technologies — under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG. SKG is supported by the Swiss Federal Office for the Environment (FOEN) and by the Natural
Environment Research Council (NERC) while holding associate status at the University of York.

Availability of data

The pollution data used for modelling is published (Lovric et al., 2020b). There is no informed consent for making the traffic data accessible or published due to internal guidelines.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. https://doi.org/10.1002/wics.101.

Abalibdahmonen, W., Sharig, A., Rasheed, Z., 2020. COVID-19: a global public health disaster. Int. J. Health Sci. 14, 7–8.

Bao, R., Zhang, A., 2020. Does lockdown reduce air pollution? Evidence from 44 cities in northern China. Sci. Total Environ. 730, 139083 https://doi.org/10.1016/j.scitotenv.2020.139083.

Bak, R., Zhang, A., 2020a. COVID-19 lockdowns: analysis of traffic and air pollution changes in China. Environ. Pollut. 253, 113411. https://doi.org/10.1016/j.envpol.2020.113411.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Bu, T., Jiang, D., 2020a. Prediction of anode lifetime in electro galvanizing lines by big data analysis. In: GALVA-Tech 2020. Vienna.

Bu, T., Jiang, D., 2020b. A study on the prediction of anode lifetime in electro galvanizing lines by big data analysis. In: GALVA-Tech 2020. Vienna.

Duret, F., Baker, J.S., Navel, V., 2020. COVID-19 as a factor in influencing air pollution. Environ. Pollut. 261, 114645 https://doi.org/10.1016/j.envpol.2020.114645.

Dutheil, F., Serret, C., Le Meur, S., Allégre, C.J., 2020. Air pollution and COVID-19: evidence from China. Sci. Total Environ. 741, 140426. https://doi.org/10.1016/j.scitotenv.2020.140426.

Endo, I., Takahashi, S., Ito, F., 2020. Review of methods and results. J. Aerosol Sci. https://doi.org/10.1016/j.jaerosci.2008.05.007.

Fattorini, D., Regoli, F., 2020. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. Environ. Pollut. 264, 114732 https://doi.org/10.1016/j.envpol.2020.114732.

Filippi, C., Mazzoli, M., Scambia, M., Filippi, C., 2020a. Effect of lockdown amid COVID-19 pandemic on air quality. Environ. Pollut. 259, 113806 https://doi.org/10.1016/j.envpol.2020.113806.

Filippi, C., Mazzoli, M., Scambia, M., Filippi, C., 2020b. Effect of lockdown amid COVID-19 pandemic on air quality. Environ. Pollut. 259, 113806 https://doi.org/10.1016/j.envpol.2020.113806.