



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 Haberl^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

ARTICLE INFO

Article history:

Received 14 August 2020

Received in revised form

15 October 2020

Accepted 17 October 2020

Available online 6 November 2020

Keywords:

PM₁₀

NO₂

Total oxidant

O_x

O₃

Random forest

Corona crisis

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 O_x (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 11.6–33.8% 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 PM_{2.5}, 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 PM_{2.5} in capital cities showed concentration drops of 20–60% during the COVID-19 crisis (Rodríguez-Urrego and Rodríguez-Urrego, 2020). It is suggested

[☆] This paper has been recommended for acceptance by Da Chen.

* Corresponding author.

E-mail address: mlovric@know-center.at (M. Lovrić).

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 (Šimić et al., 2020). Furthermore, transport pathways and source distribution can also play a role in analyzing the effects of the lockdown on pollution by means of trajectory models (Zhao et al., 2020). Therefore, a comparison of air quality in 2020 vs other years may be biased since other independent factors such as shifts in heating seasons or weather conditions can affect air quality (Schiermeier, 2020). To be able to solve independent factor some authors proposed fixed effect models (Liu et al., 2020; Venter et al., 2020). Our proposal is that the pollution level can be solved as a multivariate problem predicted by independent variables elevated from environmental variables and seasonal trends, i.e. there are many effects and they might not have fixed effects because the atmosphere is a very dynamic system. Moreover, if one wants to return a full time series, observational and fixed effect methods might fail if not accounted for environmental dependencies. For predicting the pollutant concentrations, we employed the Random Forest algorithm, a non-linear regression method which has the power to solve multivariate problems independent of the variable type. A complementary approach found in literature is a forecasting method which accounted for atmospheric and other effects but using a mechanistic instead of a data-driven machine learning approach (Menut et al., 2020).

We investigate the effects of lockdown on air quality in an urbanized area in Graz, Styria, Austria. Due to the high degree of traffic influence, we have included traffic data into our analysis. Furthermore, we have investigated in detail which of the pollutants' concentrations were influenced by the lockdown. As such, the outcome of our study serves as a guide for future interventions and their expected associated change in the pollutants' concentration changes.

2. Materials and methods

Our study contains traditional exploratory statistical analysis, including the utilization of principal component analysis (PCA) to explore key attributes. However, the primary analysis is based on machine learning (ML) models which were used to capture historical relationships between the attributes and compare the predictions to true pollution values after the COVID-19 lockdowns were imposed. We utilize historical data which matches the time frame of the lockdown for the preceding years, but also include traffic flow data to represent the drop in mobility.

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 accessed. 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 (O_x; NO₂ + O₃) was calculated and included in the analysis as an additional pollutant (Grange and Carslaw, 2019). O_x was included because it will 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 PM_{2.5} and PM₁₀ (up to 97% for Graz South), we did not take PM_{2.5} 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/>.

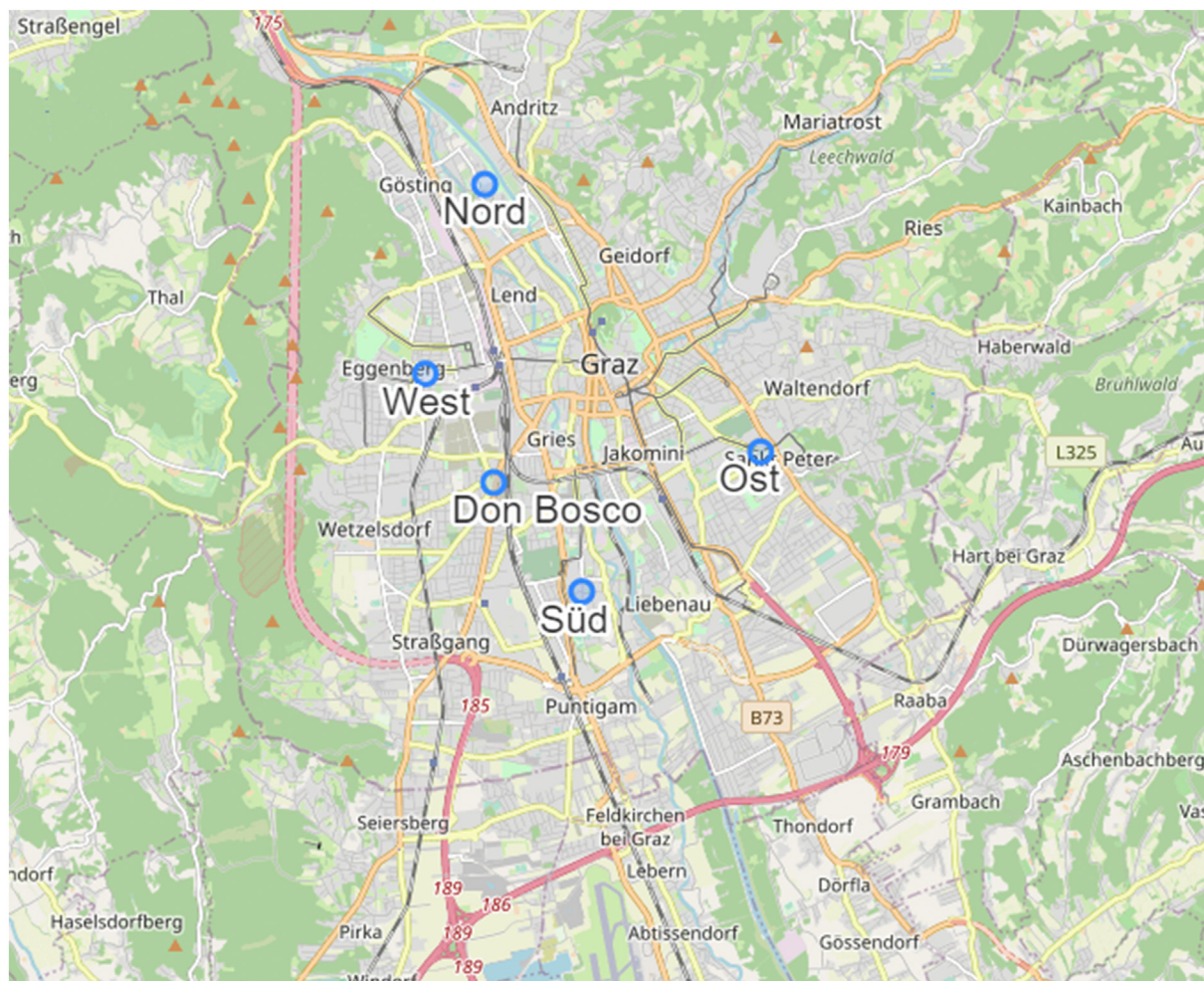


Fig. 1. A city map of Graz indicating the five measurement sites: Süd – 47.041692° N, 15.433078° E; Nord – 47.09437° N, 15.415122° E; West – 47.069506° N, 15.403728° E; Don Bosco – 47.055617° N, 15.416539° E; Ost – 47.059530° N, 15.466634° E.

Table 1

Site description for the air pollution measurement sites, data taken from Moser et al. (2019) and <http://app.luis.steiermark.at/luft2/suche.php>. The number in brackets shows the amount of missing values in data, referring to 2324 values (full number of data entries).

| Measurement | Don Bosco (D) | Nord (N) | Ost (O) | Süd (S) | West (W) |
|---------------------------------------|---------------|----------|----------|---------|----------|
| O ₃ [μg/m ³] | | x (23) | | x (39) | |
| PM ₁₀ [μg/m ³] | x (13) | x (14) | x (15) | x (31) | x (30) |
| NO ₂ [μg/m ³] | x (3) | x (8) | x (10) | x (12) | x (30) |
| %RH | x (0) | x (12) | x (1244) | x (21) | x (8) |
| Air temperature [°C] | x (0) | x (12) | x (1239) | x (45) | x (0) |
| Precipitation [l/m ²] | | x (12) | | | |
| Wind speed [m/s] | | x (12) | x (1239) | x (4) | x (0) |
| Wind direction [Deg] | | x (12) | x (1239) | x (4) | x (0) |
| Air pressure [mbar] | | x (12) | x (1239) | | |

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 ≥ 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 Šimić et al. (2020). The target variables in the models were the various pollutant concentrations (PM₁₀, NO₂, O₃, O_x) 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 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; Šimić 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; Šimić 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 (Šimić 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); Šimić 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; Šimić et al., 2020).

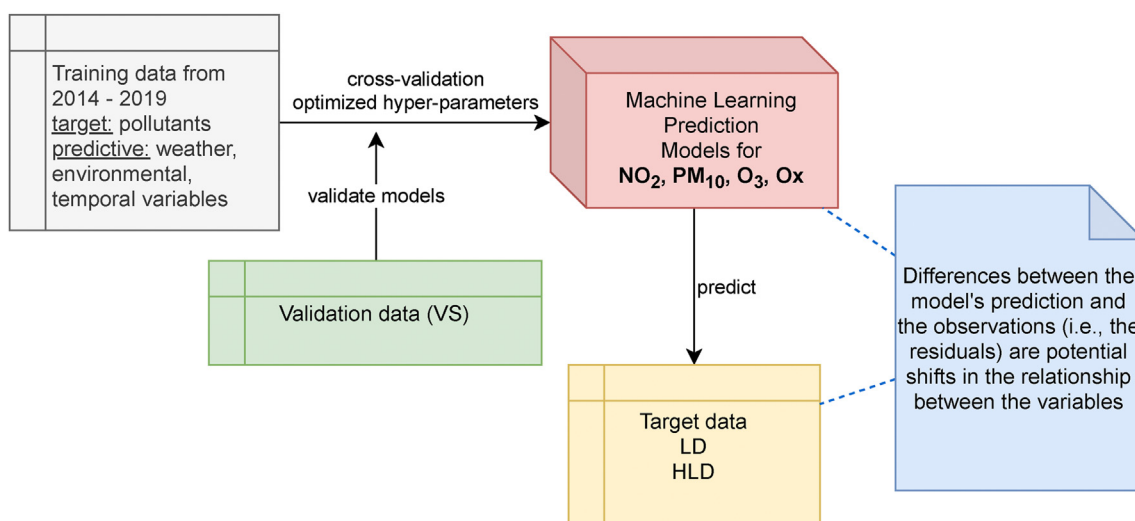
2.4. Concept validation and method comparison

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.

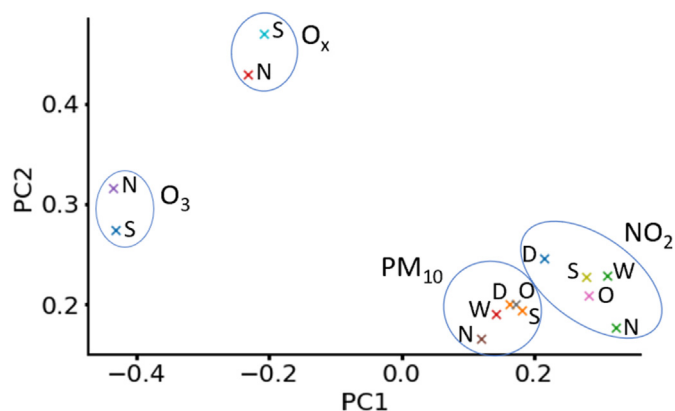


Fig. 2. A PCA loadings plot calculated from pollutant concentrations (PM_{10} , NO_2 , O_3 , O_x) from five sites in (D, N, O, S, W) Graz, Austria, described previously in Fig. 1. The pollutants form groups by means of chemical composition.

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%) (Šimić 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 $PM_{10} > O_3 \gg NO_2 > O_x$, while for the R^2 score the prediction quality increases as follows $O_x \ll NO_2 < PM_{10} < O_3$. Additional processes such as long-range transport and secondary generation helps drives PM concentrations and these processes are not as relevant to the

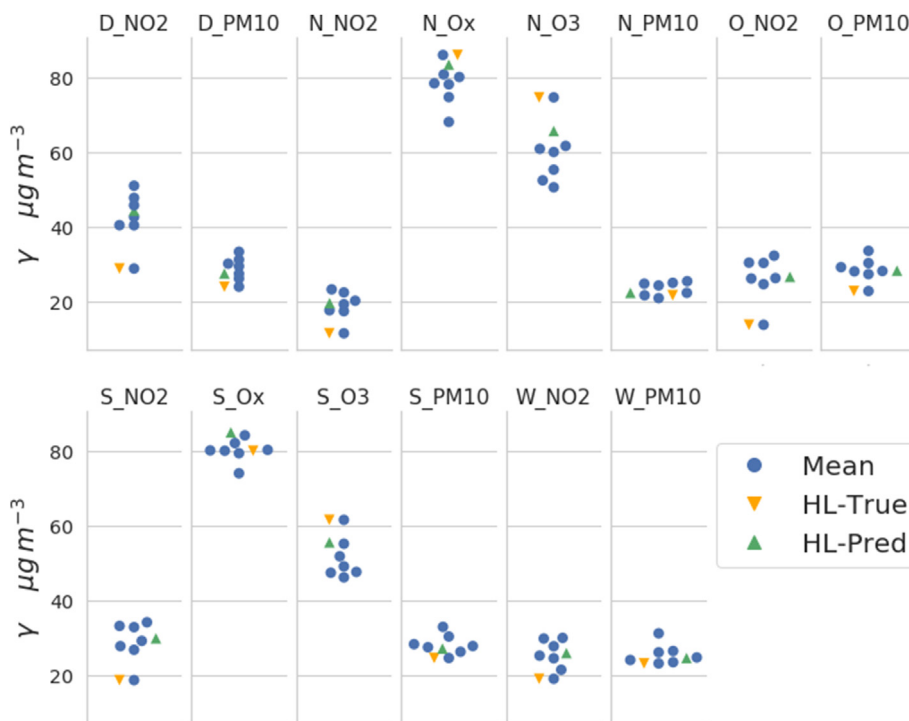


Fig. 3. Swarm plots of pollutant mean concentrations at the five sites during the hard lockdown (HLD) time frame (20th March – 14th April) through years 2014–2020. The 2020 value is colored according to whether it is true or predicted by machine learning (see Machine Learning results section).

Table 2
R², RMSE and %NRMSE value for the external validation and lockdown sets from the best model.

| D_NO2 | D_PM10 | N_NO2 | N_Ox | N_O3 | N_PM10 | O_NO2 | O_PM10 | S_NO2 | S_Ox | S_O3 | S_PM10 | W_NO2 | W_PM10 | |
|-------|--------|-------|--------|-------|--------|--------|--------|-------|-------|-------|--------|-------|--------|-------------------|
| 0.55 | 0.61 | 0.76 | 0.5 | 0.87 | 0.72 | 0.5 | 0.42 | 0.45 | -0.07 | 0.84 | 0.71 | 0.81 | 0.66 | R2 VS |
| 15.32 | 26.29 | 17.52 | 12.46 | 24.6 | 28.81 | 22.35 | 38.78 | 20.04 | 21.46 | 33.37 | 25.08 | 13.47 | 29.1 | %NRMSE VS |
| 46.64 | 29.63 | 59.89 | 8.96 | 14.01 | 27.96 | 80.1 | 37.27 | 53.97 | 13.78 | 17.35 | 28.76 | 38.59 | 28.3 | %NRMSE LD |
| 58.28 | 35.98 | 81.29 | 7.1 | 13.9 | 30.15 | 103.82 | 36.96 | 68.26 | 8.18 | 13.9 | 33.33 | 49.27 | 30.88 | %NRMSE HLD |
| 7.43 | 9.04 | 5.3 | 7.19 | 6.76 | 7.09 | 7.63 | 13.44 | 7.26 | 11.92 | 6.43 | 8.52 | 4.74 | 8.11 | RMSE VS |
| 14.48 | 7.22 | 7.83 | 7.17 | 9.38 | 5.33 | 12.3 | 8.02 | 10.9 | 10.31 | 9.47 | 6.65 | 7.91 | 6.23 | RMSE LD |
| 16.76 | 8.6 | 9.25 | 6.11 | 10.38 | 6.51 | 14.18 | 8.4 | 12.67 | 6.56 | 8.56 | 8.16 | 9.32 | 7.11 | RMSE HLD |
| 38 | 27.58 | 20.06 | 68.73 | 48.67 | 20.71 | 23.12 | 26.34 | 26.26 | 65.6 | 39.33 | 26.46 | 25.9 | 23.11 | True Mean |
| 81.8 | 78.77 | 53.36 | 106.89 | 93.71 | 65.56 | 58.14 | 111.51 | 65.05 | 97.77 | 86.13 | 78.23 | 62.28 | 69.47 | True Max |
| 8.02 | 6.63 | 1.14 | 34.14 | 3.64 | 5.6 | 1.98 | 7.59 | 2.92 | 28.67 | 1.08 | 6.93 | 3.11 | 5.58 | True Min |

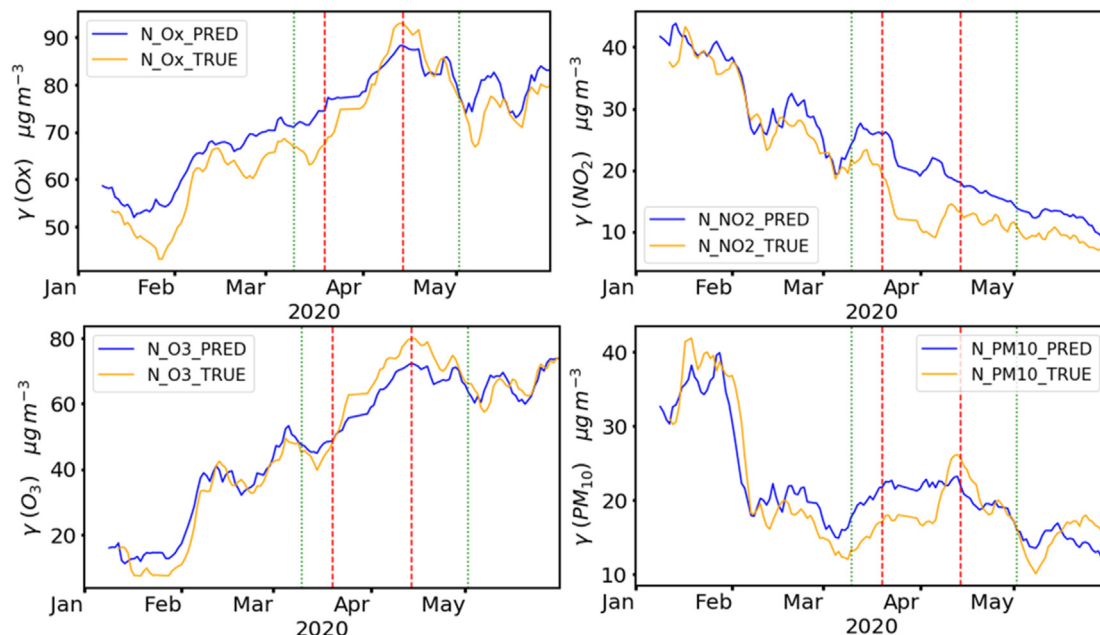


Fig. 4. Time series plots for four pollutants' concentrations measured at Graz Nord (N). Orange line (measured) is compared to their predicted values (blue). The plots present a 7-day moving average (for better visibility) for the data in 2020. Prior to the green line is the validation set (3rd January 2020–10th March 2020). The green dashed lines show the LD time frame (10th March 2020 – 2nd May 2020) and red dashed lines show the HLD time frame (20th March 2020–03-20 – 14th April 2020). Top left is O_x, top right is NO₂, bottom left is PM₁₀ and bottom right O₃. Additional plots are provided in [Supplementary Fig. 2](#). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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 (O – 0.42, D – 0.61). Better scores were achieved at less 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 NO_x 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. O_x 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

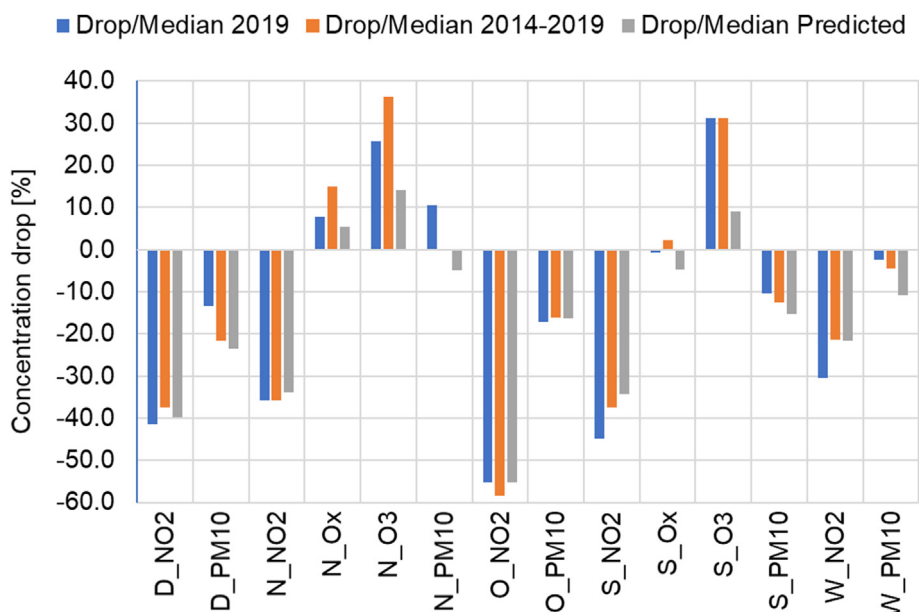


Fig. 5. Calculated pollution reduction [%] from the median HLD values in the time frame (20th March - 14th April). Median concentration drops (HLD, 2020 to 2019, i.e. short-term) are compared per pollutant to concentration drops (HLD, 2020 to 2014–2019, i.e. long-term) and the concentration drops calculated from the predicted values (Data for the plot is provided in [Supplementary Table 3](#)).

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

Table 3

top) The median pollution reductions averaged across sites (D, N, O, S, W) in 2020 when comparing the HLD median to 2019, to HLD median 2014–2019 and to HLD predicted 2020 in percentage. (Data is provided in [Supplementary Table 3](#)); bottom) Calculated traffic reduction in 2020 when comparing the HLD median to 2019 and to the 2017–2019 HLD median in percentage.

| | City average 2020 vs 2019 | City average 2020 vs 2014–2019 | City average 2020 vs predicted |
|-------------------------------|---------------------------|--------------------------------|--------------------------------|
| Reduction in NO ₂ | –41.6% | –38.1% | –36.9% |
| Reduction in PM ₁₀ | –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 NO₂ 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 O₃ 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 NO₂ up to 60% and a O₃ 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 PM₁₀ concentration drop around 60% compared to the same period in 2019. These results are contrary to ours, since Graz did not experience such a PM₁₀ drop, revealed by both historical comparison and machine learning prediction. Reasons for that 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 concordant results in the drop of NO₂ (–52.68%). A study from the UK, which employed also a historical comparison shows a reduction of 48% in NO₂ concentration and an average increase in O₃ 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 NO₂, PM_{2.5} and O₃. For NO₂ 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 PM₁₀. Furthermore, our previous research (Šimić et al., 2020) showed that among PM₁₀, NO₂, EC (elemental carbon) and OC (organic carbon); PM₁₀ was the most difficult to predict accompanied by the fact that there was a high auto-correlation for PM₁₀ (77%). In a broad analysis of PM₁₀ predictivity (Grange et al., 2018a) one can observe a huge variability for PM₁₀ model quality (R²% 53.5–70.6). However, the predictive scores show that better models can be achieved at urban traffic sites, meaning traffic-driven PM₁₀ could be easier to predict with the given variables. Even though some research shows a decline in PM₁₀ 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 unrepresented 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 PM₁₀ and NO₂, the predicted values were found to be above the measured concentrations during the lockdown. O₃ 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 PM₁₀ concentrations were lower than expected during the lockdown period in Graz, the reduction was modest compared to NO₂. 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 (Lovrić 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>.
- Alabdulmonem, W., Shariq, 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.* 731, 139052 <https://doi.org/10.1016/j.scitotenv.2020.139052>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Conticini, E., Frediani, B., Caro, D., 2020. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* 261, 114465 <https://doi.org/10.1016/j.envpol.2020.114465>.
- Delnevo, G., Mirri, S., Rocchetti, M., 2020. Particulate matter and COVID-19 disease diffusion in Emilia-Romagna (Italy). Already a cold case? *Computation* 1–16.
- Desvars-Larrive, A., Dervic, E., Haug, N., Niederkrotenthaler, T., Chen, J., Di Natale, A., Lasser, J., Gliga, D.S., Roux, A., Chakraborty, A., Ten, A., Dervic, A., Pacheco, A., Cserjan, D., Lederhiller, D., Berishaj, D., Flores Tames, E., Takriti, H., Korb, J., Reddish, J., Stangl, J., Hadziavdic, L., Stoeger, L., Goorah, L., Geyrhofer, L., Ferreira, M.R., Vierlinger, R., Holder, S., Alvarez, S., Haberfellner, S., Ahne, V., Reisch, V., Servedio, V.D.P., Chen, X., Pocasangre-Orellana, X.M., Garcia, D., Thurner, S., 2020. A structured open dataset of government interventions in response to COVID-19. *medRxiv*. <https://doi.org/10.1101/2020.05.04.20090498>. Cold Spring Harbor Laboratory Press.
- Dutheil, F., Baker, J.S., Navel, V., 2020. COVID-19 as a factor influencing air pollution? *Environ. Pollut.* 263 <https://doi.org/10.1016/j.envpol.2020.114466>, 2019–2021.
- EEA, E.E.A., 2019. Air Quality in Europe - 2019 Report.
- 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>.
- Federal Office: MeteoSwiss, 2020. Saharan dust events - MeteoSwiss. accessed 7.31.20. <https://www.meteoswiss.admin.ch/home/climate/the-climate-of-switzerland/specialties-of-the-swiss-climate/saharan-dust-events.html>.
- Grange, S.K., Carslaw, D.C., 2019. Using meteorological normalisation to detect interventions in air quality time series. *Sci. Total Environ.* 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., 2018a. Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmos. Chem. Phys.* 18, 6223–6239. <https://doi.org/10.5194/acp-18-6223-2018>.
- Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E., Hueglin, C., 2018b. Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmos. Chem. Phys.* 18, 6223–6239. <https://doi.org/10.5194/acp-18-6223-2018>.
- Hansen, K., 2020. More dust blows out from North Africa [WWW document]. accessed 7.31.20. <https://www.earthobservatory.nasa.gov/images/146407/more-dust-blows-out-from-north-africa>.
- Hinterhofer, M., 2014. Anteil der verkehrsbedingten PM10 und PM2,5 Emissionen aus Abrieb und Wiederaufwirbelung an der Feinstaubbelastung in Österreich. Technische Universität Graz.
- Kerimay, A., Baimatova, N., Ibragimova, O.P., Bukenov, B., Kenessov, B., Plotitsyn, P., Karaca, F., 2020. Assessing air quality changes in large cities during COVID-19 lockdowns: the impacts of traffic-free urban conditions in Almaty, Kazakhstan. *Sci. Total Environ.* 730, 139179. <https://doi.org/10.1016/j.scitotenv.2020.139179>.
- Krecl, P., Targino, A.C., Oukawa, G.Y., Cassino Junior, R.P., 2020. Drop in urban air pollution from COVID-19 pandemic: policy implications for the megacity of São Paulo. *Environ. Pollut.* 265, 19–21. <https://doi.org/10.1016/j.envpol.2020.114883>.
- Li, L., Li, Q., Huang, L., Wang, Q., Zhu, A., Xu, J., Liu, Ziyi, Li, H., Shi, L., Li, R., Azari, M., Wang, Y., Zhang, X., Liu, Zhiqiang, Zhu, Y., Zhang, K., Xue, S., Ooi, M.C.G., Zhang, D., Chan, A., 2020. Air quality changes during the COVID-19 lockdown over the Yangtze River Delta Region: an insight into the impact of human activity pattern changes on air pollution variation. *Sci. Total Environ.* 732 <https://doi.org/10.1016/j.scitotenv.2020.139282>.
- Liu, F., Page, A., Strode, S.A., Yoshida, Y., Choi, S., Zheng, B., Lamsal, L.N., Li, C., Krotkov, N.A., Eskes, H., Ronald van der, A., Veeffkind, P., Levelt, P.F., Hauser, O.P., Joiner, J., 2020. Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. *Sci. Adv.* 6, 2–7. <https://doi.org/10.1126/sciadv.abc2992>.
- Lovrić, M., Fadljević, L., Kern, R., Steck, T., Gerdenitsch, J., Peche, E., 2020a. Prediction of anode lifetime in electro galvanizing lines by big data analysis. In: GALVA-TECH 2020. Vienna.
- Lovrić, M., Pavlović, K., Kern, R., Grange, S.K., Vuković, M., Haberl, M., 2020b. Air Pollution 01.2014 - 05.2020 (Including COVID-19 Lockdown) Data from Graz. Austria. <https://doi.org/10.5281/zenodo.3982670>.
- Lovrić, M., Pavlović, K., Žuvela, P., Spataru, A., Lucić, B., Kern, R., Wong, M.W., 2020c. Machine learning in prediction of intrinsic aqueous solubility of drug-like compounds: generalization, complexity or predictive ability? *chemrxiv*. <https://doi.org/10.26434/chemrxiv.12746948>.
- Mahato, S., Pal, S., Ghosh, K.G., 2020. Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *Sci. Total Environ.* 730, 139086 <https://doi.org/10.1016/j.scitotenv.2020.139086>.
- McKee, M., Stuckler, D., 2020. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0863-y>.
- Menut, L., Bessagnet, B., Siour, G., Mailler, S., Pennel, R., Cholakian, A., 2020. Impact of lockdown measures to combat Covid-19 on air quality over western Europe. *Sci. Total Environ.* 741, 140426. <https://doi.org/10.1016/j.scitotenv.2020.140426>.
- Moser, F., Kleb, U., Katz, H., 2019. Statistische Analyse der Luftqualität Graz anhand von Feinstaub und Stickstoffdioxid. Graz.
- Rodríguez-Urrego, D., Rodríguez-Urrego, L., 2020. Air quality during the COVID-19: PM2.5 analysis in the 50 most polluted capital cities in the world. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2020.115042>.
- Schiermeier, Q., 2020. Why pollution is plummeting in some cities - but not others. *Nature*. <https://doi.org/10.1038/d41586-020-01049-6>.
- Shi, X., Brasseur, G.P., 2020. The response in air quality to the reduction of Chinese economic activities during the COVID-19 outbreak. *Geophys. Res. Lett.* 47, 1–8. <https://doi.org/10.1029/2020GL088070>.
- Šimić, I., Lovrić, M., Godec, R., Kröll, M., Bešlić, I., 2020. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. *Environ. Pollut.* 263, 114587 <https://doi.org/10.1016/j.envpol.2020.114587>.
- Venter, Z.S., Aunan, K., Chowdhury, S., Lelieveld, J., 2020. COVID-19 lockdowns cause global air pollution declines. *Proc. Natl. Acad. Sci. U. S. A.* 117, 18984–18990. <https://doi.org/10.1073/pnas.2006853117>.
- Viana, M., Kuhlbusch, T.A.J., Querol, X., Alastuey, A., Harrison, R.M., Hopke, P.K., Winiwarter, W., Vallius, M., Szidat, S., Prévôt, A.S.H., Hueglin, C., Bloemen, H., Wählin, P., Vecchi, R., Miranda, A.I., Kasper-Giebl, A., Maenhaut, W., Hitenberger, R., 2008. Source apportionment of particulate matter in Europe: a review of methods and results. *J. Aerosol Sci.* <https://doi.org/10.1016/j.jaerosci.2008.05.007>.
- Wang, Y., Yuan, Y., Wang, Q., Liu, C.G., Zhi, Q., Cao, J., 2020. Changes in air quality related to the control of coronavirus in China: implications for traffic and industrial emissions. *Sci. Total Environ.* 731, 139133. <https://doi.org/10.1016/j.scitotenv.2020.139133>.
- Zhao, N., Wang, G., Li, G., Lang, J., Zhang, H., 2020. Air pollution episodes during the COVID-19 outbreak in the Beijing–Tianjin–Hebei region of China: an insight into the transport pathways and source distribution. *Environ. Pollut.* 267, 115617 <https://doi.org/10.1016/j.envpol.2020.115617>.
- Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. *Sci. Total Environ.* 727, 138704 <https://doi.org/10.1016/j.scitotenv.2020.138704>.