# Assessment of the Impact of the Covid-19 Lockdown on Air Pollution over Spain using Machine Learning

Hervé Petetin[*1], Dene Bowdalo[*2], Albert Soret[*3]

[*] *Barcelona Supercomputing Center (BSC)*

{[1]herve.petetin, [2]dene.bowdalo@bsc.es, [3]albert.soret}@bsc.es

***Keywords*— air quality, machine learning, covid-19**

## EXTENDED ABSTRACT

The rapid spread of the Covid-19 pandemic over Spain recently forced the Spanish authorities to implement drastic measures of social distancing through a strict lockdown of the population starting on March 14th. As hospitalizations were still strongly increasing, a second and more stringent phase of the lockdown was implemented from the March 30th to April 9th with workers from all non-essential economical activities forced to stay at home.

This situation has impacted numerous activity sectors, including road transport, air traffic and part of the industries. As a consequence, air pollutant emissions have been greatly reduced. Although such a large change of emission forcing is expected to reduce the air pollutant concentrations in Spanish urban areas, the extent of such reductions remains uncertain. Key to this uncertainty are the highly variable meteorological conditions that can either attenuate or amplify changes of air pollution concentration originally driven by changes of emissions. Thus, assessing the impact of the Covid-19 lockdown solely based on the analysis of the concentration time series can often be misleading since at least part of the variability is driven by the meteorology.

In this study, we explore the use of machine learning algorithms for estimating the business-as-usual $NO_2$ concentrations that would have been observed without the Covid-19 lockdown based on ERA5 meteorological data and additional time features. Trained on past data, these ML models can learn the complex relationships between meteorology and NO2 concentrations, indirectly assuming an average emission forcing. By using these ML models to predict the NO2 concentrations under the current situation (with very different emission forcing), we expect the discrepancies between predictions and observations to be related to a large extent to the reduction of emissions induced by the lockdown regardless of the meteorological conditions.

In this study, the reduction of $NO_2$ pollution is investigated in all 50 provinces of Spain.

### *A. Data*

We selected one air quality (AQ) surface monitoring stations in each of the 50 Spanish provinces (Figure 1). Stations are selected according to several criteria of interest, including high population density and high data availability. $NO_2$ measurements are extracted from the BSC in-house GHOST (Globally Harmonized Observational Surface Treatment) project (Bowdalo et al., in preparation). GHOST gathers numerous publicly available atmospheric datasets together with harmonized detailed metadata and quality assurance (QA) flags. It aims at facilitating a greater quality of data analysis in the atmospheric chemistry community. In this study, we used the EEA AQ eReporting dataset, and applied a careful QA screening to the entire $NO_2$ dataset in order to remove unrealistic or very suspicious values.

Meteorological data are taken from ERA5 reanalysis dataset, whose spatial resolution is 31 km globally. We consider the following parameters at the daily scale: daily mean, minimum and maximum 2-m temperature, surface wind speed, 10-m zonal and meridian wind speed components normalized by 10-m wind speed (in order to provide information solely on wind direction), surface solar radiation downward, surface pressure and total cloud cover.

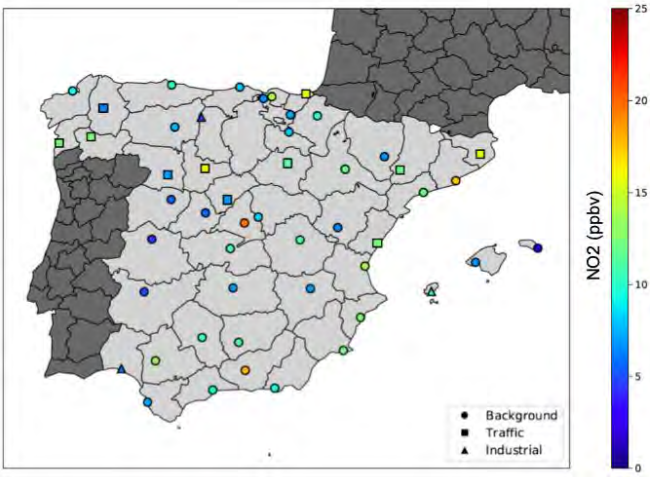


Fig. 1. Climatological (01/01/2013-06/04/2020) $NO_2$ mixing ratio at the stations selected in each of the 50 provinces of Spain (Canarias islands not shown).

### *B. Methodology*

In this study, we used the Gradient Boosting Machine (G) machine learning model. One ML model is developed at each of the 50 stations. The target is the daily mean $NO_2$ mixing ratio. Following the studies of Grange et al. (2018) and Grange and Carslaw (2019), we considered a set of features including the previously mentioned meteorological variables and several time features, namely the Julian date (as the seasonal term) and the weekday.

The complexity of using ML models for predicting pollutant concentrations in this case relies in its typical non-stationarity, as emissions often following trends due for example to the implementation of more restrictive regulation. In order to limit this potential issue, we trained the ML models only over the 3 years from 2017 to 2019, assuming that trends over that short period are relatively low compared to the variability induced by the meteorology, and let the year 2020 for testing. In this way, the performance of the ML model can be assessed looking at the results in 2020 before the lockdown (not after since emission forcing is different). To estimate the uncertainty of our ML models, we replicated our approach over the past years: training over 2013-2015 and calculation of the uncertainty over January-April 2016; then idem for 2014-2016, 2015-2017 and 2016-2018.

### *C. Results*

Preliminary results indicate that the performance of the ML models is quite heterogeneous among the different regions. In many regions, the discrepancies with observed concentrations (before lockdown) remain within the uncertainty range. However, some bias is often highlight, in some cases larger than the uncertainty. Discrepancies may be due to deficiencies in the ML model and/or changes of emissions in the vicinity of the station. However, in most regions, the predictions followed reasonably well the variability of the observed concentrations, which suggests that the ML models are capturing well the influence of the meteorological conditions.

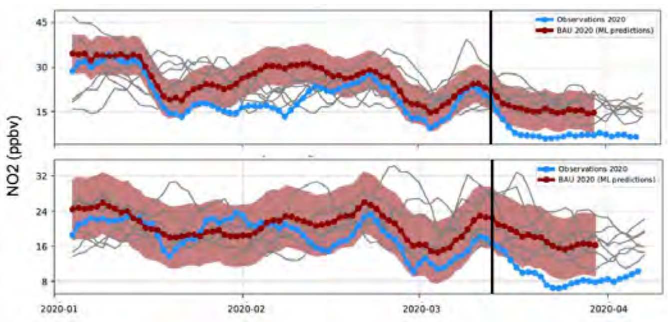


Fig. 2 Seven-days running average of $NO_2$ mixing ratios at Madrid (top panel) and Barcelona (bottom panel) from January to present day. Observed concentrations are shown in blue, predicted business-as-usual (BAU) $NO_2$ concentrations are shown in red, with a 95% confidence interval. The $NO_2$ concentrations from all previous years since 2013 are shown in grey. The lockdown period starts after the vertical black bar. Note that ML predictions were not possible after late March due to not yet available ERA5 data.

First results in Madrid and Barcelona regions are shown in Figure 2. In Madrid, before lockdown, the ML model performs quite well but with a moderate positive bias. It is not able to capture all the events, and the errors found during the first weeks of February are found to exceed the uncertainty range estimated from previous years. Including ERA5 information on the planetary boundary layer height and/or atmospheric stability may help solving this issue. After the lockdown, observed $NO_2$ mixing ratios quickly decrease, down to values around 8 ppbv, that have never been observed at this period of the year since 2013. The ML model predicts business-as-usual $NO_2$ mixing ratios of about 15 ppbv over that period. The difference with observed concentrations is found to be statistically significant at a 95% confidence level, with a best estimate of -47%.

Similar results are obtained in Barcelona where the ML model performs better as the observations are within the predicted uncertainty range before lockdown. The observed $NO_2$ concentrations dropped down to 8 ppbv while ML predictions exceed 16 ppbv, which means a reduction of 50%, in agreement with what is observed in Madrid.

In some other regions, results are more complex to interpret given the potential deficiencies of the ML models. Discrepancies between observations and business-as-usual predictions during the lockdown often remain within the uncertainty interval of the ML model, which prevents from concluding on statistically significant reductions of $NO_2$ pollution. Including more recent data may help to detect more significant changes.

### *D. Conclusion and Perspectives*

Further efforts are obviously required for analyzing the results obtained in the different regions of Spain but these preliminary results suggest that the methodology is suitable, at least in some regions, for identifying and quantifying the reduction of $NO_2$ pollution that is not due to meteorological conditions. It is worth noting that part of the discrepancies previously highlighted may be due not only to changes in the emissions but also to changes in the chemistry, although we assume this impact is lower.

Processing the most recent data, during the second phase of the lockdown, should soon provide us a more complete picture of the impact of these reduced emissions on the $NO_2$ levels in Spanish urban areas. Including additional ERA5 variables may also help to improve the skills of the ML models, which is required for confirming the statistical significance of $NO_2$ reduction in some regions.

### *E. ACKNOWLEDGEMENTS*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement H2020-MSCA-COFUND-2016-754433.

### *References*

[1] Grange, S. K. and Carslaw, D. C.: Using meteorological normalisation to detect interventions in air quality time series, Science140of The Total Environment, 653, 578–588, doi:10.1016/j.scitotenv.2018.10.344, 2019.

[2] Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models forSwiss PM 10 trend analysis, Atmospheric Chemistry and Physics, 18, 6223–6239, doi:10.5194/acp-18-6223-2018, 2018

## *Author biography*

**Hervé Petetin** was born in Caen, France in 1986. He holds an engineering diploma from the Ecole Centrale de Lille, France, a M.Sc. in Mechanics and fluid dynamics from the University of Science and Technology Lille, a M.Sc. in Atmospheric physics and chemistry from the University of Paris Est Creteil, France, and a Ph.D. in Atmospheric physics and chemistry from the University of Paris Diderot, France.

His research has first focused on the study of air quality and more specifically the aerosol pollution in large megacities, using regional chemistry-transport models and in-situ observations. As a member of the IAGOS European Research Infrastructure, he then investigated during several years the variability and trends of two important intermediate-lifetime gaseous compounds, namely ozone and carbon monoxide, in the troposphere based on airborne in-situ observations provided by the IAGOS fleet.

In 2018, he obtained a postdoctoral funding at the BSC from the STARS program (Marie-Sklodowska-Curie Action

COFUND program) for exploring the use of machine learning in air quality (including the correction of air quality forecasts).