BAYESIAN MACHINE LEARNING TO QUANTIFY IMPACTS OF COVID-19 LOCKDOWNS ON URBAN AIR QUALITY

Anonymous authors
Paper under double-blind review

ABSTRACT

The COVID-19 pandemic-induced lockdowns in 2020 caused a sharp decline in urban economic activity. This unintended experiment provides a unique opportunity to derive insights into the inter-dynamics of our environmental and human systems. Emerging literature indicates a decline in certain pollutant levels during this period. However, most of the methods adopted are simplistic, relying on historical averages as baselines without accounting for proper confounding variables such as meteorological factors. In this work, we present a Bayesian machine learning approach to provide better point estimates along with uncertainty bounds for the change in pollutant levels during COVID-19 lockdowns. Using New York City and Los Angeles data for a pilot study, we find a 21.69% [22.87%] decline in nitrogen dioxide (NO_2) and an increase of11.14% [4.15%] in ground-level ozone (O_3) levels in New York and [Los Angeles]. Further analysis with mobility data as a proxy for economic activity strengthens the role of COVID-19 lockdowns in improved urban air quality. The findings suggest the potential for mitigating environmental risks and improving urban air quality by rationalizing emissions due to economic activity.

1 INTRODUCTION

Since the beginning of the industrial revolution, humanity has experienced unprecedented economic growth. However, this prosperity has come at a high cost to the environment. For example, the concentration of greenhouse gases (GHGs) such as CO2 has increased by more than 50% (Olivier et al. (2017)) in the last 200 years, causing widespread global warming. The primary source of GHGs includes industry, transportation, and power generation, among others. While completely turning back on GHG causing economic activities is infeasible, there is an imminent need to balance. However, quantifying the tradeoff between economic activities and environmental quality is challenging in a business-as-usual case. In this regard, the present COVID-19 pandemic presents a unique opportunity to directly measure the impact on the environment, particularly air quality (AQ), when the knobs of the global economic engine have been turned off. Accessibility to a suite of data products ranging from ground sensors to remotely sensed data makes it further possible. Indeed AQ data from a globally dense network of ground-based sensors are present in near-real-time.

Ever since the first COVID-19 related lockdowns were announced in Wuhan, China, on 23rd January 2020, there is fast-emerging literature on quantifying the impacts of COVID-19 lockdowns globally. The early works, mostly focused on Chinese cities (Lian et al. (2020), He et al. (2020)), could only use a few months of data. As the pandemic spread, more studies were conducted in different parts of the world. However, upon review, most of the methods adopted are simplistic, usually relying on historical averages (Nakada & Urban (2020), Metya et al. (2020)) of pollutant levels as baselines to compare the effects of COVID-19 lockdowns. Other methods employing regression-based analysis (Venter et al. (2020)) have used methods ranging from linear regression to sophisticated neural networks. While linear regression might fail to capture the complex mapping between weather and AQ adequately, neural networks are prone to overfitting in light of limited data. Further, none of the methods provides uncertainty estimates. Therefore, in this work, a case is made for Gaussian Processes (GP) (Rasmussen (2003)), a Bayesian nonparametric approach providing uncertainty estimates (Ghanem et al. (2017)) in a principled, data-efficient, and tractable manner. Moreover, it is

robust to overfitting due to its Bayesian formalism. Linear regression method is used as a baseline to compare the performance metrics of GP regression.

Local spatiotemporal weather patterns can strongly affect ground-level pollutant aggregates (Zhang et al. (2017)). Using the previous year means to calculate AQ changes in 2020 does not account for meteorological variability that may have confounded any observable effect of COVID-19 lockdowns. In this work, we first create a GP regression model mapping the relationship between daily weather and pollutant levels. The regression model is then used to quantify the impact of COVID-19 lockdown on urban AQ (see section 2.1 Problem Formulation). Additionally, the COVID-19 Community Mobility Reports made available by Google Inc. allow us to ascertain how much of the change in pollutant levels coincides with the shift in mobility patterns in urban areas due to COVID-19. For this study, we have used data for two urban economic epicenters in the US - New York City (NYC) and Los Angeles (LA) - which were among the first regions affected by COVID-19 in the US. Consequently, NYC went into a lockdown on 20th March 2020, and a health emergency was declared in Los Angeles on 3rd March 2020, with strict restrictions.

2 METHODS AND DATA

2.1 PROBLEM FORMULATION

A Gaussian Processes (GP) regression model $f : \mathbf{X} \rightarrow \mathbf{Y}$ is learned from the input feature map $\mathbf{X} = \{(t, p, h, w)\}_{i=1}^{N} \in R^4$, to the output feature map $\mathbf{Y} = \{(q)\}_{i=1}^{N} \in R$. The symbols t, p, h, w are the daily meteorological variables, namely mean temperature, precipitation, relative humidity and wind speed while q corresponds to the daily NO_2 or O_3 values. The number N refers to the number of daily datapoints. The regression model is used to calculate what the pollutant levels would have been in 2020. The impact of COVID-19 lockdowns is then defined as the difference between the observed and weather-modeled pollutant levels in 2020.

2.2 GAUSSIAN PROCESSES

Gaussian Processes (GP) is a Bayesian machine learning framework that has been widely used for regression (MacKay (1998)) and classification tasks. For a detailed mathematical description, the reader is directed to (Rasmussen (2003)). Briefly, for a given set of training data points, there are potentially infinite functions that can fit the data. GPs offer a mathematically tractable solution to this problem by assigning a probability to each of these functions. The mean of this probability distribution represents the most probable (point) estimates. Furthermore, the probabilistic approach naturally lends to uncertainty (confidence) bounds around the point estimate during inference.

2.3 DATASET

Air Quality (AQ): ground-level sensor data for NO_2 and O_3 is obtained from the quality-controlled daily AQ dataset made available by the US Environmental Protection Agency (EPA) for the period 2015-2020.

Weather: the daily weather data (2015-2020) is obtained from the Local Climatalogical Data (LCD) product from the National Oceanic and Atmospheric Administration. The LCD product integrates data from over 10,000 ground-sensors in the US. For this study, ground-level sensor measurements are preferred over remotely sensed data because they are more sensitive to emission changes at the ground-level.

Mobility: the COVID-19 Community Mobility Reports made available by Google are used to determine in the change in mobility during the lockdown period.

2.4 DATA PREPROCESSING

The daily AQ and weather data is first combed for missing values. Second, a 7-day rolling mean is applied to reduce the high-frequency noise in the time-series data. The daily AQ and weather data is the merged by matching on dates present in both to obtain the final dataset. The data from

2015-2019 is used for training while data for 2020 is used for inference. The GP regression model is trained separately for NYC and LA.

3 RESULTS

The main findings are in shown in Figures 1 and 2. In Figure 1, the GP regression-modeled 2020 levels for NO_2 and O_3 are plotted against the observed (actual) levels. The GP modeled levels represent the weather-benchmarked pollutant levels that could have been in 2020 (in the absence of COVID-19 lockdowns). The difference between the two trajectories is the impact of COVID-19 lockdowns on pollutant levels, plotted in Figure 2. The lockdowns in NYC were in effect between 20th March and 30th September 2020. During this period, NO_2 levels were lower on an average by 21.69% ($\pm 6.02\%$), and O_3 levels were up 11.14% ($\pm 7.35\%$). For the same period in LA, the change in NO_2 and O_3 levels was -22.87% ($\pm 5.96\%$) and $+4.15\%$ ($\pm 2.71\%$). The increase in ground-level O_3 levels despite lockdowns can be explained due to the nonlinear relationship between nitrogen oxides (NO_x) and ozone. High levels of nitrogen oxides react with O_3 and remove it from the atmosphere. In urban areas where NO_x levels are generally high, lowering NO_2 levels may cause O_3 levels to increase. During this period, the change in public transportation traffic (Figure 2) was -66.47% (NYC) and -42.70% (LA).

Interestingly, between 1st October and 31st December 2020, when public transportation traffic was still down by more than 40%, the NO_2 levels were down by only 5.23% in NYC and increased by 7.71% in LA. On the other hand, the O_3 levels in LA increased by 38.78% during this period.

3.1 EFFECTIVENESS OF THE METHOD

The superior performance of GP regression is based on the assumption that it can provide better estimates of pollutant levels in 2020 (in the absence of COVID-19) by accurately mapping the relationship between meteorological variables and pollutant levels. To ascertain this, we conduct an experiment where we train the GP model on data from 2015-2018 and test it on the data from 2019. We compare the error metrics with simple linear regression and mean values from 2015-2018. Lower error metrics indicate the GP model can estimate more accurately the pollutant levels given the weather data. The combined metrics for both cities and pollutant levels are presented in Table 1.

4 IMPLICATIONS

Using Gaussian Processes, we demonstrate a more accurate approach to quantify the impacts of COVID-19 lockdowns on urban air quality. We find empirical evidence that COVID-10 lockdowns resulted in lower NO2 levels in two of the most urbanized regions in the US while O3 levels went up. Although keeping economic activities down to lockdown levels might not be feasible, there is merit in pursuing mitigation policies in this direction post-COVID-19, as underlined by the finding here. At the same time, the relationship between reduced transportation and urban air quality is not so straightforward as seen during the period from October to December 2020 when the NO2 levels went up despite public mobility still down more than 40%. Further, there is a need to consider the intertwined atmospheric chemistry where a decline in one pollutant's levels may cause an upshift in another pollutant, such as in the case of NO_2 and O_3.

REFERENCES

Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.

Guojun He, Yuhang Pan, and Takanao Tanaka. The short-term impacts of covid-19 lockdown on urban air pollution in china. *Nature Sustainability*, 3(12):1005–1011, 2020.

Xinbo Lian, Jianping Huang, Rujin Huang, Chuwei Liu, Lina Wang, and Tinghan Zhang. Impact of city lockdown on the air quality of covid-19-hit of wuhan city. *Science of the Total Environment*, 742:140556, 2020.

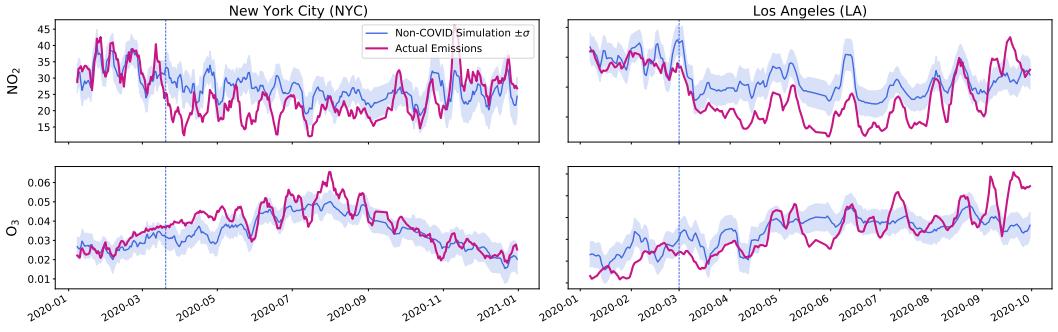

Figure 1: Comparing actual NO_2 and O_3 concentration vs estimates (with uncertainty bounds) as predicted by Gaussian Processes regression based on meteorological factors. The NO_2 concentrations show a significant reduction for both NYC and LA during March-September 2020 when lockdowns were in place. The O_3 concentration show a marginal increase for both cities due to lower NO_2 levels which inhibit ground-level ozone concentration.

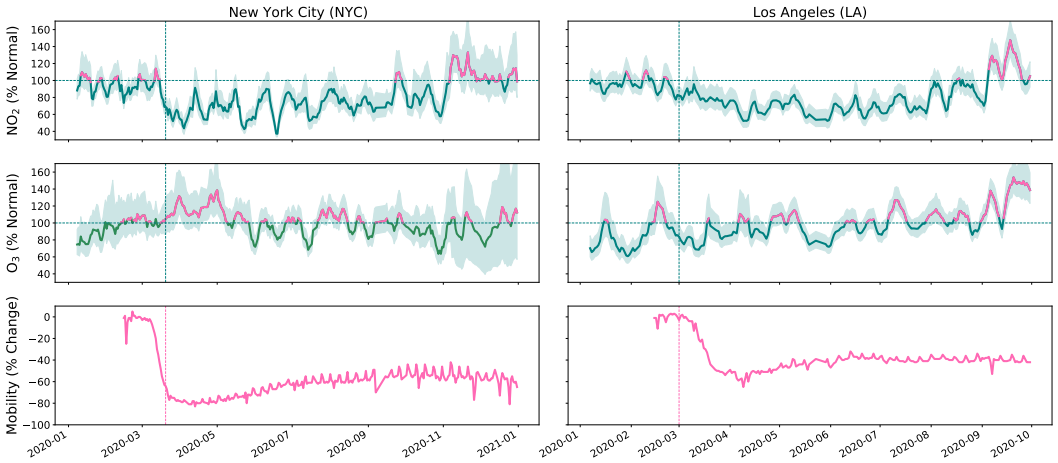

Figure 2: Change in nitrogen dioxide (NO_2) and ground-level ozone (O_3) concentration due to COVID-19 induced lockdowns compared to a non-COVID baseline for 2020. The horizontal line indicates the baseline and the vertical line points to the dates 03-20-20 (New York- NYC) and 03-02-20 (Los Angeles- LA) when a lockdown and emergency was declared in NYC and LA, respectively.

Table 1: Comparing Combined Performance Metrics

Method	**RMSE**	**MAE**	**R^2**	**Pearson CC.**
GP Regression	**4.06**	**3.19**	**0.58**	**0.77**
Linear Regression	5.31	5.47	0.34	0.48
Historic Means	8.56	8.09	0.19	0.33

David JC MacKay. Introduction to gaussian processes. *NATO ASI series F computer and systems sciences*, 168:133–166, 1998.

Abirlal Metya, Panini Dagupta, Santanu Halder, Supriyo Chakraborty, Yogesh K Tiwari, et al. Covid-19 lockdowns improve air quality in the south-east asian regions, as seen by the remote sensing satellites. *Aerosol and Air Quality Research*, 20(8):1772–1782, 2020.

Liane Yuri Kondo Nakada and Rodrigo Custodio Urban. Covid-19 pandemic: Impacts on the air quality during the partial lockdown in são paulo state, brazil. *Science of the Total Environment*, 730:139087, 2020.

Jos GJ Olivier, Jeroen AHW Peters, et al. *Trends in global CO2 and total greenhouse gas emissions: 2017 report*. PBL Netherlands Environmental Assessment Agency The Hague, 2017.

Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pp. 63–71. Springer, 2003.

Zander S Venter, Kristin Aunan, Sourangsu Chowdhury, and Jos Lelieveld. Covid-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences*, 117(32): 18984–18990, 2020.

Henian Zhang, Yuhang Wang, Tae-Won Park, and Yi Deng. Quantifying the relationship between extreme air pollution events and extreme weather events. *Atmospheric Research*, 188:64–79, 2017.