

Comment on **essd-2021-318**

Anonymous Referee #1

Referee comment on "A machine learning approach to address air quality changes during the COVID-19 lockdown in Buenos Aires, Argentina" by Melisa Diaz Resquin et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-318-RC1>, 2021

This study was majorly inspired by Grange et al 2019 to use a random forest model (RF) to train, validate and predict the air quality concentration in a megacity of Argentina. Although the methodology is not new, this study has the potential to have a significant improvement to fill the data gap given the reason that some monitored trace gas concentrations were lacking but the pollutants are becoming a concern for local authorities. (line 145- 150) I would like to propose a few major revisions for the authors of this manuscript. After done with that improvement, I think it would be strong enough to publish on earth system science data:

■

Random forest models indeed are easy and quick to train. But if this study is only focusing on predicting the time series of pollution during COVID-19, there are better and more efficient machine learning models such as ARIMA compared with random forest. The key reason for many studies that authors mentioned in this article used RF is because it could provide key components reflecting the non-linear relationship among emissions, chemical reaction, and meteorological effects. Please see figure 3 in Grange et al 2019 and figure 1 in Yang et al. It is easy to generate that Gini importance either through python or R. Since this code the authors used were based on R, I would suggest they follow the code from Grange et al 2019 to generate the Gini importance plot. After getting those plots, you can compare your RF results with the reasons from previous literature to see if it makes sense.

■

The authors also mentioned several times the meteorological impacts on air pollution. The earliest signal has been seen during COVID was the study in China from Le et al, 2020 which authors may consider mentioning this study result. Then it would also be beneficial to consider using RF models to generate new predictions by normalizing meteorological factors. This would give the third line in each panel of figure 3. You can do this by following the methodology in figure 5 of Grange et al 2019 and figure 2 of Yang et al 2021. Vu et al 2019 made an additional improvement to weather normalization which you may also consider using this methodology.

■

The description of details of why and how to interpret bivariate polar plots here is vague. The reason that Grange 2019 used this bivariate polar plot is that they wanted to prove that wind influence the dispersion of pollutants. Therefore it's better to show the meteorological impacts from the above suggestions that I mentioned first. Then applying a bivariate polar plot by combining with winds components if the meteorological impact is the dominant factor here.

■

For the relationship between NO₂ and diesel please refer to Yang et al 2021. You can also consider normalizing other anthropogenic factors besides meteorology based on the result from the first suggestion.

■

Please indicate how much data by # not percentage is used for training and how much data is used for validation/prediction. Due to some restrictions of the monitoring campaign, please indicate how the authors dealt with inadequate data for specific variables.

The following paper should have been referenced and discussed in the manuscript:

Grange, Stuart K., and David C. Carslaw. "Using meteorological normalisation to detect interventions in air quality time series." *Science of the Total Environment* 653 (2019): 578-588.

Yang, Jiani, Yifan Wen, Yuan Wang, Shaojun Zhang, Joseph P. Pinto, Elyse A. Pennington, Zhou Wang et al. "From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model." *Proceedings of the National Academy of Sciences* 118, no. 26 (2021).

Vu, Tuan V., Zongbo Shi, Jing Cheng, Qiang Zhang, Kebin He, Shuxiao Wang, and Roy M. Harrison. "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique." *Atmospheric Chemistry and Physics* 19, no. 17 (2019): 11303-11314.

Le, Tianhao, Yuan Wang, Lang Liu, Jiani Yang, Yuk L. Yung, Guohui Li, and John H. Seinfeld. "Unexpected air pollution with marked emission reductions during the COVID-19

outbreak in China." *Science* 369, no. 6504 (2020): 702-706.

The clarity and context need significant improvement to better draw out why the results are significant.