

## Comment on **essd-2021-318**

Anonymous Referee #2

---

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-RC2>, 2022

---

The manuscript is generally well written and clearly presented. However, its research outcome (i.e the impact of meteorology and regional sources on air quality in Buenos Aires, Argentina) is not new. It should investigate the interactions between input variables to understand more about Random forest model. I do recommend publishing this work if the authors can solve my major concerns as below:

### Major concerns:

#### 1. Selection of explanatory variables:

1.1. Table 3, line 200-205: Air quality strongly depends upon boundary layer height and long-range transport. Why were these variables not included in this study as input variables in the model? Please refer to a reference by Shi et al. 2021 (Science Advance, Vol 7, Issue 3, "Arupt but smaller than expected changes in surface air quality attributed to Covid-19 lockdowns").

1.2. Could the author explain why the cho CO, NO as explanatory variables for NO<sub>2</sub>? CO and NO were modelled from t<sub>2</sub>, rh<sub>2</sub>, U, V and gasoline diurnal patterns, so I guess the author also can model NO<sub>2</sub> based on these variables. Similar questions for explanatory variables for SO<sub>2</sub> and PM<sub>10</sub>, and O<sub>3</sub>.

1.3. In the model of NO<sub>2</sub>, did the author investigate interactions between input variables such as NO with t<sub>2</sub>.

1.4. In terms of O<sub>3</sub>, it strongly depends upon atmospheric temperature. Why does this variable not be included in your model?

## **2. Testing dataset:**

2.1. Figure 2, Line 189: What criteria do authors select testing dataset based on (i.e 2 weeks data before lockdown?)

2.2. In my opinion, the 2-weeks data for testing data sets is too short. Therefore, authors should do a model performance for at least one month before and after the lockdown/partial periods.

## **Minor comments:**

3.1. Table 4: I think author should include the r value between model and observation rather the r-value for diurnal cycle (r-dc)

3.2. Table 5: In BLD, it should include concentrations of pollutants between observation and model.

3.3. In discussion: Authors should plot the dependence of concentration of pollutants on meteorological conditions.