

Interactive comment on “Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning” by Roland Stirnberg et al.

Anonymous Referee #3

Received and published: 16 September 2020

This paper presents a machine-learning built model approach to analyse an extensive multi-parameter dataset at observational in a suburban area south of Paris. The focus of the manuscript is using a recently published tool (“SHapley Additive exPlanation (SHAP) values”) to analyse the machine-learning model’s predictions and then attribute drives of the statistical model.

The paper presents large amounts of information about the output from the analysis tool, but not enough focused justification or evidence is presented about how novel these interpretations are or how that they could be used for air pollution mitigation policy etc. At points, the paper even reads as if the authors are suggesting that authorities

C1

seek to mitigation against the meteorology contribution to air pollution. Could this analysis be used to make a forecasting tool if parameters were gained in real-time? If so, how long ahead would these predictions be expected to be useful for? Would this be useful in a public health context?

A core premise (in the abstract and elsewhere) is that we do not fully understand the contribution of meteorology to high air pollution episodes is true, however, this does justify the framework used here which omits two other key drivers (chemistry and emissions). Apart from a few mentions, it is not clear how are these contributions and considered in this method. Are the contributions of these processes just assumed to be part of the meteorological contributions? This needs to be a lot clearer.

The paper seems mostly focused on exploring the “SHapley Additive exPlanation (SHAP) values” approach and it is unclear whether a novel contribution has been made to the field of air pollution research. This paper may be better suited to a machine learning journal or could be re-write to be more focused on air pollution. Either of these two options would require large changes to the current manuscript.

Specific comments

Why has PM1 been the focus of this study, rather than the more health-relevant PM2.5 species? Also, how did the model perform at predicting PM10? Considering the omission of chemistry and emissions in this study, would PM10 or PM2.5 be a better candidate for study?

Line 21 - “Processes vary even within seasons”

This does not read well. Of course, processes will vary within seasons.

Line 24 - “likely causes an increase in local wood-burning emissions”

Cause and effect seem to be muddled. Maybe the authors mean to say increases in burning emission could explain increased particulates?

C2

line 25 - "The application of SHAP regression values within a machine learning framework presents a novel and promising way of analysing observational data sets in environmental sciences."

Are there implications for what we should focus on meteorology studies or observations on? What about the implications for air-quality modelling or policy? Just presenting another tool that can be used is not a notable contribution.

Line 90 - How can policymakers use this information? Improve air quality models? Focus research directions? What about it is new?

Line 90 - Why not focus on the SIRTA region, rather than Paris, which is in completely different chemistry and emissions regime? The reader needs to be convinced that the site is representative of the Paris region.

Technical comments

Please use sub/superscripts for chemical species throughout (e.g. SO₄²⁻, SO₂, PM_{2.5}).

Expand acronyms in sub-header titles (e.g. MLH).

Expand acronyms once per major section too.

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-469>, 2020.