Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2020-469-AC3, 2020

Atmospheric Chemistry and Physics Discussions

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

Roland Stirnberg et al.

roland.stirnberg@hotmail.com

Received and published: 30 November 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 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?"

Answer:

Thank you for your assessment.

The focus of this study is not on the prediction of pollutant concentrations in time, but to contribute to the advancement of the scientific understanding of how meteorology influences air pollution. The machine-learning framework presented in this study provides observation-based, quantitative estimations for the influence of various meteorological factors to PM1 at the same time, enabling their direct comparison. The model does allow for interactions between the meteorological factors, and on this basis, a separation and comparison of meteorological influences on any individual event is feasible. This is a novel aspect, as it allows to extract empirical patterns from the data set that are hard to detect using established statistical methods.

Setting up a forecasting tool is a possible extension of the machine-learning framework established within this study, but not the key objective here. This is why we only outline such possible applications and their usefulness at the end of the manuscript. So no, our analysis framework in its present form is not intended as a forecasting tool, and cannot be converted into one without more work. Hence, the reliability of such a forecast tool was not assessed. It is likely that the PM forecast would greatly depend on the reliability of the forecasted meteorological conditions. In its present configuration, however, our tool can determine an 'expected' level of air pollution under given meteorological conditions. By comparing this to actual observations, the effect of any source reductions (e.g. via policies) can be assessed. These points were added in L510-520.

The following specific changes were made in the manuscript:

- L2 & 3: “substantially contribute to” was changed to “substantially influence”. The wording “contribute to” might indeed be misleading here, as it could sound as if meteorology actively emits pollutants.

- Throughout the manuscript, the wording “meteorological contribution” was changed to “meteorological influence” or removed, if not referring to the ML model (caption chapters 4.2, 4.2.1-4.2.4, L265, L294, L295, 295, 303, 317, 329, 361, 362, 376, 424, 461, 483, 486)

-L46: the sentence “It is therefore crucial to take atmospheric and environmental processes into account during the development of efficient pollution mitigation strategies” was removed. This point is now made clearer at the end of the introduction. See changes in lines 80-85; the goal of the study is now stated more precisely and benefits in a public health context are described

-L476-479: ...As interactions between the meteorological variables are accounted for, the model enables the separation, quantification and comparison of their respective impacts the individual events. It is shown that ambient meteorology can substantially exacerbate air pollution. Results of this study point to a distinguished role of shallow MLHs, low temperatures and low wind speeds during peak PM1 concentrations in winter

-L512-515: changed to “For policy makers, the presented approach could prove beneficial in multiple ways and serve as a decision aid for air policy measures. Preventative warnings could be issued to the public if the identified meteorological conditions exacerbating air pollution are to be expected. Another application would be to attribute changes in air quality to policy measures by comparing an ‘expected’ level of air pollution under given meteorological conditions to actual observations (e.g., Cermak2009 and Knutti 2009), which may help. . .”

"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 emis-sions). Apart from a few mentions, it is not clear how are these contributions and con-sidered 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."

Answer:

The focus of this paper explicitly lies on the analysis of the influence of meteorological conditions on PM1 concentrations. We are fully aware that meteorology alone cannot explain PM1; one of our aims is to ultimately be able to ‘remove’ the effect of meteorology, and retain the effects of emissions (and to a lesser degree, chemistry), which to some extent can be influenced directly by policy (this was added in L99-102). As mentioned in other answers above, pollutant concentrations have been shown to be exacerbated or decreased by certain meteorological conditions (e.g., Dupont et al., 2016). It is shown that the model is able to capture a large fraction of the occurring variation of daily PM1 concentrations, which shows that the variables chosen as inputs are indeed important drivers. Even without explicitly considering emissions and chemistry, the model explains between 50-60% of the day-to-day PM1 variability. Thus, for the location and data set analysed here, the influence of meteorological variability on PM1 is at least as large as the influence of the variability of emissions and chemistry. Hence, given the key objective of this study, the presented framework is suitable for the analysis by capturing key meteorology-based processes. The detailed analysis presented in chapter 4.4 emphasizes that the temporal trends of PM1 concentrations are largely well captured.

Some of the meteorological parameters inherently contain information on chemistry and emissions. For example, RH, solar radiation, and temperature can influence local transformation processes, as detailed in L44-60. Temperature also contains inherent information on the strength of residential heating (L250). Wind direction indicates whether clean air from the west or more polluted air from the northeast is influencing the PM1 measurement. These mechanisms are mentioned in the introduction (L42-59)

and the result section (chapter 4.2).

To convey these points more clearly to the reader, the following changes were made:

- L85: added “atmospheric”, changed “determining” to “influencing” - L60: Added “...while moisture in the atmosphere can stimulate secondary particle formation processes. . .” - L136-141: added in method section (chapter 2.2): “Following the objective of this study, a set of meteorological variables is chosen as inputs for the ML model that either influence PM concentrations directly via dilution (MLH, wind speed (ws), and wet scavenging of particles (precipitation)) and particle transport (wind direction as u, v components, air pressure (AirPres)), as a proxy for emissions (e.g. from residential heating: temperature at a height of 2 m (T)), and as a proxy for transformation processes (total incoming solar radiation (TISR), relative humidity (RH), T).

"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 learn-ing 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."

Answer:

The novelty and also the advantage of the machine-learning framework is that all meteorological influences on PM1 concentrations are quantified at the same time, and interactions between the meteorological variables are captured. On this basis, their influence on any individual event can be separated and quantified (as done in chapter 4.4). These aspects are novel and taken together, exceed the potential of past observation-based analyses.

It was not the aim of this study to explore the applicability SHAP values and it is unfortunate if this impression is conveyed by the current state of the manuscript. Therefore, extensive changes to the manuscript have been made to sharpen the scientific contri-

bution of this manuscript and to more carefully emphasize scientific contributions.

Still, it is important to note here that much of the methodology chapter is dedicated to the ML algorithm and the SHAP values to make sure that the results chapter can be followed by readers not familiar with these techniques. An evaluation of the model such as in chapter 4.1 is critical to ensure that the model is able to reproduce empirical patterns.

Large parts of the abstract, introduction and the conclusion section were altered to shift the focus from the SHAP approach to the scientific findings. The following specific changes were made in the manuscript:

- L510: removed “To our knowledge, this is the first time that the SHAP-framework for explainable machine learning is applied in atmospheric sciences”

- Headline 3.2: added “to infer processes” to stress the purpose of the SHAP values

- L211: was changed to “The interactions of input features contribute to the model output and thus reflect empirical patterns that are important to deepen the process understanding.”

- L215: deleted from the manuscript “SHAP values are a novel tool to better understand multivariate natural systems, in particular when applied in state-of-the-art machine learning models as GBRT. So far, SHAP values have been used in the fields of computer science (Antwarg et al. 2019) and medical science (Lundberg et al., 2018b; Li et al., 2019a; Lundberg et al.,2020), but have yet to be applied to study environmental systems.”

-L96: Removed “With the use of SHAP values, a detailed insight to the decisions of the statistical model can be provided, hence allowing an advancement of previous ML approaches (Friedman, 2001; Lundberg et al., 2018a).”

-L508-511: Removed “The GBRT approach in combination with the SHAP regression values presented here provides an intuitive tool to assess meteorological drivers of

air pollution and to advance the understanding of high pollution events by uncovering different physical mechanisms leading to high-pollution episodes."

-L248: added "...as suggested by Fig. 5d..." to state more clearly that this constitutes a new finding

-L404-405: added "The physical explanation behind this pattern would be that lacking wet deposition and low wind speeds increase particle numbers in the atmosphere, while northeastern winds advect further particles. Given that there is now a large number of particles available, the accumulation effect of a low MLH is more efficient"

See also changes in L96-103, which now more clearly pinpoint the purpose of the study.

Specific comments

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

Answer:

The available ACSM instrumentation does process only PM1 particles. PM1 is highly relevant for human health, affecting the respiratory system. Smaller particles can penetrate deeper into the lungs compared to larger particles and potentially cause more damage Studies show that health impacts of PM1 are similar (Yang et al., 2018, DOI: 10.1016/S2542-5196(17)30100-6) or worse than PM2.5 (Chen et al. 2017, DOI: 10.1016/j.envint.2018.08.027). In addition, a study by the WHO indicates that BC is a good indicator for human health, which is most prominent for particles smaller 1μm (see https://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2012/health-effects-of-black-carbon-2012). A comparison to PM10/PM2.5 is currently not feasible since no simultaneous measurements

of PM1, PM2.5, PM10 and meteorological parameters at the same site are available

- Added to L111: "..., a highly health relevant fraction of PM including small particles that can penetrate deep into the lungs (Yang et al., 2018; Chen et al., 2017a)"

"Line 21 - "Processes vary even within seasons"This does not read well. Of course, processes will vary within seasons."

Answer:

Sentence was removed from the manuscript.

"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?

Answer:

Yes, this was the intention. To make this more clear, the sentence was changed to "likely triggers increased local wood-burning emissions, which increase PM1 concentrations"

"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."

Answer:

This sentence was removed from the manuscript and replaced by "The identification of these meteorological conditions that increase air pollution could help policy makers to issue warnings to the public or install preemptive measures by specifically accounting for meteorological variability that influences PM1 concentrations. Furthermore, the presented framework has the potential to assess the effectiveness of air pollution mea-

sures." L8 was changed to . . ."Based on the model, an isolation and quantification of individual meteorological influences for process understanding is achieved..."

See also changes in the introduction (L98-106) and conclusion (L502-510).

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

Answer:

Extensive changes in the manuscript have been made in L96-103 (see also previous answers). In addition, potential applications and the new insights were emphasized in various parts of the conclusion section (L475-480, L502-510, L515-520).

- L482-485: changed to "For policy makers, the presented approach could prove beneficial in multiple ways and serve as a decision aid for air policy measures. Another application would be to attribute changes in air quality to policy measures by comparing an 'expected' level of air pollution under given meteorological conditions to actual observations (e.g.,Cermak2009 and Knutti 2009), which may help. . ."

"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."

Answer:

The results relate to the measurement site, which is representative of the Paris region background values. This was added in L127: "PM1 measurements are representative of background pollution levels of the region of Paris (Petit2015 et al., 2015)". The sentence was rephrased in L94 "govern pollution concentrations at the measurement site" instead of "lead to high pollution events in Paris"

Technical comments

"Please use sub/superscripts for chemical species throughout (e.g. SO42-,

SO2,PM2.5)."

Answer:

This was adjusted accordingly.

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

Answer:

This was adjusted accordingly.

"Expand acronyms once per major section too."

Answer:

Given the limited number of acronyms, the authors propose to extend them only at the first mention.

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