Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2020-469-AC2, 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

"The research work employs field measurements of Particulate Matter smaller than1μm in diameter (PM1), routine meteorology data and propose a machine learning framework in air-pollution forecasting. The authors address the significant challenge of Interpretability in machine learning using the SHapley Additive exPlanation (SHAP)regression values. A general comment is related to the aim of the proposed work. Although the role of meteorology on PM concentrations is well studied, the paper proposes a novel method/tool of explainable machine learning in atmospheric sciences. The results support the use of explainable machine learning as a statistical

modeling framework in operational air quality forecasting. The authors comment on this in the conclusion section but could highlight the ability of the proposed framework earlier in the manuscript. "

Answer:

Thank you for your assessment. In order to highlight the capability of the framework in air quality forecasting, the following statement was added to the introduction (L105): "Furthermore, the proposed ML framework can be viewed as a first step towards a data-driven, prognostic tool in operational air quality forecasting, complementary to CTM approaches.."

"While the results of the study are of local interest the proposed modelling framework has a high replication potential in areas with limited PM1 field measurements and therefore has a general implication in atmospheric science. Some concern is related to the use of meteorological data for the period of July to mid-November 2016. It is useful to include some descriptive statistical analysis of the meteorological data for all sites in order to compare and highlight the suitability of using meteorological data from the Paris Charles de Gaulle Airport."

Answer:

Please note that only MLH was substituted during that time, all the other variables were not affected by the instrument failure. An appendix was added. Figs A1 and A2 now provide a comparison of MLHs measured at Sirta vs. MLH measured at Charles de Gaulle airport for available data of the year 2016.

"Furthermore, the locations of all measurement sites should be included in the map of the area of study and use more appropriate location mark labels."

Answer:

Map (Figure 1) was changed accordingly.

"The authors could provide some descriptive statistical analysis of the PM1 field measurements. This analysis could provide thresholds of high-pollution events in the region(e.g. similar to the >95 percentiles used in the paper)."

Answer:

Table showing mean, interquartile range, 95 percentile was added in section 4.3 Several thresholds were tested to define high-pollution events (e.g., >75%, >90%, >95%). It was found that the more extreme the events, the clearer the meteorological influence, which points to a relatively narrow set of meteorological characteristics responsible for high-pollution events. The decision to finally use the 95th percentile for the analysis was because this is a typical threshold for extreme value analysis. Below is shown the plot for extreme events >75th percentile; patterns are similar, but less distinct compared to >95th percentile as shown in the manuscript.

Please note that a more extensive description of the statistics of the PM1 field measurements is also provided in Petit et al., 2014 (DOI: https://doi.org/10.5194/acp-14-13773-2014), Petit et al., 2015 (DOI: https://doi.org/10.5194/acp-15-2985-2015), Petit et al., 2017 (DOI: https://doi.org/10.1016/j.atmosenv.2017.02.012).

"The overall ability of the proposed framework could be also evaluated using exceedances forecast verification metrics (e.g. Probability of Detection, False Alarm Ratio etc.) for certain PM1 thresholds. This analysis could be complementary to the analysis of high-pollution case-studies and role of meteorological conditions of high-pollution events."

Answer:

As the main objective of this study is to advance the understanding of meteorological drivers, a regression model was set up and validated to reproduce the temporal development of PM concentrations. The model that was set up for this task does not do a classification, therefore calculating the Probability of Detection or False Alarm Rate

might not be suitable for validation. Hence, while these are certainly good suggestions for a classification framework, they are out of the scope of the manuscript and could be included in future work on this topic.

"The paper presentation and structure is clear and supports the discussion of the results. The authors give proper credit to earlier published work and discuss their findings appropriately. The figures in the manuscript support the discussion of the results. "

Answer:

Thank you!

"In Figures 9 and 10 the color-bar of the PM1 composition could be misleading. It is advised to change to avoid confusion with the feature contributions color scales."

Answer:

The colors used for major PM1 species are consistent with previous literature (e.g., Petit et al. 2014, Petit et al. 2015, Dupont et al. 2016), so the colors of the meteorological variables were changed instead. In addition, a horizontal line was added to more clearly separate the PM1 composition plot.

"The authors should also check of consistency of abbreviations throughout the manuscript. For example, Mixed-layer height (MLH) in some figures is abbreviated as BLH."

Answer:

Thank you, BLH was changed to MLH in all affected plots.

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

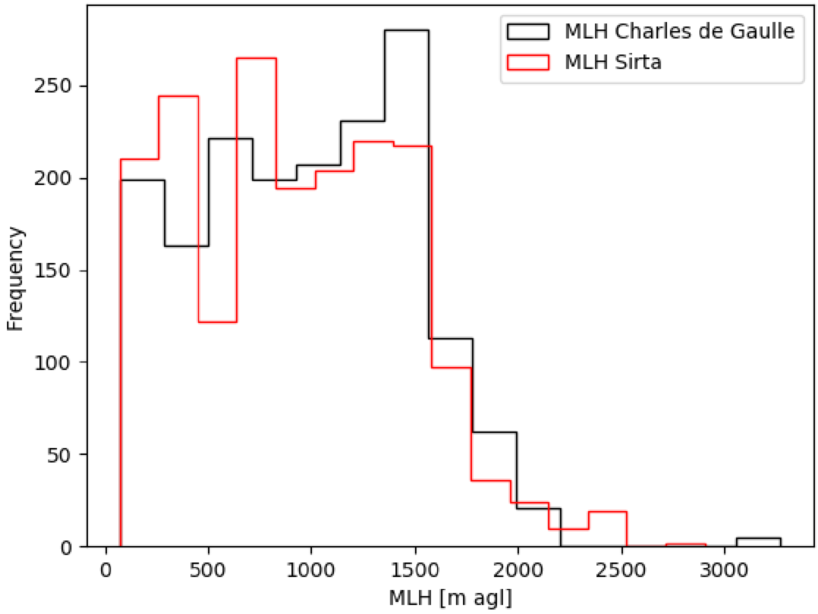
MLH Charles de Gaulle
MLH Sirta
Frequency
MLH [m agl]
0
50
100
150
200
250
0
500
1000
1500
2000
2500
3000

Fig. 1.

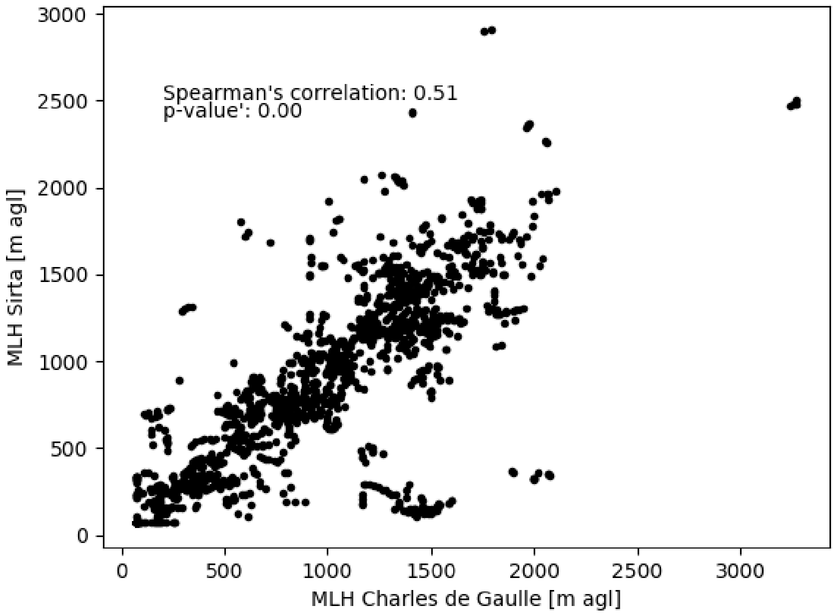
Spearman's correlation: 0.51
p-value': 0.00
MLH Sirta [m agl]
MLH Charles de Gaulle [m agl]
0
500
1000
1500
2000
2500
3000

Fig. 2.

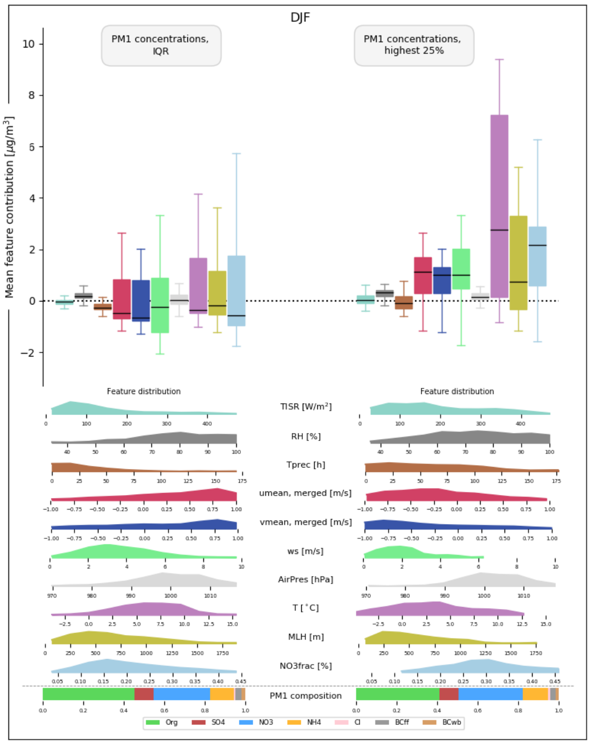
DJF
PM1 concentrations, IQR
PM1 concentrations, highest 25%
Mean feature contribution [μg/m³]
Feature distribution
Feature distribution
TISR [W/m²]
RH [%]
Tprec [h]
umean, merged [m/s]
vmean, merged [m/s]
ws [m/s]
AirPres [hPa]
T [°C]
MLH [m]
NO3frac [%]
PM1 composition
Org
SO4
NO3
NH4
Cl
BCff
BCwb

Fig. 3.