Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2021-743-RC1, 2021

Comment on acp-2021-743

Anonymous Referee #1

Referee comment on "Understanding aerosol microphysical properties from 10 years of data collected at Cabo Verde based on an unsupervised machine learning classification" by Xianda Gong et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-743-RC1, 2021

Review of manuscript for Atmosphere

The paper 'An unsupervised machine-learning-based classification of aerosol microphysical properties over 10 years at Cabo Verde', by Gong et al. investigates aerosol properties and their relation to properties relevant for cloud formation and puts the results also in the perspective of air mass origin. This topic is very relevant for a better understanding of the interplay between aerosol and clouds. In particular, this study investigates data sets for a long time period and for a region of interest (influenced by both marine and dust sources, area with not so many observations). The manuscript is therefore well fitting into the scope of Atmospheric Chemistry and Physics.

Overall, the manuscript is well written and I can recommend publication for ACP after some revision, described below.

I have a slight preference that sections 4 and 5 are re-organised. The synopsis presents already to a good part the conclusions, and the conclusions more the future work. I would prefer to have a well-structured section 4 as conclusion and a shorter section 5 for the future work outlook.

General comments

As machine learning is in the title, I would have expected a more detailed introduction to it. However, for the understanding of the article, the necessary descriptions are given. But the authors should explain also the difference between supervised and un-supervised machine learning algorithms.

The light absorbing carbon data has not been included in the aerosol type classification. However, such a long data set would be worth to study or use in more detail. Have the authors tested to include the LAC data in the clustering?

The authors dispose of a 10 years long data set. Although a break-down by year or seasons and respective description would probably result in a too long paper, it would could have been worth to analyse this for some distinct topics. E.g., the interesting result of high Nccn numbers during dust periods with low critical diameter (as discussed on page 17, lines 1-6). Maybe a look per year or by season would have brought some additional insights to this.

Specific comments

Page 2, line 2: please check also for a few more recent dust INP articles (eg, Hoose et al, doi:10.5194/acp-12-9817-2012.Kanji et al, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0006.1, Boose et al., https://doi.org/10.5194/acp-19-1059-2019)

Page 3, line 3: the machine learning algorithms are described to deliver 'faster, accurate results' this is a comparison, please mention compared to what.

Page 3, lines 1 to 10: the aerosol classification is done by machine learning algorithms. Not mentioned is why other methods, applied often like, eg, multivariate analytical method, principal component analysis were not applied here.

Page 3, lines 13-17: would it be possible to integrate a wind rose graph here? In Gong et al 2020b there is one, however, not for the whole time period investigated here. Please refer to a suitable reference (like done in Gong et al 2020b). In addition, for the air mass origin analysis it is mentioned later that the boundary layer height was derived in previous studies. The paper of Gong et al 2020b referred to, is however only for a short period. The study here is however over 10 years. How can the authors assure that the used assumption for the boundary layer height is overall valid?

Page 3, lines 18ff: the aerosol inlet is at 32m; it is however not mentioned in this text how long the total inlet tubing to the measurement container is (mentioned however in Gong et al 2020b). However, this is important to judge the sampling set up. The authors should also mention that this long tubing and related particle losses is accounted for in their particle loss corrections (it is I assume).

Page 3, line 19: '… to minimize the influence of sea spray aerosol…': Please explain bit more, why in particular the sea spray aerosol should be minimized, or refer, eg, to set ups at GAW stations like Mace Head, Ireland how they set up such measurements, at which height?

Page 4, line 8: The authors mention that MPSS and APS were calibrated regularly. Please mention briefly how, where.

Page 4, lines 25ff: derivation of scattering coefficient; assumption for the refractive index. The scattering coefficient is not presented later in the manuscript. This might be either skipped or the authors describe why this derivation can be useful for their analyses.

Page 5, section 2.5, backward trajectories: the authors should give some more details on the initialising parameters for the HYSPLIT model. E.g., which data were used for the meteorological fields, which height resolution was applied, what was the spatial grid size resolution?

Page 6, section 3.1.1 and related figure 1: the authors mention that the number concentration for the supermicron particles show a high variation (1 to above 100 (here, the unit is missing)) – but in the figure the scale only goes up to 50 cm-3. Further, in the caption in Figure 1 the submicron range is given for 10 to 1000 nm, but in section 2, the MPSS measures only from 20 nm onwards. Also, please mention the time resolution of the PNSD data – hourly, daily averages?

Page 6, section 3.1.2: line 20: replace 'concentrations' with values. It's the absorption coefficient, not eBC concentration; Please adapt also accordingly in the caption of Figure 2.

Page 6, section 3.1.3, CCNC time series: please mention on which time basis the values are presented – hourly, daily? Please mention this also in the caption of Figure 3. Further, CCN values for a SS=0.7% are missing for around December 2015 to March 2016 – why? Also, please mention in the caption of Figures 4 and 5 if the shown values are for the whole CCN measurement period (I assume to be so).

Page 9, line 3: Barbados: mention briefly where it is located

Page 9, line 9: Hoppel minimum and related supersaturation 0.3%: this is valid for the presented data set, please write accordingly

Page 9, line 13: comparing Nccn at 0.3% ss with Nhm : the mentioned scatter plot: either insert a 'not shown' here or show it (maybe in appendix); also, please clarify if the correlation is valid for the whole CCN data set period, or how it changes if you look at the correlations per month.

Page 11, lines 9-10: how would the number of derived aerosol types differ if the authors would have chosen a different set/number of size range bins? Have the authors tested this? The chosen 5 size ranges take by themselves already a hypothesis of aerosol type classification. Does this not pre-define the result of the classification?

Page 12, Figure 6: please clarify briefly if the given relative frequency, integrated over the whole, totals finally 1 (then it would not be [%]) or 100 % / same for Figure 8.

Page 13, line 6: the authors argue that the high concentrations of very small particles indicate new particle formation events. It sounds like as if only NPF events are responsible for these concentration numbers. Could also other mechanisms like, eg, simple transport (from upper troposphere), could be responsible?

Page 15, section 3.2.2: the authors mention only briefly the freshly-formed cluster in this section. However, from the seasonal cycle in Figure 9 there are clear variations worth to be discussed.

Page 15, lines 26-30 would also fit into the introduction, in order to tell the reader why this paper makes a significant additional contribution, compared to the previous Gong et al papers for Cabo Verde.

Page 18, line 6-8: please clarify in text that the observed particles during dust periods with comparable hygroscopicity like particles during marine periods – that thes were most likely not dust particles, but that the new particle formation happened within the dusty air mass origin events.

Page 19, lines 1-7: please clarify if the kappa values discussed here for months October through March are for all the 10 years of PNSD observations (and accordingly with derived

Nccn numbers as described with the method earlier).

Technical comments

Title: machine-learning: with or without hyphen? Because, otherwise in the manuscript it is written without hyphen

Page 2, lines 1 to 17: Please check time applied

Page 3, line 31: '.. for a more detailed explanation…' or '… for more detailed explanations …'

Page 4, line2: skip 'a' before 'density' / line 3: '… chloride are …' / line 4: '… of mineral dust are … and within a range of …' / line 5: '… shape factor and density of 1.17 and 2000 kg m-3 were …'

Page 4, line 20: 'extent'

Page 6, line 13: '… particle number concentration (…) in number per cubic…'

Page 13, line 13: '… were present a similar …' ; discard the 'were'

Page 17, line 4: '… from new particle formation in an air mass …' or '… in air masses ..' / 'a phenomenon'

Page 19, line 22: 'K-means' can be skipped here