Review of Dada et al "*Sources and sinks driving sulphuric acid concentrations in contrasting environments: implications on proxy calculations*"

The manuscript addresses an important issue on predicting sulphuric acid concentrations when the measurements are not available. Especially finding an applicable proxy for night-time concentrations would be a significant improvement to existing literature. The manuscript introduces different variations of the proposed proxy and they seem to fit nicely on the measurements in selected locations. However, the procedure how the proxy variations were derived and the conditions where the measurements were made need to be described in more detail before the applicability of the proxies can be evaluated and I can recommend the manuscript for publication.

Major comments:

The proxies for individual campaigns were derived from the same data they are predicting, these proxies need to be verified on independent data before they can be generalized even on different conditions in the same sites. In addition, the data were collected from short periods, except for Hyytiälä, and it would be helpful if there would be some discussion on how representative the measurements are compared to annual level or long term seasonal averages of all variables in the sites. Bootstrap resampling is good method in the case where not so much comparable data are available but it is not enough for constructing a generalizable tool if the measurements are not representative.

Derivation of night-time proxies in Hyytiälä should be revisited. I would suggest calculating separate proxies for dark time without global radiation included, or similarly than in China, as the chemistry is different during the dark hours. The manuscript suggests that the night-time formation of sulphuric acid is mostly driven by Criegee intermediates and thus the coefficient k_2 in China was seen to be significantly higher than for daytime and that might be the case also in Hyytiälä.

Specific comments:

Page 2 line 76: proved->suggested

Page 2 lines 91-93: Bold statements, considering the comments in this revision regarding generalizability

Page 3, lines 102-104: Were all the measurements made on the same platform?

Page 3, lines 130-134: I have recently learnt that calibrating CI-APi-ToF is not an easy task (Talk by Ylisirniö et al. EAC2019). Were the instruments calibrated such that the results between sites are comparable and are the measured concentrations of realistic magnitude?

Page 4, lines 145-155: CS was reported in Hyytiälä with RH correction and in other sites no such correction is defined. The CS measures should be consistently defined if the results are being generalized.

Page 5, lines 183-184 and Figures S3-S7: Why Pearson correlation coefficients? The data are most probably not normally distributed and they contain outliers, which violate the basic assumptions of Pearson correlation.

Page 5, lines 203-209: How the sink term $k_3[H_2SO_4]^2$ is defined? It needs to be clarified here for usability of the proxy

Pages 5-6, Equations: Overall, the notation of the equations is somewhat confusing. First term is clear, does the second term refer similarly as the first one that it is k_2 times ozone concentration times Alkene concentration times SO_2 concentration? In addition, does $[H_2SO_4]$ in third term refer to sulphuric acid concertation or that the CS is calculated for sulphuric acid? Does in last term $[H_2SO_4]^2$ refer to squared concentration, and if yes, drawn from where? I suggest clarification of the equations.

Page 6, lines 242-249: It is not surprising to see that the Petäjä proxy had some difficulties, as it is constructed only with data from Hyytiälä. Already in Mikkonen et al. (2011) it was seen that the Petäjä proxy is not always working well outside of Hyytiälä. Thus, it would be interesting to see comparisons on proxy from Mikkonen et al., which has been shown to work in varying environments.

Page 6 lines 251-254: The predictor variables in the proxy contain high measurement uncertainty. Does the fminsearch procedure take that account?

Page 6 lines 254-257: I am happy to see uncertainty estimation for the coefficients made with bootstrap! Though some details on bootstrap procedure should be provided, e.g. how many resamples were drawn?

Page 6 lines 260-265: How does the AIC reflect the probability of over- or under-fitting in these analyses? As calculating log-likelihood for AIC might be sensitive for number of observations was it checked that the N was the same for all proxies in certain site? With multiple instruments in use, there might be gaps in data in different time points.

Page 7, line 273 and Figure 1: Are the numbers of data points the same in each subplot?

Figure 2 and related text in chapter 4.1: Do I read the figure correctly that the proxy values from 23-02 are missing? If this is due to missing global radiation, this could be corrected by the suggestion above to derive separate night-time proxy.

Page 7, line 308: "…proves the truthfulness…" is quite an overstatement

Figure 5: Why the scale is from 10^2 when the data starts from 10^5? Overall, the observed concentrations seem rather low for urban environment. Were the conditions somewhat unusual during the measurement campaign?

Page 9, lines 388-389: Clarify how the predicted fractions were drawn for table 2 and fig 9

Table 2: 27^{th} percentile?

Figure 10: Global radiation distribution is missing. The basic statistics could also be given in (supplement) table. Sulphuric acid concentration in Megacity seems also low.

Page 10, lines 438-440: It is stated that the coefficients did not vary substantially, I might disagree. But regardless of that, did you try to pool the data from different sites an calculate a combined data proxy? Naturally with Equation 4 which could be calculated for all sites. Would this give a more generalizable proxy?

Discussion and suggestions section: It would be helpful to give here the direct equations for calculating the proxies in each site. It would probably increase the future use of the derived proxies. Equations could also be an appendix.

References

Mikkonen, S., et al. (2011) A statistical proxy for sulphuric acid concentration, Atmos. Chem. Phys. 11 11319-11334. doi:10.5194/acp-11-11319-2011

Ylisirniö et al. (2019) Solving the Mystery of Wildly Differing Results in Calibrating FIGAERO ToF-CIMS Desorption Temperature to Saturation Vapour Pressure. Proceedings of EAC2019, **O1_A6_M06**