Anonymous Referee #2

Received and published: 9 September 2011

In this manuscript, the authors seek to derive a proxy for sulfuric acid vapor concentration ([H2SO4]) from measurements taken at six sites. A number of different proxy functions (both linear and non-linear) have been tested. The best formulation of the [H2SO4] proxy uses global solar radiation, SO2 concentration, condensation sink (CS) and relative humidity (RH) as predictor variables. Nevertheless, the role of the CS (and RH) in the proxy was found to be only minor, since similarly accurate proxies could be constructed with global solar radiation and SO2 concentration alone. The derived power dependence of [H2SO4] on key parameters has been shown to be far different from those of theoretical values. While [H2SO4] proxy functions have been derived in previous studies, this study extends such studies by using data from more sites. The content of the manuscript is within the scope of ACP. Below I give my comments and suggestions for improvement.

We thank the reviewer for careful revision of our manuscript.

1. It is hard to understand that [H2SO4] is determined by global solar radiation and SO2 concentration alone. The authors state that this could be attributed to SO2 being an indicator for anthropogenic pollution which represent sinks for the OH radical. Please provide some chemical mechanisms connecting anthropogenic pollution with OH concentrations.

On page 16, starting on line 19, we speculate on the possible reasons for the weak dependence of [H2SO4] on [SO2]. As we state in line 22, we feel that the main explanation for this is that SO2 represents both a possible source of H2SO4 (through gas phase oxidation, ref eqns R1-3) and as a sink term. The reviewer suggests that this sink could be as a sink of OH, and there is certainly some validity to this statement. Besides SO2, OH reacts with several different trace gases of anthropogenic origin e.g. CO, VOCs, CH4 and NOx (e.g. Austin et al., 2002) which can be highly correlated with SO2 in certain urban settings:

e.g. OH + NO2 -> HNO3

Clarification for this is now added to the revised manuscript in page10, lines 19-23.

In addition, anthropogenic pollution can include large amounts of particulate matter, which increases the condensation sink. This explains why the success of the simple version of the proxy, since [SO2] and CS can co-vary. Spearman correlation between [SO2] and CS is in the magnitude of 0.5, depending slightly on the measurement site.

Austin, J., Brimblecombe, P., Sturges, W.: Air pollution science for the 21st century. Elsevier Science Ltd., Oxford, 2002

Can the Proxy L3 be applied to power plant plumes where SO2 concentration is very high and remote and marine air where SO2 concentration is very low? What about under the conditions that CS is dominated by sea salt, dust, and biomass burning emissions?

Proxy L3 works well in Atlanta, where SO2 concentrations can be exceptionally high but if it works directly in power plant plume we cannot say for sure. Niwot Ridge and Hyytiälä are sites with low SO2 concentrations and the proxy works well in both sites but without data from marine environment we cannot say if the proxy can be used there.

The proxy does not separate the sources of CS thus we cannot be sure what happens in these special conditions. We inserted clarification for this in the Conclusions of the revised manuscript, see comment for Referee 1.

*2. CS and RH. Theoretically there is no doubt that CS is a key parameter determining [H2SO4]. The dependence of CS on RH is nonlinear, especially when RH>_70-80%. Also the effect of RH on CS depends on the composition of particles. To use CSdry*RH to approximate wet (i.e., total) CS is questionable. What is the justification to use CSdry*RH as the total CS? Could the errors associated with this approximation affect the results?*

CSdry*RH was used because it fits the data best. We are aware that the dependence is not linear and thus used the variables first as individual parameters (Proxy N4). However, it showed out that Proxy N5 gave better results than N4. We constructed several different kinds of subsets from the data by limiting RH to different ranges and noticed that the results did not change significantly.

As we stated in the text, RH corrections by Birmili et al. and Laakso et al. are constructed only for Hyytiälä. If we could get a correction function which is valid globally we could use that instead of CS^d*RH^e in the proxy construction

3. As the authors mentioned, [H2SO4] proxy functions have been derived in previous studies with the limited data. It is necessary to discuss the difference between the present proxy functions and previous ones. How much is the improvement?

The proxies used in Petäjä et al. and Hamed et al. are basically the same as the Proxy L1 in our study, with only difference that we took the reaction rate k by DeMore et al. (1997) in the proxy. This is now stated in the revised manuscript in page 10, lines 12-14.

As Tables 4 and 7 shows, the improvement is significant.

4. I would be useful if the authors can estimate the uncertainty of predicted [H2SO4] based on their derived proxy functions. The limitation of recommended proxy formulation should be discussed.

The uncertainties of the proxies are now estimated with the average absolute errors relative to the dependent-variable mean, given by Willmot et al. (2009). See page 15, lines 13-22 in the revised manuscript.

Limitations of the use of the proxies are also discussed in the revised manuscript, see comment for Reviewer 1.

5. p20148, L5-6. Also Table 2. Are the values given in Table 2 for whole period of measurements (day and night)? Since [H2SO4] is generally very small at night. It is more relevant to consider the daytime data only for the statistical analysis presented in Table 2.

The values given in Table 2 are calculated from 10 minutes averages of all data, including day and night values. This is due to the fact that we wanted to show the overall conditions in the measurement sites. Furthermore, the mean values do not change drastically if we use only daylight hours instead: mean [H2SO4] for Atlanta, Hyy03, Hyy07, mel, nwr, spc for daylight hours are 1.38E+07, 2.10E+06, 5.43E+05, 8.71E+06, 2.29E+06, 7.51E+06 respectively while for the whole data they are 1.29E+07, 1.41E+06, 4.29E+05, 6.42E+06, 1.82E+06, 5.39E+06 respectively.

6. p20151, second paragraph and Figure 3. Figure 3 is exactly the same as Figure 2. Need to provide the right Figure 3.

We apologize for the confusion with the figures. The correct figure is now inserted into the revised manuscript.

7. p 20152, L25-27. Why? In the same paragraph, the authors cited several references to show that steady-state assumption should hold. Could the authors give some theoretical arguments under what conditions that the steady state assumption is unrealistic and why?

The argument that the steady state assumption might not hold in some conditions is based only on the data and the result that the proxy derived from the theory is not the best one. We cannot say for sure what causes the non-equilibrium state of the balance equation and thus cannot give the theoretical arguments requested. However, as the reviewer noted we state quite clearly in the same paragraph that the steady state assumption should hold at least in remote areas but all our data is not from remote locations.

8. p 20153, L10-14. Please provide some details of the nonlinear fitting procedures. Why this method can take account the non-equilibrium conditions?

Details of the nonlinear fitting procedures can be found from various text books, including Bates & Watts mentioned in the text. However, we inserted following text into the same paragraph of the revised manuscript (page 11, lines 15-22):

"Nonlinear regression is usually needed when there are physical reasons for believing that the relationship between the response and the predictors follows a particular functional form. The general form of nonlinear regression model is given by

$$y_i = f(x_i, \beta) + \varepsilon_i$$

where y_i are the measured response observations, f is a known nonlinear function of the measured predictor variables x_i, β are the estimated parameters of the model, ε_i are the random residuals of the model which are usually assumed to be uncorrelated with mean zero and constant variance."

If the conditions are in equilibrium, the dependence between the response variable and the predictors should be linear. Then all powers of the proxy function should be unity. As the powers vary a lot from unity we can state that the model takes into account the non-equilibrium conditions.

9. p 20153, L24-25. Could this an indication of the problem in the fitting method? I think this should be reflected in the abstract.

It is not a problem, it is a result. The equilibrium requirements for the balance equation are not fulfilled and thus the powers differ from unity.

10. Equation N5c. What are the values of Ii used in this study?

Values li are added as Appendix table in the revised manuscript.

11. If possible, the authors should publish the data they used in their analysis as supplementary materials so that interested readers can independently look into the issues and assess the derivations.

The data used in this study is used by kind permission of the institutes who have conducted the measurements and we do not have permission to distribute the data freely. If the reviewer is interested on the data, she/he should contact following persons:

Atlanta: Peter McMurry

Hyytiälä: Markku Kulmala or Tuukka Petäjä

Hohenpeissenberg and SPC: Christian Plass-Duelmer

Melpitz: Wolfram Birmili

Niwot Ridge: Jim Smith, Michael Boy or Lee Mauldin