Review of “Characterization of satellite based proxies for estimating nucleation mode particles over South Africa” by Sundström et al.

**General Comments:**

This paper is a straightforward application of the existing technique of Kulmala et al., ACP (2011) to estimate nucleation mode particles over South Africa from MODIS and OMI data products. Since nucleation mode particles are too small to be directly detected via ground-based or space-borne optical instruments, a proxy is computed that balances retrieved concentrations of gas-phase particle precursors against the existing aerosol condensational sink as represented by particle concentrations in the larger, optically-active size range. The difference between the two papers is that this study focuses on South Africa, while Kulmala et al., ACP (2011) focuses both on Hyytiala as well as extends the work to create global nucleation proxy maps. The present paper also develops a slightly different method for approximating the condensational sink (CS) from ground-based measurements and AERONET retrievals. The new method for approximating condensational sink is found to not substantially improve over the existing assumption that CS=AOD.

In reviewing Kulmala et al., ACP (2011), that Referee #1 noted that “After some changes, [that] manuscript could be suitable for publication in ACP. Not for the goodness of the results but in order to encourage developing more suitable satellite products for the analysis of fine particles. [Her] suggestion is that the main conclusion should be reformulated such that it is not possible to get adequate estimation results for nucleation mode particles with current satellite products.” These comments also apply to the present manuscript by Sundström et al., which also shows the extremely poor skill of this technique when using the MODIS and OMI data (Table 3 and Figures 5 and 10). Yet, statements in the abstract and conclusions sections imply that these satellite-based proxies are rather good at showing the potential for nucleation events – this conclusion is not supported by the results!

Publishing null results is important because it prevents wasteful duplication of effort and it motivates future work to either improve the proxy method or supplant it with another technique. As such, a paper discussing the reasons for the poor proxy skill and perhaps linking these results to sources of uncertainty and proxy sensitivity would be a welcome addition to ACP. However, with the abstract, discussion, and conclusions as presently stated in this manuscript, I cannot recommend publication.

**Specific Comments:**

1) Throughout the manuscript, linear regression is used to compare the proxy and measurement variables, and goodness of fit is assessed using a Pearson correlation coefficient, r, and associated p-value.

First, while the correlation coefficient (r) does give us a metric for assessing the linear dependence between the proxy and measurement variables, the coefficient of determination ($R^2$) is more meaningful in evaluating the skill of the proxy by representing the proportion of total variation in the measurement captured by the proxy. In all instances in the manuscript except for the comparison between dry scattering coefficient and condensation sink (Figure 3), the $R^2$ values are around 0 - 0.3. The most direct and important comparison (Figure 10) shows an $R^2$ value of 0.10. This means that the proxy is only able to explain 0-30% of the observed variability in the measurement variables, and that a majority of the variability remains unexplained. Given this high level of uncertainty, slight improvements in the correlation coefficient do not really show an increase in the skill of the proxy, as is suggested, e.g., on Pg. 25846, Lines 6-8. While it's easy to square the values of r now reported, I suggest the authors report the $R^2$ values instead throughout the manuscript as a more direct metric for assessing the ability of the proxy to represent the observations.

In addition, scatter plots of all one-to-one fit comparisons with regression lines should be uploaded as supplementary information. As stated by Referee #1 in reviewing Kulmala et al. (2011): "Drawing a line through a random sample and claiming that there exists significant correlation is bad statistics and in some cases even deceptive." Being able to visualize the regressions used to generate Tables 2-3 is essential for understanding how the data are distributed, and including them in the SI ensures that they don't clutter the main paper.

Second, the p-value tells us whether or not we can reject the null hypothesis that $r = 0$. As illustrated in the plot below (computed assuming the test statistic follows a chi-squared distribution for simplicity), the minimum value of r needed to reject the null hypothesis rapidly decreases as the sample size increases. This can mean that for N greater than about 50-100 points, the correlation can be very weak (i.e., lacking scientific or explanatory significance), but still be statistically significant. As such, this p-value statistic is not really meaningful and should be removed. Instead, the number of points, N, used in the regression should be reported in both the tables and figures.

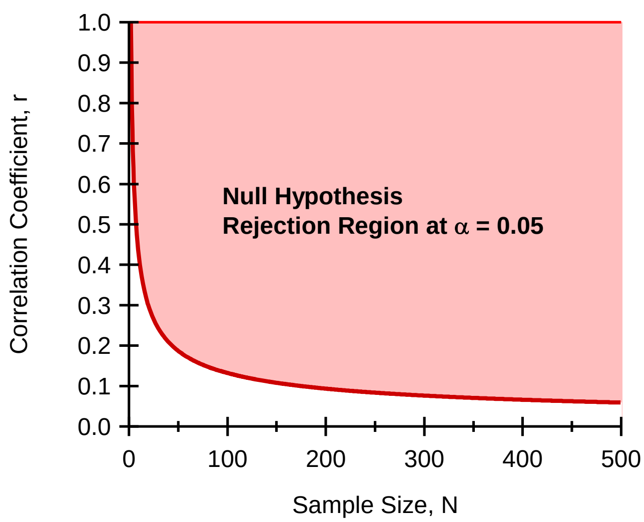


**Figure 1: The shaded area above the curve shows the values of r where the null hypothesis (i.e., r=0) would be rejected plotted as a function of sample size for alpha = 0.05. For N > 50, even a weak correlation coefficient is sufficient to reject the null hypothesis.**

2) There needs to be a little more honesty in assessing the performance of the proxy. The following statements are not supported by the current results, and therefore should be removed or new supporting data or reanalysis be included to support the claims:

Page 25826, Lines 10-12: *"However, when the AOD in the proxy sink was replaced by an estimate from linear bivariate fit between AOD and CS, the agreement with the actual nucleation mode number concentration improved somewhat."* I presume that this relates to the YORK fit vs. the LSQ fit in Figure 5. It's true that the LSQ fit completely fails (presumably due to extreme outliers outside of the plot area), while the YORK fit seems to follow the data better, but it is not clear that this translates into a meaningful improvement in capturing the nucleation mode number concentration.

Page 25826, Lines 16-19: *"Best agreement between the satellite and in situ based proxies were obtained for $NO_2$/AOD and UV-B/$AOD^2$, whereas proxies including $SO_2$ in the source term had lower correlation."* Using the numbers from Table 3 to compute $R^2$ values, the $NO_2$/AOD and UV-B/$AOD^2$ proxies are able to explain 10-12% and 3-6% of the observed variation in nucleation mode number concentration, respectively. This is contrasted with the metrics including $SO_2$, which are able to explain at 1-5% of the variability. So, while this statement is technically correct, I find it to be misleading to the reader in that it suggests that there is indeed agreement between the proxy and the observed nucleation mode number concentration. There is only 10-12% agreement at most.

Page 25845, Lines 8-10: *"A distinct improvement in the quality of the proxy components was obtained when different satellite products were selected to those utilized by Kulmala et al. (2011)."* This statement does not appear to be true. r values from K2011 and present study as follows:

K2011: UV-SO2/CS^2 = 0.54, UV/CS^2 = 0.49, UV-SO2/AOD^2 = 0.25; UV/AOD^2 = 0.23

Present Study (Table 3): UV-SO2/AOD^2 = 0.09-0.21; UV/AOD^2 = 0.17-0.25

Page 25846, Lines 6-8: *"Some improvement, however, was obtained (0.21 < R < 0.34) when the AOD was replaced by the estimated sink from the York fit (Fig. 5)."* Using the numbers in Table 3 to compute $R^2$ values, the use of the York fit versus AOD improves the percentage of explained variance from 10%→12%, 1%→5%, 1%→4%, and 3%→6% for the four different proxies in the order given in Table 3, respectively. It is not clear to me that this is a meaningful improvement.

Page 25846, Lines 20-21: *"In general this study showed that the satellite proxies seem to be able to show the potential for nucleation events in a statistical sense. Actual data from non-event days would have been needed to carry out such study."* This sentence is just not supported by the data. The statistics are heavily skewed toward event days with no data from non-event days. Therefore, it's not possible to be able to validate or invalidate the ability of the proxy in distinguishing event from non-event days. The second sentence is correct – actual data from non-event days would be needed to draw a conclusion.

Page 25846, Lines 22-24: *"More studies of the satellite based proxies in different type locations and environments are needed to improve the proxies, and especially the sink term, further."* This study and that of Kulmala et al. (2011) have shown that these proxies have very low skill when using column-integrated satellite measurements. This is probably due in large part to the uncertainties and coarse resolution associated with using these column-integrated measurements, and less due to regional peculiarities that might be uncovered by the authors' suggested path forward. Consequently, I doubt that increasing the number of locations and environments studied while following the same set of methods as these two studies will actually improve the proxies. Rather, I would think the best way to improve the proxies would be to improve the satellite inputs. I'd like to encourage you to include some discussion in the paper on measurement uncertainty and the sensitivity of the proxies to these uncertainties, which could serve to underpin future measurement design considerations.

Page 25846, Lines 24-26: *"The next step is to study the satellite based proxy approach in China, where, in addition to the elevated $NO_2$ and $SO_2$ column densities, the AOD signal is also strong."* I don't understand the meaning and purpose of this statement, which both singles out the entire country of China as being particularly polluted (which is not supported by any discussion in this paper), and seems to imply that the proxies have worked so well in South Africa that no further work is needed and it's time to move on to more complicated regions -- China. This statement should be removed.

3) Table 2 and the discussion on Page 25841 indicates that there are weak correlations between the in situ and the satellite-based proxies. If you use only the ground-based, *in situ* $NO_x$, $SO_2$, UV-B, and N(Dp>100nm) to compute the proxies (Equations 1-3), how well does it correlate with the nucleation mode number concentration? This sort of analysis must be included in the discussion because it places an upper limit on the skill of the proxy in capturing nucleation mode number concentration. If the collocated *in situ* measurements don't produce reasonable proxies of nucleation mode number concentration, then the much more uncertain, coarser satellite retrievals will not be able to do any better.

4) Kulmala et al. (2011) discuss multiple potential proxies corresponding to different assumptions related to the exponent, n, in their equations 6-12. What is the reason for only exploring a single regional proxy as given by Equation 1 in this work?

5) There is a lot of detail about specific point sources given on Pages 25838-25839 including some discussion of the ore types in the smelters. This level of specificity and discussion doesn't seem relevant to this paper, which is concerned with characterizing the satellite proxies using ground-based data. This section should be tied more directly into how it informs the proxy analyses or it should be removed.

6) There needs to be a more extensive discussion of uncertainties and sensitivities (as mentioned above). What are the uncertainties of each of the satellite measurements that feed into the proxies? How do these uncertainties translate into the overall proxy uncertainty through error propagation? The regional proxy could be more sensitive to errors in AOD because it's a higher-order term – is that what dominates the overall proxy uncertainty or is it UV or $SO_2$? What kind of measurement precision or accuracy is needed from next-generation satellite sensors in order to achieve reasonable proxies?

**Minor Comments:**

Tables 2 and 3 and all inline text: Report as $R^2$ instead of R. Remove p values. Add scatter plots w/ regression lines for each correlation coefficient to the supplementary material.

Figure 3: Add a histogram plot for each showing the relative error centered about the regression line, since the log-log plot makes it hard to see how the points are distributed about the regression line. Report as $R^2$ instead of R. Also, is there really only one significant figure in the regression pre-exponential constant?

Figure 4: Almost all of the points fall below AOD=0.8 and sigma=200. Please rescale the axes so that this is more clear. Report as $R^2$ instead of R.

Figure 5: Again, almost all the points are less than AOD=0.5 now. Please rescale the figures. Are there points not shown that would skew the regression line in the Botsalano panel? I wouldn't know how the LSQ curve would diverge from the visible data without some extreme outliers. Also, add $R^2$ values to each of the regression lines and consider if it makes any sense to report equation coefficients for regression lines that do not, at a minimum, explain a majority of the variance.

Figure 6: Add interquartile ranges to each median profile. For the low MODIS AOD cases, why are there no points above 3.3 km?

Figure 7: Emphasize in the caption that the $SO_2$ density is only in the PBL, while the $NO_2$ column density is the entire troposphere, and the AOD is presumably over the entire atmospheric column. The caption does not indicate that these are not all over the same vertical scale. Also, please list the locations that the points correspond to.

Figure 10: Report as $R^2$ instead of R. Remove p-value.

**References:**

Anonymous Referee #1. "Interactive Comment on 'The first estimates of global nucleation mode aerosol concentrations based on satellite measurements' by M. Kulmala et al.". Atmos. Chem. Phys. Discuss., 11, C8097-C8100, 2011. www.atmos-chem-phys-discuss.net/11/C8097/2011/