Author's Response

Based on the reviews the manuscript has undergone major changes and is now more or less completely rewritten. Therefore it is not meaningful to provide a detailed list about the changes. The major changes in the manuscript are (as listed in Answer to Rev 2.):

- Abstract is completely rewritten
- Introduction (Sect. 1): the text is somewhat modfied and obvious spelling mistakes are corrected.
- Data, Sect. 2: New paragraph is added where the uncertainties of the OMI NO_2, SO_2 and UV-products are described. Also other parts of text has been somewhat modified.
- Theoretical concepts in Sect 3: Proxies, condensation sink, and aerosol extinction. Some modifications in the text are made.
- Section 4: Results. This Section has undergone major changes and it is now re-organized. E.g. the detailed description of the emission sources as well as the seasonal variation of the different satellite parameters are removed, and the focus is now more on the performance of proxies obtained both using in situ and satellite data. Also the comparison between nucleation event- and non-event days is removed since the statistics were too much skewed (towards the event-days).
    - Subsection 4.1., comparison of CS and AOD: the text has been modified.
    - 4.2., In situ proxies and comparison with nucleation mode number concentration N_{nuc} . This is a new subsection where we calculate the proxies from the in situ data and evaluate their performance in predicting N_{nuc}.
    - 4.3., Proxies using satellite data. This Section has been further divided into three Subsections. In Sect. 4.3.1. the spatial pattern of the satellite NO_2, SO_2 and AOD as well as the proxies is defined from the four years of satellite data. In Sect. 4.3.2. the satellite parameters are compared with in situ data, and in Sect. 4.3.3. proxies calculated using satellite data are compared with in situ N_{nuc} .
- Section 5, Conclusions are rewritten completely.

Below are the responses for the two Reviewers:

Authors' answer to Reviewer 1

We thank the reviewer for critical and constructive comments on our manuscript. Please find below Authors' answer to Reviewer 1.

Review, general comment:
The authors apply a very interesting methodology which was originally introduced by Kulmala et al. (2011). The goal is to estimate nucleation mode particles over South Africa by using proxies which are constructed with geophysical parameters derived from satellite data. I very much support the development and reporting of such type of studies. It is a real challenge to gain information about processes related to new particle formation derived from satellite measurements and relate it to ground based measurements. Also, to use a combination of sensors onboard A-Train constallation as data source is an adequate input and forward-looking for such purposes. Generally, the use of synergistic observation in combination with in-situ data enable to launch excellent science.

However, the work presented here obvioulsly discloses the inadequateness of the currently suggested proxies for describing the processes in focus. The results should be carefully and critically explored, which has not been done. A creative analysis of other proxies which could substantially influence the results is missing. Although the region of interest was changed and in addition the formulation of proxies was slightly changed it is clearly shown that results don't

improve significantly. When reading the current manuscript it seems that the authors would like to introduce these results as an improvement as compared to the earlier article by Kulmala et al. (2011) (which I believe is not intended at all by the authors). The results presented here demand further discussion if it is possible to derive the envisaged goal from using these proxies and most importantly how results can be refined. In my opinion the presented approach is in the early development stage and defenitely requires further treatment. Furthermore, I would recommend to include more critical and constructive aspects in the overall discussion, e.g. to consider additional properties. I would like to encourage the authors to rewrite the manuscript to do justice to the complexity of the given research topic. In summary, I cannot recommend the manuscript in its present form for publication in ACP.

Authors' Answer:
We agree with the Reviewer that in its current form the manuscript might give an impression to the reader that the results presented in the manuscript would be an improvement to the work done by Kulmala et al. (2011), which indeed was not our intention. The main point in the manuscript is to test the performance of the proxies using actual satellite data, which was proposed in Kulmala et al (2011), but not carried out in practice. We have now critically evaluated the manuscript and it has been rewritten in many parts. More discussion have been added e.g. about the uncertainties related to the satellite -based proxies, which remains one of the major issues in this kind of applications. We have also included a new section where the performance of the proxies are tested using in situ data, to see how well the proxies overall are able to predict the number concentration of nucleation mode particles over South Africa.

We feel that the critical comments from both Reviewers have improved our manuscript, and clarified the presentation of the results.

Response to Reviewer 2

We thank the Reviewer for the throrough revision of our manusrcipt and constructive comments. These suggestions improved the presentation and quality of the original manuscript. Please find below Authors' response (**A**) to Reviewer's comments (**R**):

General Comments:

R1: 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.

A1: Based on the Reviewer's comments we agree that we need to clarify the manuscript in many parts. For example, the slightly different method for approximating the condensational sink (CS) from ground-based measurements and AERONET retrievals has been pointed out several time by the Reviewer as one of the major results, which was not our intention. Based on Reviewer's comments it is also obvious that we need to emphasize some of the differences between this study and Kulmala et al. (2011), and modify the text accordingly so that it is clear to the reader that this study is not just a "wasteful duplication of effort" over different location, but instead this study could provide some useful information about using satellite data in such applications. The major differences to Kulmala et al. (2011) study are that while they derived the formulas for the satellite-based proxies and showed preliminary satellite-based proxy maps, we use the actual satellite-data and compare it against in situ measurements. Such analysis was not presented in Kulmala et al. (2011). All the comparisons carried out in their paper between proxies, nucleation mode number concetrations, CS, and AODs had been obtained using in situ data, not actual satellite data. Hence, this is the first manuscript where the satellite-based proxies are compared against in situ measurements.

The Reviewer also raised good points especially related to the in situ-based proxies and the uncertainties related to the satellite proxy-approach (later in the specific comments), which we now have added to the manuscript. As a summary, the major changes made in the manuscript are:

- Abstract is completely rewritten
- Introduction (Sect. 1): the text is somewhat modfied and obvious spelling mistakes are corrected.
- Data, Sect. 2: New paragraph is added where the uncertainties of the OMI NO_2, SO_2 and UV-products are described. Also other parts of text has been somewhat modified.
- Theoretical concepts in Sect 3: Proxies, condensation sink, and aerosol extinction. Some modifications in the text are made.
- Section 4: Results. This Section has undergone major changes and it is now re-organized. E.g. the detailed description of the emission sources as well as the seasonal variation of the different satellite parameters are removed, and the focus is now more on the performance of proxies obtained both using in situ and satellite data. Also the comparison between nucleation event- and non-event days is removed since the statistics were too much skewed (towards the event-days).
    - Subsection 4.1., comparison of CS and AOD: the text has been modified.
    - 4.2., In situ proxies and comparison with nucleation mode number concentration N_{nuc} . This is a new subsection where we calculate the proxies from the in situ data and evaluate their performance in predicting N_{nuc}.
    - 4.3., Proxies using satellite data. This Section has been further divided into three Subsections. In Sect. 4.3.1. the spatial pattern of the satellite NO_2, SO_2 and AOD as well as the proxies is defined from the four years of satellite data. In Sect. 4.3.2. the satellite parameters are compared with in situ data, and in Sect. 4.3.3. proxies calculated using satellite data are compared with in situ N_{nuc} .
- Section 5, Conclusions are rewritten completely.

Specific Comments:

R2: *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.

A2: Initially we choose to report the R- and p-values similarly as in Kulmala et al. (2011) to be able to compare the results. However, we have now replaced those values with R^2, as the Reviewer suggested.

We agree that even though the correlation between nucleation mode number concentration and satellite-based-proxies slightly increased when using the CS-estimate from the York fit, it did not significantly improve the overall performance of the satellite-based proxy, which lead to the impression the reader might get e.g. from the sentences in p. 25846, lines 6-8. Since this is not a major improvement, we have removed this part of the text. The major issues when estimating CS using satellite AOD are still the ocurrence of elevated aerosol layers and the relative uncertainty related to the rather low AOD values often observed over South Africa. A better improvement could potentially be obtained if there would be a sufficient number of coincident vertical aerosol extiction profiles available.

R3: *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.*

A3: The scatterplots have been added as supplementary material.

R4: *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.*

A4: We refer here to the answer A2; we have removed the p-values and added the number of points as suggested. Overall, when using actual satellite data in comparisons with the in situ measurements that have been carried out during some limited time period, the number of coincident in situ-satellite observations can become quite low, as in many cases in this study, which is not good from the statistical point of view. We agree that the number of observations should be pointed out more clearly.

R5: *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.

A5: We refer here to answer A2. One of the major issues is still the fact that AOD is a measure over total atmospheric column and CS is defined from the in situ measurements. Based on the Reviewer's general comments the abstract has been now rewritten and this sentence has been removed.

R6: *Page 25826, Lines 16-19: "Best agreement between the satellite and in situ based proxies were obtained for NO2/AOD and UV-B/AOD2, whereas proxies including SO2 in the source term had lower correlation." Using the numbers from Table 3 to compute R2 values, the NO2/AOD and UV-B/AOD2 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 SO2, 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.*

A6: The sentences in the abstract **(**Page 25826, Lines 16-19) referred to the agreement between in situ and satellite-based proxies, not correlation between the proxies and nucleation mode number concentration. However, based on the Reviewer's general comments the abstract has been now rewritten and this sentence has been removed.

R7: *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

A7: The the correlation coefficients of R=0.54 and R=0.49 in Kulmala et al. (2011) have been obtained using in situ data measured at Hyytiälä, not satellite data. Also, the correlation coefficients of 0.25 and 0.23 have been obtained using in situ UV- and SO_2 -measurements, and AOD from the AERONET sunphotometer, not from satellite. Hence, in this work we show the first comparisons of the satellite-based proxies and the nucleation mode number concentration using actual satellite data.

R8: *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 R2 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.*

A8: We refer again to answer A2. Despite the slight improvement by replacing AOD with the York-fit estimate, the elevated aerosol layer and the relative uncertainty of AOD are still the most significant factors affecting the performance of AOD as a substitute to CS. We have removed this part of the text.

R9: *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.*

A9: Since the statistics are indeed heavily skewed toward the event days and conclusions can not be made whether the proxies could predict new particle formation or not, we have removed this chapter.

R10: *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.*

A10: Both Hyytiälä (Finland) used in Kulmala et al. (2011) and South Africa are locations where the AOD is often very low (< 0.1). In such locations where the relative uncertainty of satellite based AOD is high, and where an elevated aerosol layer can easily double the columnar AOD value, the estimation of (surface) CS is difficult, as is seen in this study. Carrying out a comparison of CS and AOD over an area where the observed satellite-based AOD range is wider, and overall the AOD is higher (and hence the relative uncertainty smaller), we might get somewhat better estimate of the CS. Also in such environment an elevated aerosol layer most probably would not have that large contribution to the total column AOD as it has over this study area. On the other hand, it is true that the satellite NO_2 and especially SO_2 product might still not have the sufficient accuracy that would be needed to improve the performance of the satellite based proxies.

R11: *Page 25846, Lines 24-26: "The next step is to study the satellite based proxy approach in China, where, in addition to the elevated NO2 and SO2 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.*

A11: We have modified this part of the text as it is clearly misleading. Referring to answers A10 and A15, the relative uncertainties especially in satellite AOD and SO_2 are very high over South Africa since the observed values are low over the major part of the area. E.g. within our study area the typical MODIS AOD values are abt. ~0.1, which would convert to the relative uncertainty of abt. ~65%. China was mentioned here because we have there similar in situ measurements available that were carried out in South Africa. It has been shown in several studies that over China the AOD over densely populated areas could vary around 0.5. In those cases the relative uncertainty of AOD would be ~ 25%, which would also reduce the relative uncertainty associated to the satellite-based proxies. Another question is then how well overall the proxies would work over China for predicting the nucleation mode number concentrations.

R12: *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 NOx, SO2, 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.*

A12: This is a good suggestion. We have now added a new section (the in situ proxy – nucleation mode number concentration comparisons to the manuscript, and discussion considering what could be expected from the satellite based proxies based on the results with the in situ proxies, as well as the diurnal variation of each in situ-based proxy parameter.

R13: *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?*

A13: Due to the uncertainties related to the satellite data we choose to consider only the cases of Nn,1. In the n^{th} order proxy terms the uncertainties would get even higher than they are now.

R14: *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.*

A14: The text on p. 25838-25839 has been modified, and too specific discussion has been removed.

R15: *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 SO2? What kind of measurement precision or accuracy is needed from next-generation satellite sensors in order to achieve reasonable proxies?*

A15: This is an important point. According to Tanskanen et al. 2006 the OMI UV irradiance uncertainty is about 7%, but can increase during some episodic aerosol plumes up to 20 %. The OMI NO_2 tropospheric column uncertainty has been resported as ~ 0.75 x 10^{15} molec./cm^2 (Boersma et al., 2011, Bucsela et al., 2013). Over South Africa this would convert to about 7 -25 % uncertainty, depending on how high the NO_2 column values at each location are. For OMI PBL SO_2 the uncertainty for one column observation is very high, the noise reported by Krotkov et al. (2008) can be even 1.5 DU, but when averaging over longer time period, and/or larger spatial area the noise can be reduced to 0.3-0.6 DU. Over South Africa this means e.g. that over the background areas the observed SO_2 column densities can be about the same magnitude as the noise, and over hot spots the uncertainty can be abt. 60% for a single observation, and abt. 20% when averaged over longer time period. The AOD uncertainty according to Levy et al. is 0.05+15%, which means that the relative uncertainty for AOD=0.1 would be 65%, and for AOD=0.25 35%. Hence, over background areas where both AOD and SO_2 are low, the satellite-based SO_2 UV/AOD² -proxy can have an uncertainty of over 90%. On the other hand, over source areas where both NO_2 and AOD are slightly elevated the NO_2/AOD proxy would have an uncertainty of ~50%. Overall over South Africa the uncertainty in satellite-based proxies is high. Over areas where e.g. both NO_2 and AOD are elevated, the relative uncertainty in the satellite-based proxy values would be expected to be somewhat lower than in this study.

We have added this discussion to the text.

Minor Comments:

R16: *Tables 2 and 3 and all inline text: Report as R2 instead of R. Remove p values. Add scatter plots w/ regression lines for each correlation coefficient to the supplementary material.*

A16: R2 is reported instead of R, and p values are removed as suggested. Scatter plots with regression lines have been added to the supplementary material.

R17: *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 R2 instead of R. Also, is there really only one significant figure in the regression pre-exponential constant?*

A17: We have reported R^2 instead of R, but we did not quite understand what was meant by the histogram plot. We feel that the figure, as it is now, is informative enough and the needed differences can be seen on the log-log-scale too.

R18: *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 R2 instead of R.*

A18: Changed as suggested.

R19: *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.*

A19: There are no points that skew the regression line in the Botsalano panel. Now that the axes are rescaled, the scatter of the points can be seen more clearly. R^2-values have been added to the regression lines, and the equations have been removed.

R20: *Figure 6: Add interquartile ranges to each median profile. For the low MODIS AOD cases, why are there no points above 3.3 km?*

A20: The interquartile ranges have been added to the profiles as suggested. To avoid the figure becoming too busy, the calipso profiles were vertically averaged into 200 m height bins. The points in the red profile above 3.3 km are missing since they did not have the required quality control flag.

R21: *Figure 7: Emphasize in the caption that the SO2 density is only in the PBL, while the NO2 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.*

A21: The OMI SO_2 Planetary Boundary Layer (PBL) product is a retrieval of SO_2 total column density throughout the whole atmosphere, and "PBL" refers only to the *a priori* profile assumed in that product (the *a priori* profile has SO_2 predominately in the PBL). This has been now clariefied in the text (in the caption as well as Sect. 2). The names of the in situ measurement stations have been added to the caption.

R22: *Figure 10: Report as R2 instead of R. Remove p-value.*

A22: Changed as suggested.

Characterization of satellite based proxies for estimating nucleation mode particles over South Africa

A.-M. Sundström[1], A. Nikandrova[1], K. Atlaskina[1], T. Nieminen[1], V. Vakkari[2,1], L. Laakso[2,3], J.P. Beukes[3], A. Arola[4], P.G. van Zyl[3], M. Josipovic[3], A.D. Venter[3], K. Jaars[3], J.J. Pienaar[3], S. Piketh[3], A. Wiedensohler[5], E.K. Chiloane[3,6], G. de Leeuw[2], and M. Kulmala[1]

[1]{Department of Physics, University of Helsinki, Helsinki, Finland}

[2]{ Finnish Meteorological Institute, Helsinki, Finland.}

[3]{Unit for Environmental Science and Management, North-West University, Potchefstroom, South Africa}

[4] {Finnish Meteorological Institute, Kuopio, Finland.}

[5] {Leibniz Institute for Tropospheric Research, Leipzig, Germany}

[6] {Eskom Holdigns SOC Ltd, Sustainability Division, South Africa}

Abstract

~~In this work satellite observations from the NASA's A-Train constellation were used to derive the values of primary emission and regional nucleation proxies over South Africa to estimate the potential for new particle formation. As derived in Kulmala et al. (2011), the satellite based proxies consist of source terms (NO_2, SO_2 and UV-B radiation), and a sink term describing the pre-existing aerosols. The first goal of this work was to study in detail the use of satellite aerosol optical depth (AOD) as a substitute to the in situ based condensation sink (CS). One of the major factors affecting the agreement of CS and AOD was the elevated aerosol layers that increased the value of column integrated AOD but not affected the in situ CS. 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. The second goal of the work was to estimate how well the satellite based proxies can predict the potential for new particle formation. For each proxy the highest potential for new particle formation were observed over the Highveld industrial area,~~

~~where the emissions were high but the sink due to pre-existing aerosols was relatively low. 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. Even though the OMI SO_2 boundary layer product showed reasonable spatial pattern and detected the major sources over the study area, some of the known minor point sources were not detected. When defining the satellite proxies only for days when new particle formation event was observed, it was seen that for all the satellite based proxies the event day medians were higher than the entire measurement period median.~~

Proxies for estimating nucleation mode number concentrations and further simplification for their use with satellite data have been presented in Kulmala et al. (2011). In this paper we discuss the underlying assumptions for these simplifications and evaluate the resulting proxies over an area in South Africa based on comparison with a suite of ground-based measurements available from four different stations. The proxies are formulated in terms of sources (concentrations of precursor gases (NO_2 and SO_2), and UV-B radiation intensity near the surface), and a sink term related to removal of the precursor gases due to condensation on pre-existing aerosols. A-Train satellite data are used as input to compute proxies. Both the input data and the resulting proxies are compared with those obtained from ground-based measurements. In particular a detailed study is presented on the substitution of the local condensation sink (CS) with satellite aerosol optical depth (AOD) which is a column-integrated parameter. One of the main factors affecting the disagreement between CS and AOD is the presence of elevated aerosol layers. Overall, the correlation between proxies calculated from the in situ data and observed nucleation mode particle number concentrations (N_{nuc}) remained low. At the time of the satellite overpass (13-14 LT) the highest correlation is observed for SO_2/CS (R^2=0.2). However, when the proxies are calculated using satellite data, only NO_2/AOD showed some correlation with N_{nuc} (R^2=0.2). This can be explained by the relatively high uncertainties related especially to the satellite SO_2 columns and by the positive correlation that is observed between the ground-based SO_2 and NO_2 concentrations. In fact, results show that the satellite NO_2 columns compare better with in situ SO_2 concentration than the satellite SO_2 column. Despite the high uncertainties related to the proxies calculated using satellite data, the proxies calculated from the in situ data did not predict significantly better $N_{nuc.}$ Hence, overall improvements in the formulation of the proxies are needed.

1 Introduction

Aerosol particles are key constituents in the Earth-Atmosphere system that can alter climate through their direct and indirect effects on the Earth's radiation budget. Aerosols affect the radiation budget directly by scattering and absorbing solar radiation, and indirectly by acting as cloud condensation nuclei or ice nuclei and modifying clouds' radiative properties and lifetimes. However, the quantification of the aerosol effects on climate is very complex and large uncertainties still exist due to the high spatial and temporal variability of aerosol mass and particle number concentrations (e.g. IPCC, 2013). Besides the climatic effects, aerosols affect human life by reducing the air quality and visibility as well as affecting human health especially in urban areas. Particulate air pollution has been associated with adverse cardiovascular and pulmonary diseases, and even with rises in the numbers of deaths among older people (e.g. Seaton et al., 1995 Uttel et al., 2000, Schnelle-Kreis, 2009).

Primary aerosol particles are emitted directly ~~to~~into the atmosphere; e.g. sea spray aerosol, desert dust, aerosol generated from biomass burning~~,~~ and fossil fuel combustion ~~or re-suspended dust from disturbed or un-vegetated soil~~. Secondary particles are ~~on the other hand nucleated via~~formed from precursor gases through gas-to-particle conversion. The ~~key phenomena associated with secondary atmospheric aerosol system are the~~ formation of new ~~nanometer-size aerosol~~ particles ~~and their subsequent growth to climatically relevant sizes (D_p ~ 100 nm), i.e. to sizes when their direct and indirect contribution to Earth's radiation budget becomes significant. The atmospheric aerosol formation~~ is strongly connected to the ~~formation~~presence of sulphuric acid and other vapours of very low volatility, as well as the magnitude of solar radiation (e.g. Kulmala et al., 2008, Kulmala et al., 2005). On the other hand pre-existing aerosol particles act as a sink for the vapours inhibiting new aerosol formation (e.g. Kulmala et al, 2008). These new nanometer-size aerosol particles grow through condensation and coagulation to sizes where they may act as cloud condensation nuclei (particle diameter D_p ~ 50 nm) or where they are large enough (D_p > ~ 100 nm) to scatter solar radiation and thus affect the Earth's radiation budget.

Several studies have shown that nucleation occurs frequently in the continental boundary layer and free troposphere from clean to polluted environments (Kulmala et al., 2004, Kulmala et al. 2008 and references therein). Laakso et al. (2008) and Vakkari et al. (2011) have studied new particle formation over moderately polluted savannah ecosystems in South Africa and found that nucleation takes place in the boundary layer almost every sunny day throughout the year with a frequency of as high as 69% of all ~~analyzed~~analysed days

(Vakkari et al. (2011)). Hirsikko et al. (2012) extended the studies in South Africa to a polluted measurement site and found an even higher frequency for the nucleation event days (86%), which is among the highest event frequencies reported in the literature so far. Hirsikko et al. (2013) also studied the causes for two or three consecutive daytime nucleation events, followed by subsequent particle growth during the same day. They concluded e.g. that the multiple events were associated with SO_2 rich air from industrial sources.

Satellite instruments have been providing global observations of the Earth's atmosphere for three decades (e.g. Lee et al., 2009, Kokhanovsky and de Leeuw, 2009~~).~~., Burrows et al., 2011). Information about the spatial distribution of aerosols and trace gases can be obtained from multiple instruments with various temporal and spatial resolution and coverage. Passive remote sensing instruments such as NASA's Ozone Monitoring Instrument (OMI) onboard the AURA platform or the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua platforms use ~~the~~ solar radiation to detect either trace gases or aerosol and cloud properties. ~~The trace~~Trace gas remote sensing techniques using OMI are based on the trace gas absorption features in the UV-region (wavelength λ ~ 200-400 nm), whereas the remote sensing of aerosol particles is mainly based on measurements in the UV/visible and near infrared regions (λ ~ 500-2000 nm). Since the aerosol measurements ~~utilizes~~utilize only the optically active size range of the solar spectrum, the detectable aerosol sizes are limited to particles ~~diameter~~with diameters greater than about 100 nm. Nucleation mode particles (smaller than about 25-30 nm in diameter), therefore, cannot be detected directly using ~~the~~ satellite instruments. In 2011 Kulmala et al. introduced proxies, i.e. ~~satellite based~~ parameterizations for estimating the number concentrations of ~~nucleated particles smaller than about 25-30 nm~~nucleation mode (N_{nuc}) simplified for the use with satellite data. These simplifications were made assuming that in ~~diameter~~situ parameters could be replaced with satellite-based observations. Their study was the first attempt to estimate the global nucleation mode aerosol concentrations ~~based on~~using data derived from satellite measurements. The ~~satellite~~ proxies ~~are~~were defined ~~as the ratio of the source and sink~~in terms of sources and sinks. The nucleation source terms ~~consisted~~consist of ~~trace~~precursor gas column densities (NO_2 or SO_2) and UV-radiation ~~(~~intensity near the surface (all from OMI as opposed to in situ data in the initial proxies) whereas the sink ~~due to pre-existing aerosols was estimated using~~term, i.e. the condensation sink in the original proxy formulation related to the aerosol surface area concentration is assumed to be proportional to the aerosol optical depth (AOD, from MODIS). ~~-~~ More recently Crippa et al. (2013) formulated a new

proxy algorithm for ultrafine particle number concentrations based on satellite-derived parameters. They used multivariate linear regression approach to derive the proxy, where the source terms consisted of SO_2, UV (from OMI), and NH_3 (from Tropospheric Emission Spectrometer, TES). The sink term was formulated using MODIS (collection 5.0) AOD and the Ångström ~~coefficient~~exponent, which expresses the spectral dependence of AOD on the wavelength. However, ~~caution is needed if using the satellite-based Ångstrom coefficient especially over land. For MODIS collection 5~~ there ~~have been reported~~are issues with the Ångström coefficient (e.g. Mielonen et al., 2011), and thus ~~it~~this parameter is ~~not~~no longer included ~~anymore~~ in the most recent MODIS collection 6.0 land parameters (Levy et al., 2013).

In this work we evaluate the simplifications and underlying assumptions of the method introduced ~~at~~in Kulmala et al. (2011) ~~is used~~ to ~~determine the~~estimate the number concentration of nucleation mode particles from satellite ~~based proxies for estimating the potential for new particle formation over~~-derived data. The study area is the north-eastern part of South Africa (25-28S, 25.5 -30.5E, Figure 1.). Even though the ~~study~~ area is not very large, it comprises lots of contrasts from the emission point of view; the cities of Johannesburg and Pretoria, as well as highly industrialized areas especially east from the cities, ~~and on the other hand~~versus a very clean background in the western part of the study area. The study period considered ~~in this paper was~~is Jan 2007- Dec 2010. There are also four different measurement stations located within the region of interest, where observations of various in situ parameters were available. ~~This work comprises of two parts:~~

This work comprises of two parts:

1) A detailed investigation of ~~using~~replacing the ~~satellite based AOD to describe the sink, and comparison with the in situ~~ condensation sink (CS~~).~~, defined below in Eq. 8), a local parameter evaluated from in situ observations, with the AOD, a column-integrated aerosol property available from satellite.
2) ~~To estimate~~ The estimation of how well ~~the~~ satellite ~~based~~ data can be used to compute proxies ~~can predict the new~~for nucleation mode particle ~~formation over the study area~~number concentrations. This comprises ~~of~~ the analysis of both the satellite-~~-~~and in situ-based proxy components and the proxies, as well as the comparison of the proxies with the ~~in situ observations, including also the event classification data.~~

~~Even though the connection between the AOD and CS is very important, it has not been studied in detail before, starting from the theoretical definitions of the parameters and~~

~~continuing to the comparison of the~~measured concentration of nucleation mode particles. The influence of the uncertainties in ~~situ and~~the satellite~~ based observations. Also, the satellite based proxies have not been extensively compared with in situ measurements, and especially with the new particle formation event classification data.~~

2) -derived quantities on the proxy is also evaluated.

2 Data

In this study, a variety of data was used from satellite instruments and ground-based stations (see Table 1 for a summary). Satellite data used originate from NASA's Afternoon-Train (A-Train) constellation~~ was used~~. The A-Train constellation consists of seven satellites that are on a same polar-orbiting track and follow each other closely enabling near-simultaneous observations of a variety of atmospheric parameters. The equatorial ~~overpassing time~~overpass for the A-Train satellites is around 1:30 p.m. local time. In this study ~~the~~we use OMI ~~instrument aboard NASA's AURA satellite was used to get the NO_2 and SO_2 column densities as well as the amount of UV-B radiation. The measurement range of OMI covers spectral region from 264 to 504 nm, and the nominal spatial resolution is 13 x 24 km^2. Global coverage is achieved in one day. For this study the selected~~ Level 2 ~~OMI parameters were~~products, i.e. the NO_2 tropospheric column (Bucsela et al., 2013), the SO_2 planetary boundary layer (PBL) product (Krotkov et al., 2006, Krotkov et al., 2008), and ~~for UV-B~~ the 310 nm irradiance (UV-B) at surface at local noon (Tanskanen et al., 2006). It is noted that the OMI SO_2 PBL product ~~is a retrieval of~~ describes the SO_2 ~~throughout~~concentration integrated over the whole atmospheric column, and PBL refers to the a priori profile assumed in the retrieval of this product. The OMI L2 ~~parameters~~products are provided with a nominal spatial resolution of 13 x 24 km^2. For the current is study they were re-gridded ~~into~~onto a 3 km x 3 km geographical grid~~ and four year medians (2007-2010) were calculated~~ as in Fioletov et al. (2011). In this way the effective spatial resolution could be increased despite that the instrument resolution is coarser than the grid. For NO_2 and SO_2 only those observations were used where the (radiative) cloud fraction was below 20%.

~~The AOD was obtained from MODIS onboard Aqua platform, which~~According to Lamsal et al. (2014), and references therein, the uncertainty in the OMI NO_2 tropospheric column concentrations is ~~measuring key~~about $0.75 \cdot 10^{15}$ molec./cm^2, whereas Krotkov et al. (2008) report that the SO_2 PBL product could be associated with noise as high as 1.5 DU. However averaging the SO_2 columns over longer a period and/or over a larger spatial area could reduce the noise to 0.3-0.6 DU. For OMI UV-B irradiance the relative uncertainty is on

average 7%, but could be higher e.g. due to some episodic aerosol ~~and cloud properties in 36 spectral bands ranging from 0.4 to 14.0 μm. The spatial resolution of MODIS is 250-1000 m, depending on the measurement channel. The swath width is 2330 km, and global coverage is achieved~~ plumes (Tanskanen et al., 2006).

The AOD used in ~~one day. In~~ this study is the ~~recently released~~ MODIS Aqua collection 6.0 AOD product at 3 km spatial resolution ~~was used~~ (Levy et al., 2013). The relative uncertainty for the MODIS AOD over land is reported as 0.05+15%. For selected cases also ~~data~~vertical aerosol extinction profiles from ~~The~~the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) (Winker et al., 2007) ~~were~~are used ~~to get the vertical profiles of aerosol extinction.~~.

The in situ data ~~were~~used in this study are collected at four different stations in South Africa: Elandsfontein (ELA), Marikana (MAR), Botsalano (BOT), and Welgegund (WEL). All ~~the~~of these stations are located in the north eastern part of the country ~~and locations are~~ shown in Fig. 1. Depending on the station, the measured parameters included e.g. particle size distribution, extinction coefficient and trace gas concentrations. More detailed description of the in situ measurements can be found e.g. in Kulmala et al. (2011), Beukes et al. (2014), Hirsikko et al. (2012), Venter et al. (2012), Vakkari et al. (2011), Laakso et al. (2012), and Vakkari et al. (2013). ~~–~~Also data from the Aerosol Robotic Network (AERONET, http://aeronet.gsfc.nasa.gov, Holben et al., 1998) ~~was used~~ at the Elandsfontein station is used. AERONET is a global ground-based sunphotometer network, providing observations of aerosol optical, microphysical, and radiative properties that are available in a public domain. The aerosol optical properties in the total atmospheric column are derived from the direct and diffuse solar radiation measured by the Cimel sunphotometers. ~~Table 1 summarizes the data used in this work.~~

3 Proxies

Regional scale nucleation is associated with photochemistry, and typically ~~has~~occurs over a spatial scale of hundreds of kilometres (Kulmala et al., ~~(~~2011~~);~~ and references therein). The number concentration of nucleation mode particles on a regional scale can be estimated ~~as a ratio~~from consideration of ~~source term proportional to UV-radiation and~~ the sources, i.e. precursor gas concentrations such as sulphur dioxide ~~concentration~~(SO_2), the UV-radiation

intensity needed to initiate the photochemical reaction, and a sink term which reduces the nucleation potential by removing precursor gases (Petäjä et al., 2009):

$$P_{Nnuc,regional} = \frac{UV \cdot [SO_2]}{CS^2} \quad (1)$$

Primary emissions occur in the vicinity of local sources such as industrial or urban areas. For nucleation from primary emissions two proxies are defined as (Kulmala et al., (2011)):

$$P_{N_{nuc,prim.}} = \frac{[NO_2]}{CS}, \quad (2)$$

$$P_{N_{nuc,prim.}} = \frac{[SO_2]}{CS}, \quad (3)$$

In each of the proxies the source terms are estimated from the satellite measurements by replacing the SO_2 and NO_2 concentrations at the surface with the ~~total~~ column densities from the satellite. The amount of global UV radiation is also available from satellite measurements e.g. as a local noon irradiance at ~~surface at~~ 310 nm wavelength (UV-B-radiation~~). On~~) at the ~~other hand for~~ surface. For the sink ~~term there is no equivalent column integrated~~ parameter ~~from satellites. For satellite based sink parameter~~(CS), Kulmala et al. (2011) proposed ~~aerosol optical depth (~~to use the AOD~~) that~~ which describes the total aerosol extinction in ~~an~~the atmospheric column. The relation between ~~actual~~the CS and the AOD will be discussed in ~~more detail in~~ the following section. By replacing CS with AOD the simplified proxy for using satellite ~~based proxies can be expressed~~data for primary nucleation ~~as~~becomes:

$$P^{Sat.}_{N_{nuc}} = \frac{[NO_2]_{column}}{AOD} \quad (4)$$

$$P^{Sat.}_{N_{nuc.}} = \frac{[SO_2]_{column}}{AOD} \quad (5)$$

~~, and for~~For regional nucleation ~~as~~ the proxy expressed in terms of satellite data becomes

$$P^{Sat.}_{N_{nuc.}} = \frac{UV[SO_2]_{column}}{AOD^2} \quad (6)$$

In addition we also considered

$$P_{N_{nuc.}}^{Sat.} = \frac{UV}{AOD^2} \quad (7)$$

as a potential proxy for the number concentration of nucleation mode particles.

3.1 Condensation sink and aerosol extinction

As indicated in the previous section, Kulmala et al. (2011) proposed AOD as a substitute for CS. Both parameters are also roughly proportional to the aerosol surface area distribution. According to e.g. Lehtinen et al. (2003) the condensation sink is defined as

$$CS = 2\pi\rho_{diff}\int_0^\infty D_p\beta_M(D_p)n(D_p)dD_p \quad (8)$$

, where D_p is the particle radius, $n(D_p)$ is the particle number size distribution function, ρ_{diff} is the diffusion coefficient, and $\beta_M(D_p)$ is the transitional correction factor.

Aerosol optical depth describes quantitatively the column-integrated extinction of solar light caused by atmospheric aerosols and it is one of the standard aerosol parameters that is retrieved from the satellite radiance observations. At ~~some constant~~a height z and for a wavelength λ the aerosol extinction is defined as

$$\sigma_{ext,z,\lambda} = \frac{1}{4}\pi\int_0^\infty Q_{ext}(\lambda, D_p, m)D_p{}^2 n(D_p)dD_p \quad (9)$$

, where Q_{ext} is the extinction efficiency describing aerosols ability to scatter and absorb solar light. At a fixed wavelength the extinction efficiency is a complex function of aerosol size and complex refractive index m (which in turn depends on the aerosol particle composition). Also the particles shape affects somewhat on Q_{ext}, but this is not considered in this study. If the particles are assumed to be spherical, Q_{ext} can be calculated using a computer code based on the Lorenz-Mie theory (Mishchenko et al., 2002). AOD is obtained by integrating σ_{ext} over the total atmospheric column.

~~If the satellite retrievals provided the aerosol size distribution, CS could be derived from the satellite observations. Since this is not the case, the sink of existing aerosols is estimated using the column integrated value of σ_{ext}, i.e. AOD. Fig. 2 illustrates the theoretical~~The differences between CS and σ_{ext} (at a ~~constant atmospheric level~~certain height) as a function

of particle size are illustrated in Fig. 2. Both parameters are derived using the same aerosol size distribution (Fig. 2, left panel). The σ_{ext} is calculated using a refractive index of m=1.48+0.003i and wavelengths of 0.55 and 0.45 μm. As Fig. 2 shows, particles with D_p about 0.05-0.1 μm have the largest contribution to CS, whereas for σ_{ext} the largest contribution is coming from particles with D_p about 0.2-0.8 μm. The notable difference between the two quantities is that particles $D_p < 0.1$ μm can have a contribution to CS which is several orders of magnitude larger ~~contribution to CS~~ than that to σ_{ext}. On the other hand, σ_{ext} is significantly more sensitive to ~~the~~ particles with $D_p > 1.0$ μm than CS. It is clear that e.g. a large change in number concentration of the smaller particle sizes would change the value of total CS when integrated over the size distribution, but would have ~~only~~ a minor effect on the value of σ_{ext}, and vice versa, if e.g. the number concentration of large particles increased there would be little effect on CS. It is noted that in addition to the theoretical differences the possibility of elevated aerosol layers affect the column integrated values of σ_{ext}, i.e. the AOD, which must be considered when comparing the satellite based AOD with in situ CS.

The ~~sensitivity~~response of σ_{ext} on changes in the particle ~~sizes~~size distribution depends to a certain extent on the particle composition and the measurement wavelength. If the particle absorption is ~~significantly~~ high (i.e. the imaginary part of m ~0.1i), the contribution of particles $D_p < 0.1$ μm to σ_{ext} would be somewhat higher than in Fig. 2. Shorter wavelengths increase the sensitivity to smaller particles, but as Fig. 2 illustrates, a 0.1 μm decrease in wavelength does not improve the sensitivity significantly. Much shorter wavelengths would be needed to increase the sensitivity of σ_{ext} to particles $D_p < 0.1$ μm, but such measurements could not be carried out in a real atmosphere.

4 Results

The proxies as defined in Sect. 3 are formulated in terms of parameters which are either obtained from ground-based in situ measurements (Eqs. 1-3) or from satellite data (Eqs. 4-7). In this section the performance of these proxies is critically evaluated and in particular each of the satellite-based parameters is critically examined.

4.1 Comparison of condensation sink and aerosol optical depth

Replacing CS with AOD is perhaps the most crucial assumption when determining the proxies using satellite data, as indicated in Kulmala et al. (2011). ~~The theoretical~~ Apart from the sensitivity of these parameters for different particle sizes discussed in Section 3.1, other differences ~~between~~play a role such as the vertical variation of the aerosol ~~extinction~~concentrations, the particle size range considered and ~~CS are illustrated in Fig. 2, but when comparing satellite AOD with CS other issues might arise related to for example, the in situ measurement devices and~~the dependence of aerosol ~~vertical distribution.~~particle size on relative humidity. CS is determined from ~~the~~ measured dry particle size ~~distribution~~distributions with a correction for ambient humidity. CS at Botsalano and Marikana has been estimated from submicron size distribution while at Elandsfontein ~~the~~ size ~~distribution~~distributions up to 10 µm ~~was~~were used. ~~On the other hand, as~~ In contrast, the AOD is an integrated quantity ~~through the total atmospheric column, it has contribution~~with contributions from all optically active aerosols ~~at all~~throughout the whole atmospheric ~~levels~~column. To assess the effect of these different factors on the relation between the AOD and CS, the following comparisons ~~were carried out~~are made:

1) In situ CS ~~and~~with nephelometer aerosol scattering coefficient
2) In situ nephelometer aerosol scattering coefficient ~~and column integrated~~with AOD from AERONET
3) In situ CS ~~and column integrated aerosol extinction~~with AOD from both AERONET- and satellite measurements.

Coincident measurements of ~~situ~~ size distributions to derive the CS and ~~in situ~~aerosol scattering coefficients from a nephelometer ~~were~~are only available ~~at~~from the Elandsfontein measurement station. The comparison between CS and scattering coefficient ~~was intended~~serves to ~~mainly illustrate the theoretical differences in CS and~~eliminate effects of the vertical variation of the aerosol ~~extinction because both measurements were carried out at ground level and elevated aerosol layers would not affect~~concentrations on the comparison. The nephelometer measures the dry particle scattering at 0.525 µm wavelength and the results are presented at ~~a~~ Standard Temperature and Pressure (STP) atmosphere. ~~Also the~~The maximum particle size ~~was~~is limited to ~~a~~ D_p ~ 10 µm. It ~~should be~~is noted that the nephelometer considers only aerosol scattering, and not the total extinction which would also require ~~the~~ information ~~about~~on absorption. However, the contribution of absorption to the total aerosol extinction is generally much smaller than scattering. Laakso et al. (2012)

reported that at Elandsfontein the absorption was increased during the coldest months (May-Oct.) due to ~~the~~ biomass burning ~~season and,~~ domestic burning of coal for heating ~~as well as~~and cooking, contributing about 15-20% ~~of~~to the total aerosol extinction whereas during the warmer months (Nov.-Apr.) absorption contributed ~10% of the total aerosol extinction. ~~Due to~~To take the seasonal variation of absorption into account, the ~~comparison between one hour mean~~ CS and the scattering ~~coefficient was carried out~~coefficients were compared separately for the periods May-Oct. and Nov.-Apr~~. For~~. The results in Fig. 3 show that for both periods ~~significant correlations between in situ~~ scattering coefficients and CS were ~~found:~~ well-correlated with R^2=0.~~82 (p<0.001)~~67 for Nov.-Apr~~,~~., and R^2=0.~~84 (p<0.001)~~71 for May-Oct. ~~(Fig. 3).~~ The ~~correlations~~R^2 values were higher than ~~what has been reported at~~those from measurements at a clean continental boreal forest measurement site in Hyytiälä, Southern Finland~~, by~~ (R^2=0.38, Virkkula et al~~.~~ (., 2011). ~~They obtained a correlation coefficient of 0.62 between CS and the scattering coefficient.~~

The next step ~~was~~is to compare the nephelometer scattering coefficient to the ~~column integrated aerosol extinction~~AOD to ~~find out how much~~evaluate effects of the possible occurrence of elevated aerosol layers and/or boundary layer mixing ~~might affect the comparison~~. Also the presence of large dust particles might have some effect on the comparison due to the limited particle size in the nephelometer inlet. In this comparison we first compare with AERONET measurements of AOD at Elandsfontein, which are more accurate than those retrieved from satellite data. As Fig. 4 shows, ~~lower~~the correlations ~~were obtained~~ between ~~column integrated~~the AERONET AOD and the in situ scattering coefficient (~~correlation coefficient~~ warm season~~;~~ R^2=0.~~68~~46, cold season~~;~~ R^2=0.~~49)~~24) are lower than ~~what were obtained~~those between the ~~in situ~~ CS and the scattering coefficient. This indicates that the elevated aerosol layers and boundary layer mixing might affect more than the theoretical differences when estimating the sink of pre-existing aerosols by using the AOD.

~~Finally~~For the ~~in situ~~comparison of CS ~~was compared~~ with the AOD retrieved from MODIS ~~AOD. The~~, daily AOD values ~~used in the comparison~~ were ~~the~~used which are spatial averages ~~calculated from~~ of the observations within 3 km radius ~~around the~~from each measurement station. As Fig. 5 shows, the CS- vs. satellite AOD data ~~were~~are scattered all over the graph and although there is a tendency of increasing CS with increasing AOD there is no apparent correlation ~~was lower than in the previous comparisons in Figs. 3 and 4, varying between (~~0.~~17 and 03~~≤ R^2 ≤0.~~25 depending on the location. In addition, the standard

~~least-squares method (LSQ) did not seem to give a proper slope for the coincident CS and AOD values. Therefore~~06). As an alternative, a bivariate method ~~by (~~York et al~~. (~~., 2004) was applied ~~that accounts~~to account for the uncertainties associated to both CSs and MODIS AODs in the fitting. For CS the uncertainty was assumed to be 10~~%, which should be a reasonable estimate referring to the work by~~ % (Petäjä et al~~. (~~., 2013~~). For~~) and for MODIS AOD an uncertainty of 0.05+15% was used ~~as suggested in (~~Levy et al~~. (2013). ~~., 2013). This means that for low AOD the relative uncertainty is rather high, e.g. for AOD=0.1 the relative uncertainty would be 65%. As Fig. 5 shows the bivariate method gave very different ~~slopes than LSQ at each station for the relationship between CS and AOD, which might be partly related to the differences in the size distribution upper limits at Elandsfontein and Marikana/Botsalano. If the coincident CS-AOD observations at all three stations were combined, the bivariate fit gives y= 0.194x-0.008~~results than LSQ.

At Marikana and Elandsfontein the largest observed AODs ~~were~~are not related to lagest CS, which could ~~indicate~~be due to the presence of elevated aerosol layers. ~~At Marikana the median MODIS AOD was 0.15 for the whole measurement period, and as Fig. 5 shows, the CS values were less scattered when AODs were below the median. Also the correlation coefficient for coincident CS and AOD ≤ 0.15 was higher (R=0.4, p<0.001) than for the whole range of observations. To estimate how much the~~In a recent study by Giannakaki et al., (2015) data from a ground-based lidar at Elandsfontein are analyzed and the results show that the mean contribution of elevated aerosol layers ~~might affect the~~ to the AOD is 46%. To estimate the effect of elevated aerosol layers on the CS-AOD comparison at Marikana, CALIPSO observations of aerosol vertical extinction profiles ~~were studied~~are used. All ~~the~~ CALIPSO daytime overpasses between 8.2.2008 and 17.5.2010 within 50 km from the Marikana station were considered. ~~Since~~Due to the small CALIPSO swath width ~~is narrow~~ only 48 days of data ~~were obtained. The~~are available. At Marikana the median MODIS AOD is 0.15 for the whole measurement period, and as Fig. 5 shows, the CS values are less scattered when AODs are smaller than the median. Therefore the vertical aerosol extinction profiles from CALIPSO ~~were~~are studied separately for the cases where MODIS AOD ≤ 0.15 and AOD > 0.15. As Fig. 6 shows, for higher AODs the median extinction profile ~~indicated~~indicates an elevated aerosol layer, which supports the result that high AODs also at Marikana are likely to be associated with an elevated aerosol layer.

4.2 ~~Spatial~~ Proxies defined from the in situ data and ~~seasonal characterization of the satellite~~ comparison with N_{nuc}

The proxies are first computed using in situ measurements from Marikana and Elandsfontein following Eqs. 1-3 to evaluate how well each of them could predict the nucleation mode number concentration within our study area. It is noted that due to different instrumentation N_{nuc} from Marikana consists of particles with D_p < 30 nm, but at Elandsfontein N_{nuc} consists of particles with D_p 10-30 nm. In addition, CS at Marikana is defined from submicron particles whereas at Elandfontein CS is defined from particles with D_p<10 µm.

Figure 7 shows the diurnal variation of each of the in situ proxy components and the number concentration of nucleation mode particles. At Marikana the N_{nuc} median peaks about 10 a.m., and at Elandsfontein about an hour later. At the time of the satellite overpass the median of N_{nuc} is lower than before noon at both locations, and about the same order of magnitude. The diurnal variation of NO_x-NO and SO_2 concentrations show somewhat different characteristics at Marikana than at Elandsfontein, The morning and evening peaks of NO_x-NO at Marikana are most likely associated with household combustion and traffic whereas the single SO_2 peak in the morning is most likely related to the industrial emissions and the break-up of the inversion layers that form quite regularly in the South African Highveld (Venter et al., 2012). At Elandsfontein, where the major emission source is heavy industry, an increase in the NO_x-NO and the SO_2 concentration medians are seen about 10 a.m. The median of SO_2 concentration decreases in the late afternoon while the median of NO_x-NO concentration does not vary much. At the time of the satellite overpass the NO_x-NO and SO_2 medians are much higher at Elandsfontein than at Marikana. Results show also that at the time of the satellite overpass NO_x-NO and SO_2 are positively correlated; at Elandsfontein R^2=0.58, and at Marikana R^2=0.32 are obtained. At Elandsfontein CS does not show any clear diurnal variation and it is systematically lower than at Marikana. Also at Marikana the diurnal variation of the CS is rather weak during the daytime but a peak in the median is seen in the evening.

Figure 8 shows the diurnal variation of the in situ proxies at Marikana and Elandsfontein. The comparison of the diurnal variation of the proxies and N_{nuc} indicates that the proxy-N_{nuc} relation depends on the time of the day. At the time of the satellite overpass (13-14 LT) the highest correlation with N_{nuc} at Marikana is obtained with the SO_2/CS-proxy (R^2=0.22, Fig.9), but at Elandsfontein the correlation remains below 0.1. At Marikana the correlation of N_{nuc} with $SO_2 \cdot UV./CS^2$ - proxy (Eq. 1) is less good at the time of the satellite overpass but at 9-10

a.m. R^2 =0.25. On the other hand the (NO_x-NO)/CS and UV/CS^2 proxies do not perform well in predicting N_{nuc}. Also, it is noted that at the time of the satellite overpass all the proxy values show much higher median values at Elandsfontein than at Marikana while the median for N_{nuc} is about the same at both locations. At Elandsfontein somewhat better correlations with N_{nuc} are observed if only the source terms of the proxies are considered. For example, the values of R^2 between N_{nuc} and SO·UV are 0.35 at 10-11 LT, and 0.14 at 13-14 LT, respectively, but when the sink-term CS^2 is included in the proxy there is no correlation. At Marikana CS doesn't have as high influence on the proxy performance as at Elandsfontein.

The difference with the results reported for Southern Finland (Kulmala et al. (2011)) is that in our study SO_2 has a strong effect on the performance of the proxy: without SO_2 the UV/CS^2 – term does not correlate with N_{nuc}. Given that the satellite data are associated much higher uncertainties than the in situ measurements, these in situ-based results can be considered as some kind of upper limits for the overall performance of the proxies computed using satellite data (Eqs. 4-7).

.

4.3 Proxies using satellite data

4.3.1 Spatial pattern of the satellite-based proxies

Each of the satellite based parameter ~~was~~is analyzed from Jan. 2007 to Dec. 2010. ~~Fig. 7~~Figure 11 shows the four year medians of SO_2 and NO_2 column densities obtained from the OMI ~~measurements~~instrument, as well as the AOD at 550 nm from MODIS Aqua observations. ~~The NO_2 and SO_2 column medians are calculated from data, where more than 80% of the photons originate from the cloud free area of the pixel, whereas the AOD represents only cloud free cases. The clear sky noon UV-B irradiance depends on the season and hence has only small spatial variation at the time of the satellite overpass during the year. Highest amount of UV-B radiation is obtained naturally during summer (December-February), and lowest during winter (June-August).~~

~~The use of long time series and fine 3 x 3 km~~ Daily satellite data is used to define the satellite-based proxies over the study area (Eq. 4-7). Figure 12 shows the four year median spatial patterns for the four satellite-based proxies. The spatial ~~grid reveals OMI "sub-pixel"~~ patterns of these four proxies are quite different and in particular there is large difference between the spatial ~~patterns from the satellite data. Some similarities in the distribution of~~ variation of the regional proxies and that of the proxies for nucleation from primary

emissions. As expected, the latter strongly reflect the spatial distributions of the precursor gases with high concentration over the Highveld industrial area, where the values of NO_2 and SO_2 columns are high and the sink (AOD) is low. For the NO_2 ~~and SO_2 column densities could be observed. Both parameters showed the highest median values over the Highveld Mpumalanga industrial area east from~~/AOD proxy also elevated values are observed over the Johannesburg ~~and Pretoria (around the Elandsfontein measurement station), where most of South Africa's coal-fired power station, as well as some metallurgical smelters and a large petrochemical plant are situated (Lourens et al., 2011). None of these large point sources currently apply de-SO_x or de-NO_x technology. For NO_2 the maximum median values south of Elandsfontein were about 10 times larger than over the background areas in the west of the study area. Elevated values for the NO_2 column density were also observed over Johannesburg as was previously indicated by Lourens et al. (2012) as well as over the Vereening area where another large coal-fired power station, numerous petrochemical operations and various metal smelters are located. The maximum medians for SO_2 PBL column densities (about 2.5 DU) were observed over area with several metallurgical smelters, North West from Elandsfontein. A local hotspot of SO_2 was also observed over the Vereening area. However, in contrast to the NO_2 values, the SO_2 column densities over Johannesburg were not notably higher than over the background areas. Overall the SO_2 column densities outside the two clear hotspots had little variation, which is different to the spatial pattern of greenhouse gas emission inventory data shown in Vakkari et al. (2011). It seems that the satellite data is unable to detect some of the weaker well known SO_2 sources e.g. over the Rustenburg area. As Fig. 7 illustrates the spatial pattern for AOD was different from the other two parameters. Overall the four- year median AOD was low (approximately 0.1) over major part of the study area. Highest medians (approximately 0.25) were observed over Johannesburg and Pretoria area, but somewhat elevated values were also observed over Rustenburg region that closely resembled the shape of the southern portion of the western Bushveld Igneous Complex (wBIC) geological deposit. Nine ferrochromium, platinum and base metal smelters occur within a 40 km radius within this southern part of the wBIC. Each of these smelters is loaded with ores from at least two different mines operating within the wBIC. A similar concentration of smelters also occurs in the area coinciding with the northern part of the SO_2 hotspot identified over the Mpumalanga Highveld (Fig. 7c). However, these smelters in the Mpumalanga Highveld are fed with ores that are not mined in their immediate vicinity. From the difference observed in AOD between these two areas with high metal smelter densities it can be deduced that the higher AOD measured in the area near~~

~~Rustenburg is due to larger particles being emitted from mining activities. All the smelters in South Africa apply bag filter or wet venturi scrubbing to remove particulate matter, therefore the smelters do not make significant contributions to the AOD. All the coal-fired power stations and petrochemical operations also apply off-gas cleaning technologies, and hence relatively low four-year AOD medians (approximately 0.1–0.15) were also observed over the heavily industrialized Mpumalanga Highveld area where the highest NO_2 and SO_2 column densities were observed~~ Pretoria area while for the other proxies a local minimum occurs over these cities.

~~Fig. 8 illustrates the seasonal anomalies (season median – four-year median) of~~ All the four satellite proxies show larger values at Elandsfontein than at Marikana, which is consistent with the results obtained for the in situ proxies. Based on the in situ results the SO_2-related proxies are expected to predict N_{nuc} at the time of the satellite overpass better than the other proxies. Comparison of the spatial patterns of each proxy calculated using satellite data in the vicinity of the in situ measurement stations shows that there are not very much difference between the spatial pattern of SO_2- and NO_2-related proxies.

The propagation of relative uncertainty associated with the proxies using satellite data can be estimated by comparing the uncertainties related to each satellite parameter (Sect. 3) and the observed median values shown in Fig. 11. For example, over background areas where both AOD and SO_2 are low, the SO_2 ·UV-B /AOD² -proxy can have an uncertainty of over 90%. On the other hand, over source areas where both NO_2 and AOD are slightly elevated the NO_2/AOD proxy would have an uncertainty of about 50%. Generally over South Africa the uncertainty in satellite-based proxies is high, especially over areas where both low values of NO_2, SO_2 and AOD are frequently observed.

4.3.2 Comparison of satellite and in situ proxy components

Before evaluating the performance of the proxies using satellite data, first the quality of the parameters used in these proxies should be examined. The CS/AOD comparison was discussed in Sect. 4.1. Here we compare satellite data for NO_2, SO_2 and ~~AOD. Overall it seems that the NO_2 and SO_2 have somewhat similar seasonal variation, but over Highveld they seem to have opposite anomalies during winter and summer. AOD also varies in a different phase. One of the major factor affecting the trace gas and AOD seasonal variation is the large scale biomass burning season between June and September- mid-October, which~~

~~e.g. causes the NO_2 column peak during winter (JJA) over major part of the study area. During winter months an anti-cyclonic flow pattern dominates the Highveld area and the pollutants are re-circulated, resulting in a build-up of pollutants originating from all sources. In addition, domestic burning of coal for heating contributes to the increase of primary pollutants while prevailing stable meteorological conditions as well as the lack of precipitation enhances the effect (Laakso et al., 2012, and references there in).~~

~~The AOD values increased significantly during spring (SON). The later peak of AODs is likely to be partly related to the long range transported aerosols from the large scale biomass burning in southern Africa. Maritz et al. (2014) indicated that the peak in biomass burning within an approximate 1000 km radius around the study area considered in this paper is in September. The large scale fires intensify during June-August in central Africa and shift southward during spring. It is also seen recognisable how the NO_2 anomaly decreases over the northern part of Mpumalanga and the Johannesburg-Pretoria areas as the anti-cyclonic transport of pollutants weakens.~~

~~During summer the humidity is relatively high and it forms part of the rainy season in the studied region. Wet deposition of pollutants is therefore more efficient than during the dry season (May-Oct). Therefore the lowest column densities of NO_2 are observed between December and February. Also SO_2 column densities are generally lower than the four years median during summer. However, over Mpumalanga and the eastern part of Gauteng an increase was observed. During summer the AOD is still larger than the four year median. This might partly be explained by the prevalent wet and cloudy meteorological conditions that can affect the sampling and quality of the AOD data. During Autumn (March-May) the NO_2 and SO_2 column densities increase over Gauteng and western Mpumalanga.~~

~~4.3 Comparison of the satellite data and~~ UV-B with in situ ~~measurements~~

~~In order to evaluate how well the satellite based proxies actually can describe the potential for new particle formation in the boundary layer, a number of comparisons with the in situ data were carried out.~~data at each of the measurement stations. The satellite data for each station ~~was~~is collected within a 12 km (NO_2, SO_2, UV-B) or a 3 km (AOD) radius from the station ~~location. The in situ data were~~ and the results are compared with hourly means of the in situ data extracted between 13-14 ~~local time~~LT, i.e. ± 30 min within the approximate ~~A-Train satellites overpass time. In Sec. 4.1 it was shown that in some cases the sink of pre-existing~~

~~aerosols within the boundary layer could be overestimated when using AOD as a substitute to CS due to elevated aerosol layers. However, when the satellite based source terms were compared with the in situ data, significant correlations were found between the satellite NO_2 column densities and the in situ NO_x-NO as well as the satellite UV-B irradiance and the in situ global radiation at each measurement station. The highest correlation for NO_2 were obtained at Marikana ($R=0.72$, $p<0.001$), and lowest at Elandsfontein ($R=0.51$, $p<0.001$). For UV-B and global radiation the correlations were between 0.78-0.88 ($p<0.001$). In Kulmala et al. (2011) the satellite based SO_2 was not included in the proxy analysis because the quality of the data was poor~~satellite overpass.~~. In this study the middle -troposphere SO_2 data was replaced by the OMI boundary layer product (Sec. 3), which significantly improved the quality of the data and the spatial variation thereof (Fig. 7). However, for all the stations lower correlation between the satellite and in situ based SO_2 measurements were obtained than for the other source parameters. The highest correlation was obtained at Welgegund ($R=0.47$, $p=0.002$), but at Marikana there was practically no correlation at all ($R=0.1$, $p=0.16$) which indicates that the satellite is not able to present the SO_2 sources over the Rustenburg area as already discussed in Sec. 4.2.~~

The satellite NO_2 column densities and the in situ NO_x-NO concentrations are reasonably well correlated as are the satellite UV-B irradiances and the global radiation measured at each station. The highest correlation for NO_2 were obtained at Marikana ($R^2=0.55$), and lowest at Elandsfontein ($R^2=0.26$). For UV-B and global radiation the correlations were $0.61 \leq R^2 \leq 0.77$. In Kulmala et al. (2011) a constant value was assumed for the satellite-based SO_2 when defining the global proxy maps, because the SO_2 product they used (middle tropospheric SO_2) did not show reasonable spatial pattern. In this study the middle -troposphere SO_2 data was replaced by the OMI boundary layer product (Sec.

~~4.4 Satellite based proxies~~

~~The proxies for primary emissions and regional nucleation were defined for each season between 2007 and 2010 using the equations introduced in Sec. 2. Fig. 9 shows the four year median spatial pattern for the satellite based proxies. The highest potential for primary and regional nucleation was found over the Highveld industrial area, where the values of NO_2 and SO_2 columns were high but the sink (AOD) was low. For the NO_2/AOD proxy also elevated values were observed over the Johannesburg-Pretoria area while the other proxies showed local minimum over these cities. Even though the wBIC is one of the major pyrometallurgical complexes in South Africa, the observed proxy values were low when compared to values~~

over Highveld industrial area. The SO_2 related proxies and UV-B/AOD^2 even showed local minima over the Marikana-Rustenburg area, due to the higher AOD values (Fig. 7a).

The proxy values from both satellite observations and in situ measurements were defined for Elandsfontein, Botsalano, and Marikana measurement stations. The correlation between the satellite and the in situ proxies were calculated after converting the proxy values into log-scale (Table 2). A relatively high correlation for the NO_2/AOD proxy was obtained for Marikana and Elandsfontein, but at Botsalano the agreement was weak. This might be partly related to the low number (N=9) of coincident satellite and in situ proxy observations. In addition the satellite AOD was extremely low at Botsalano (about 0.05), and it varied relatively more than the CS values, which resulted in stronger satellite proxy variation. This indicated that the satellite based proxies might show too high variability (noise) over background areas where the AOD is very low and close to the satellite detection limits. Also the UV-B/AOD^2 showed reasonable correlation with the in situ data while the correlation for the SO_2 related proxies was rather low at all of the stations. This was expected because of the relatively low correlation between the satellite and in situ SO_2.

The satellite proxy values were also compared with the actual number concentration of nucleation mode particles (N_{nuc}, D_p<30 nm) at stations where the DMPS measurements were available. Fig. 10 presents the comparison of satellite based NO_2/AOD and N_{nuc} at the Marikana station, where the number of coincident satellite proxy and number concentration observations were the highest. The correlation ($R=0.31$, $p<0.001$) was higher than for the other satellite based proxies, for which there were practically no correlation at all (Table 3). At Welgegund and Botsalano all the correlations with the nucleation mode number concentrations and satellite proxies were below 0.1.

To test if the correlation between N_{nuc} and satellite based proxies could be improved, AOD was replaced with a sink estimate obtained from the York method CS-AOD linear fit (Sec. 4.1). Even though some improvement in the correlations between satellite based proxies and N_{nuc} were obtained at Marikana station (Table 3), the correlations remained overall rather low.

4.5 Satellite based proxies and event day classification

The event day classification scheme by dal Maso et al. (2005) divides days into three main categories; event, undefined, and non-event days, based on the investigation of particle size distribution measurements. The event days can be further categorized into different sub-

classes. The nucleation event classification was available at Marikana, Botsalano and Welgegund measurement stations. The event classification data was used to define the satellite based proxies at the station and the surroundings based on the stations event and non-event days. Unfortunately there were no temporally overlapping event classification data for the three stations, and hence the study was made separately for each station. As already showed in Vakkari et al. (2013) and Hirsikko et al. (2012), the nucleation event frequency is very high in South Africa. Table 2 summarizes the number of event- and non event days defined at the three stations between Jan. 2007 and Dec. 2010. In this study the different event subclasses were not considered, and the days were only classified as event, non-event, and undefined. Also, the satellite data were not considered from those days when the event classification was not available at the station.

In an event day N_{nuc} is high, and hence, it is also expected that the satellite proxy values should be elevated. As Table 4 shows, the difference between the number of event and non-event days was significant at the measurement stations, e.g. at Marikana only two days were classified as non-event during the study period. When considering the satellite based proxies at the stations it was found that all the non-event days were cloudy, and hence the satellite based proxies could not be determined for those days at the stations. Also event and undefined days consisted some cloudy days, but there were also days when satellite data were available. In the majority of the cases when the proxy values could not be determined, the reason was that the MODIS cloud screen procedure had declared pixels cloudy and hence the AOD was missing. Fig. 11 illustrates the satellite proxy medians, and differences between the event and non-event days based on the event classification carried out at Marikana station. For event days slightly higher proxy medians were observed at Marikana, even though non-event proxies did not have any contribution to the whole measurement period median. More pronounced positive event day proxy anomalies were seen on the south side of the station. For non-event days lower median values were observed over the cloud-free areas where the proxies could be determined. Nearest non-event day observations were about 45 km to the west of Marikana. It is noted that these differences between the measurement period and event day medians were detected only when looking data over longer time period. Based on the daily values of the satellite proxies it was not possible to define whether there was an event at the station or not. At least some non-event day data from the satellites would have been needed to study the differences on a daily basis between events and non-events.

~~The OMI UV-B irradiance at surface was the only satellite based data included in the proxies that could be obtained for both cloudy and cloud-free days. From the OMI irradiance retrievals it is also possible to define a cloud modification factor (CMF) which is the ratio of observed UV-B irradiance and modelled clear-sky UV-B irradiance at the surface. In practice CMF describes how much cloudiness alters the UV-B radiation that reaches the surface. On cloud- free days CMF =1.0, but on cloudy and overcast days CMF values decrease depending on the spatial cloud coverage and thickness of the cloud deck. Since UV-B radiation plays a central role in new particle formation events (e.g. Hirsikko et al. (2012), Petäjä et al. (2009)), CMF was retrieved for Marikana and Botsalano stations and divided into three classes according to the event classification data. Fig. 12 illustrates the distribution of CMF in event, undefined, and non-event days. In each of the three classes the CMF was most often between 0.95 and 1.0 indicating either clear skies or some scattered cumulus clouds. This CMF data indicated that during non-event days the proportion of CMF < 0.7 might be slightly higher than in the two other categories. However, it should be noted that the number of CMF observations between event and non-event days differed considerably (768 vs. 20) which explains e.g. the gaps in the non-event CMF distribution.~~

3), which improved the characterization of the SO_2 spatial variation (Fig. 10). However, the relative uncertainty in the satellite-based SO_2 remains still high, unless the data is averaged over a long time period/ large spatial area. At all three stations lower correlation between the satellite and in situ based SO_2 measurements were obtained than for the other source parameters, at Marikana there is practically no correlation. Similar results were obtained when the satellite- and in situ-based proxies were compared (Table 2, figures in the supplementary material). Overall large differences exist between the satellite proxies and in situ proxies.

Since at Marikana and Elandsfontein the in situ data showed correlation between the (NO_x-NO) and the SO_2 concentrations, the satellite NO_2 column density is also compared with the in situ SO_2. Results show that in fact the OMI NO_2 compares better with the in situ SO_2 than the actual OMI SO_2 product. At Elandsfontein R^2=0.25, and at Marikana R^2=0.31 are obtained between the satellite NO_2 column and in situ SO_2 concentration.

4.3.3 Comparison of satellite-based proxies with N_{nuc}

To further evaluate the performance of the satellite-based proxies, they are compared with the in situ N_{nuc}. Only data from Elandsfontein and Marikana are included in the comparinson since the number of coincident N_{nuc} and satellite proxy observations was too low at the other stations. As expected, neither of the two satellite-based SO_2 – proxies are able to predict N_{nuc}. Interestingly, the only case where weak correlation is obtained between a proxy using satellite data and N_{nuc} is for the NO_2/AOD (Fig 12). This result is very different than what is expected based on the comparison of the in situ proxies and N_{nuc}. In fact, the connection between NO_2/AOD and N_{nuc} is most probably related to the correlation between the satellite NO_2 column density and the in situ SO_2 concentration. If the source term in the SO_2·UV-B/AOD^2 proxy was replaced by NO_2·UV-B, the correlation with N_{nuc} at Elandsfontein would be R^2=0.23, and at Marikana R^2=0.06. This implies that over areas where SO_2 and NO_2 are affected by some common factors, e.g. emission sources, the satellite NO_2 could be a better estimate for the source term than SO_2.

5 Conclusions

~~Four different satellite based~~ This work explores the use of proxies ~~for estimating the number concentrations~~using satellite data to obtain information on the concentration of nucleation mode aerosol particles ~~and the potential for new particle formation were derived over South Africa. A distinct improvement in the quality of the proxy components was obtained when different satellite products were selected to those utilized by~~ (N_{nuc}). These proxies have been formulated using relations derived from data on ground-based nucleation and precursor gases, which were simplified for the use of satellite data in Kulmala et al~~.~~,. (2011). ~~The recently released MODIS coll. 6 AOD product at 3x3 km^2 spatial resolution revealed small scale AOD variation that could not be observed when using coarser spatial grid. This was especially important when considering~~ simplifications and associated assumptions are critical examined. In this study ~~area,~~data were used over part of South Africa where ~~the AOD is generally low (0.05-0.1). The four year medians of the OMI data (NO_2, SO_2, and UV-B) were also calculated in a 3x3 km grid using the method of Fioletov et al. (2011), which also revealed "sub-pixel" features in the spatial pattern. Even though the OMI SO_2 product was changed to a planetary boundary layer data set and the spatial pattern generally improved, the satellite data still seemed to miss some of the well known smaller scale sources within the study area.~~ground-based observations are available from four experimental sites, for comparison with both the satellite-based parameters used in the proxy formulations and for

comparison of the proxies with ground-based measurements of the nucleation mode aerosol particle number concentrations. For the computation of the proxies, data from the A-train satellites are used. The NO_2, SO_2 and UV-B radiation are obtained from the OMI instrument and AOD from the MODIS instrument. The NO_2 and UV-B data are the same than what was used in Kulmala et al. (2011), but the AOD was upgraded to the newest collection 6 3 km product. Also the SO_2 product was changed to the planetary boundary layer product (OMI SO_2 PBL) that represents the total column values with a priori assumption that the emissions are mainly in the boundary layer. The satellite observations are also extensively compared with in situ data.

Based on the proxies derived from the in situ data it is expected that the SO_2-related proxies would be the best predictors of N_{nuc} within the study area at the time of the satellite overpass (13-14 LT). It is also noted that even though the in situ NO_2/CS proxy did not do well in predicting N_{nuc}, a positive correlation between the SO_2 and NO_2 concentrations is found at the measurement stations (at 13-14 LT). The R^2 between in situ SO_2/CS and N_{nuc} is 0.22 and this value could be considered as some kind of "upper limit" for the satellite proxies, for which uncertainties are much higher than for the in situ proxies. Using ground-based data, Kulmala et al. (2011) reported that SO_2 had only moderate influence on the performance of the $SO_2 \cdot UV/CS^2$ proxy in Southern Finland. The overall correlation between this proxy and N_{nuc} over South Africa was even lower (R2= 0.13) than over Southern Finland (R^2=0.29), yet our results clearly indicate a strong influence of SO_2 on the performance of the proxy. If the SO_2 was excluded from the proxy, no correlation with in situ proxies and N_{nuc} was found.

Kulmala et al., (2011) emphasized that the most crucial assumption in deriving the satellite based proxies was the replacement of the CS with AOD. ~~Overall the AOD was near the satellite detection limits in many locations, which might result in "artificially" high values of the satellite based proxies. Despite the theoretical differences between CS and AOD, good correlation was obtained between in situ scattering and CS, but significantly lower correlations were obtained when the in situ scattering or CS were compared with aerosol extinction integrated over the total atmospheric column, i.e. AOD. One of the most probable reasons was the elevated aerosol layers, which were often related to the increased AOD values. When the other satellite based proxy components were compared with the in situ data, the SO_2 column densities did not show that good agreement compared to the correlation with the NO_2 column density and the radiation. This was also observed when comparing the in situ and satellite based proxies - the highest correlations were obtained for NO_2/AOD and UV-~~

~~B/AOD^2. When the satellite based proxies were compared with actual nucleation mode number concentrations the correlations were low ($0.06 \leq R \leq 0.31$). Some improvement, however, was obtained ($0.21 \leq R \leq 0.34$) when the AOD was replaced by the estimated sink from the York fit (Fig. 5).~~ This assumption is further evaluated in the current study using several tests. A fundamental reason for differences between CS and AOD is the intrinsic dependence on different aerosol size ranges, with CS more sensitive to very small particles (smaller than about 200 nm) and AOD more sensitive to particles larger than that. Yet, good correlation is obtained between measured scattering coefficients for dry aerosol and CS evaluated from collocated particle size distribution measurements. When the in situ scattering coefficients or CS are compared with collocated AOD measurements the correlation decreases. This may be due to several effects. In particular the presence of elevated aerosol layers and/or large dust particle increases the AOD but does not affect the CS. However, overall the AOD is rather low (<0.1) over the major part of the study area, which means that these values are also associated with substantial relative uncertainty, which needs to be accounted for when deriving the satellite-based proxies.

~~South Africa is a challenging area to study the satellite proxies within, since the nucleation event frequency is very high. Generally the CS is very low and not very high source species is needed for an event to occur. During the study period (Jan 2007-Dec 2010) only few days were classified as non-event at the measurement stations. In addition all the non-event days were cloudy and the satellite proxies could not be determined. Despite of this, for each satellite based proxy a difference was seen between the entire measurement period median (namely event and unclear days) and event day median. At each measurement station where the event classification data was available the event day satellite proxy medians were larger than the entire measurement period median. From a single day satellite proxy values it was not possible to define whether there was an event or not.~~

~~In general this study showed that the satellite based 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. More studies of the satellite based proxies in different type of locations and environments are needed to improve the proxies, and especially the sink term, further. 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.~~

Even though the OMI SO_2 PBL data product showed a distinct improvement in describing the spatial patterns of SO_2 as compared to the dataset used in Kulmala et al. (2011), the satellite-based SO_2 did not describe well the day-to-day variations at the measurement stations. In addition, the observed SO_2 column values were often close to the noise level associated with a single column retrieval reported by Krotkov et al. (2008). The only relation between a satellite-based proxy and N_{nuc} was obtained for NO_2/AOD (at Elandsfontein R^2=0.24, and at Marikana R^2=0.09). The result is different than what was expected based on the in situ proxies. The most probable explanation is the positive correlation between the ground-based NO_2 and SO_2 concentrations within the study area. It is found that in fact the satellite NO_2 column correlates better with in situ SO_2 concentration than the satellite SO_2 column, where no correlation was found.

Overall this study shows that the uncertainties related to the satellite products remain a major issue in this satellite-based proxy approach, especially over areas like South Africa, where the AOD and the SO_2, and NO_2 concentrations are generally relatively low. Throughout the whole study the relative uncertainties related to the satellite-based proxies were well above 50%. In For the NO_2/AOD proxy the largest relative uncertainties were often related to AOD. Otherwise SO_2 was clearly the most uncertain component in the proxies calculated using satellite data. Despite these uncertainties related to the satellite data, the in situ data did not do significantly better in predicting N_{nuc} within our study area. This indicates that overall improvements in the formulation of the proxies are needed.

Acknowledgements

This work is supported by Academy of Finland (1251427, 1139656, Finnish Centre of Excellence in Atmospheric Science 272041), European Research Council (ATMNUCLE~~)~~ ~~and~~), The European Integrated project on Aerosol Cloud Climate and Air Quality Interactions (EUCAARI~~). ~~), and the ESA projects Aerosol-cci (ESA-ESRIN project AO/1-6207/09/I-LG and ALANIS-Aerosols (Contract no. 4200023053/10/I-LG, STSE-ALANIS Atmosphere-Land Interactions Study Theme 3: Aerosols). Eskom and Sasol supplied logistical support for measurements at Elandsfontein, while the town council of Rustenburg supplied support to the measurement at Marikana. The OMI NO_2, SO_2 and UV data were obtained from the NASA Mirador service maintained by Goddard Earth Sciences Data and Information Services Center (GES DISC). The OMI surface UV data were obtained from the NASA Aura

Validation Data Center (AVDC). The MODIS Aqua data were provided by NASA LAADS Web, and the CALIPSO data was obtained from –NASA Atmospheric Science Data Center (ASDC).

References

Beukes, P., Vakkari, V., van Zyl, P. G., Venter, A., Josipovic, M., Jaars, K., Tiitta, P., ~~Laakso, H.,~~ Kulmala, M., Worsnop, D., Pienaar, J~~., Järvinen, E., Chellapermal, R., Ignatius, K., Maalisck, Z., Cesnulyte, V., Ripamonti, G., Laban T., Skrabalova, L., du Toit, M~~., Virkkula, A., and Laakso L.: Source region plume characterisation of the interior of South Africa, as measured at Welgegund, ~~Atmos. Chem. Phys., in preparation, 2014~~Clean Air Journal, Vol. 23, No. 1, 2013.

Bucsela, E. J., Krotkov N., Celarier, E., Lamsal, L., Swartz, W., Bhartia, P., Boersma, K., Veefkind, J.P., Gleason, J., and Pickering, K.: A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI. Atmos. ~~meas. tech. discuss., 6, 1361-1407~~Meas. Tech, 6, 2607-2626 , doi: 10.5194/amt-6- 2607-2013, 2013.

Burrows, J.P., U. Platt and P.Borrell (Editors) The Remote Sensing of Tropospheric Composition from Space, 536 pp., Springer-Verlag Berlin Heidelberg 2011. ISBN: 978-3-642-14790-6 doi: 10.1007/978-3-642-14791-3., pp. 359-313.

Crippa, P., Spracklen, D., and Pryor, S.C.: Satellite-derived estimates of ultrafine particle concentrations over Eastern North America. J. Geophys. Res., 118, 9968-9981, 2013.

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P. & Lehtinen, K. E. J.: Formation and growth of fresh atmospheric aerosols: eight years of aerosol size distribution data from SMEAR II, Hyytiälä, Finland. Boreal Env. Res. 10, 323–336, 2005.

Fioletov, V.E., McLinden, C.A., Krotkov, N., Moran, M.D., and Yang, K.; Estimation of SO_2 emissions using OMI retrievals. Geophys. Res. Lett., vol. 38, L21811, doi:10.1029/2011GL049402, 2011.

Giannakaki, E., Pfüller, A., Korhonen, K., Mielonen, T., Laakso, L., Vakkari, V., Baars, H., Engelmann, R., Beukes, J. P., Van Zyl, P. G., Josipovic, M., Tiitta, P., Chiloane, K., Piketh, S., Lihavainen, H., Lehtinen, K. E. J., and Komppula, M.: One year of Raman lidar observations of free tropospheric aerosol layers over South Africa. Atmos. Chem. Phys. Discuss., 15, 1343-1384, doi:10.5194/acpd-15-1343-2015, 2015.

Hirsikko, A., Vakkari, V., Tiitta, P., Manninen, H.E., Gagné, S., Laakso, H., Kulmala, M., Mirme, A., Mirme, S., Mabaso, D., Beukes, J.P., and Laakso, L.: Characterisation of sub-

micron particle number concentrations and formation events in the western Bushveld Igneous Complex, South Africa. Atmos. Chem. Phys, 12, 3951-3967, 2012.

Hirsikko, A., Vakkari, V., Tiitta, P., Hatakka, J., Kerminen, V.-M., Sundström, A.-M., Beukes , J.P., Manninen, H.E., Kulmala M., and Laakso, L.: Multiple daytime nucleation events in semi-clean savannah and industrial environments in South Africa: analysis based on observations. Atmos. Chem. Phys, 13, 5523-5532, 2013.

Holben, B., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J. , Nakajima, T. , Lavenu, F. , Jankowiak, I. and Smirnov, A.: AERONET - A Federated Instrument Network and Data Archive for Aerosol Characterization. Rem. Sens. Environ., vol. 66, 1-16, 1998.

IPCC, Intergovernmental Panel on Climate Change, Fifth Assessment Report: Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Kokhanovsky, A.A., and de Leeuw Gerrit (editors): Satellite Aerosol Remote Sensing over Land, Springer, Berlin, Germany, 2009.

Krotkov, N., Carn, S., Krueger A., Bhartia, P., and Yang, K.: Band residual difference algorithm for retrieval of SO_2 from the Aura Ozone Monitoring Instrument (OMI). IEET T. Geosci. Remote, 44, 12591266, 2006.

Krotkov, N., McClure, B., Dickerson, R., Carn, S., Li, C., Bhartia, P., Yang, K., Krueger, A., Li, Z., Levelt, P., Chen, H., Wang, P., and Lu, D.: Validation of SO_2 retrievals from the Ozone Monitoring Instrument over NE China. Journal of Geophys. Res., Vol. 113, D16S40, doi:10.1029/2007JD008818, 2008.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V.-M., Birmili, W., and McMurry, P.H.: Formation and growth rates of ultrafine atmospheric particles: a review of observations. J. Aerosol Sci., 35, 143-176, 2004.

Kulmala, M., Petäjä, T., Mönkkönen, P., Koponen, I.K., Dal Maso, M., Aalto, P.P., Lehtinen, K.E.J., and Kerminen, V.-M.: On the growth of nucleation mode particles: source rates of condensable vapor in polluted and clean environments. Atmos. Chem. Phys. 5, 409-416, 2005.

Kulmala, M., and Kerminen, V.-M.: On the formation and growth of atmospheric nanoparticles. Atm. Res., 90, 132-150, 2008.

Kulmala, M., Arola, A., Riuttanen, L., Sogacheva, L., de Leeuw, G., Kerminen, V.-M. and Lehtinen, K.E.J.: The first estimates of global nucleation mode aerosol concentrations based on satellite measurements. Atmos. Chem. Phys. 11, 10791-10801, 2011.

Laakso, L., Laakso, H., Aalto, P.P, Keronen, P., Nieminen, T., Pohja, T., Siivola, E., Kulmala, M., Kgabi, N., Molefe, M., Mabaso, D., Phalatse, D., Piennaar, K., and Kerminen, V.-M.: Basic characteristics of atmospheric particles, trace gases and meteorology in a relatively clean Southern African Savannah environment, Atmos. Chem. Phys., 8, 4823-4839, 2008.

Laakso, L., Vakkari, V., Virkkula, A., Laakso, H., Backman, J., Kulmala, M., Beukes, J.P., van Zyl, P.G., Tiitta, P., Josipovic, M., Pienaar, J.J., Chiloane, K., Gilardoni, S., Vignati, E., Wiedensohler, A., Tuch, T., Birmili, W., Piketh, S., Collett, K., Fourie, G.D., Komppula, M., Lihavainen, H., de Leeuw, G., and Kerminen, V.-M.: South African EUCAARI measurements: seasonal variation of trace gases and aerosol optical properties. Atmos. Chem. Phys., 12, 1847-1864, 2012.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W.H., Pickering, K. E., Bucsela, E. J. , Gleason, J. F., Martin, R.V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J.J., Knepp, T. N.: Evaluation of OMI operational standard NO_2 column retrievals using in situ and surface-based NO_2 observations. Atmos. Chem. Phys.,14, 11587–11609, doi:10.5194/acp-14-11587-2014, 2014.

Lee, K. H., Li, Z., Kim, Y. J., ~~and~~& Kokhanovsky, A.: Atmospheric ~~Aerosol Monitoring~~aerosol monitoring from ~~Satellite Observations: A History of Three Decades.~~satellite observations: a history of three decades. In Atmospheric and ~~Biological Environment Monitoring, DOI 10.1007/978-1-4020-9674-7_2~~biological environmental monitoring, Springer ~~Science + Business Media B.V., 13-38.~~Netherlands, 2009.

Lehtinen, K. E. J., Korhonen, H., Dal Maso, M., and Kulmala, M.: On the concept of condensation sink diameter. Boreal Env. Res., 8, 405-411, 2003.

Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., and Hsu N.C. : The Collection 6 MODIS aerosol products over land and ocean. Atmos. Meas. Tech., 6, 2989-3034, 2013.

Lourens, A. S. M., Beukes, J. P., Van Zyl, P. G., Fourie, G. D., Burger, J. W. , Pienaar, J. J., Read, C. E., and Jordaan, J. H.: Spatial and Temporal assessment of gaseous pollutants in the Highveld of South Africa, S. Afr. J. Sci., 107, ~~8 pp.,~~ No. 269, 8 pp., doi:10.4102/sajs.v107i1/2.269, 2011.

Lourens, A. S. M., Butler, T. M., Beukes, J. P., Van Zyl, P. G., Beirle, S., Wagner, T., Heue, K-P., Pienaar, J. J., Fourie, G. D., Lawrence, M.G.: Re-evaluating the NO_2 hotspot over the South African Highveld, S. Afr. J. Sci., 108(9/10), Art. #1146, 6 pages. DOI: 10.4102/sajs. v108i11/12.1146, 2012.

Maritz, P., Beukes, J. P., Van Zyl, P. G., Conradie, E.H., Liousse, C., Galy-Lacaux, C., Castéra, P., Ramandh, A., Mkhatshwa, G., Venter, A.D. and Pienaar, J.J.: Spatial and temporal assessment of organic and black carbon at IDAF sites in South Africa, submitted to Environ. Mon. Assess., 2014.

Mielonen, T., Levy, R.C., Aaltonen, V., Komppula, M., de Leeuw, G., Huttunen, J., Lihavainen, H., Kolmonen, P., Lehtinen, K.E.J., and Arola, A.: Evaluating the assumptions of surface reflectance and aerosol type selection within the MODIS aerosol retrieval over land: the problem of dust type selection. Atmos. Meas. Tech., 4, 201–214, 2011.

Mishchenko, M. I., L. D. Travis, and A. A. Lacis: Scattering, Absorption, and Emission of Light by Small Particles. Cambridge University Press, Cambridge, 2002.

Petäjä, T., Maudin III, R.L., Kosciuch, E., McGrath, J., Nieminen, T., Paasonen, P., Boy, M., Adamov, A., Kotiaho, T., and Kulmala, M.: Sulfuric acid and OH concentrations in a boreal forest site. Atmos. Chem. Phys., 9, 7435-7448, 2009.

Petäjä, T., Vakkari, V., Pohja, T., Nieminen, T., Laakso, H., Aalto, P.P., Keronen, P., Siivola, E., Kerminen, V.-M., Kulmala, M., and Laakso, L.: Transportable Aerosol Characterization Trailer with Trace Gas Chemistry: Design, Instruments and Verification. Aerosol Air Qual. Res., 13: 421–435, 2013.

Schnelle-Kreis, J., Küpper, U., Sklorz, M., Cyrys, J., Briedé, J. J., Peters, A., and Zimmerman, R.: Daily measurements of organic compounds in ambient particulate matter in Augsburg, Germany: new aspects on aerosol sources and aerosol related health effects. Biomarkers, 14 (SI), 39-44, 2009.

Seaton, A., MacNee, W., Donaldson, K., Godden, D.: Particulate air pollution and acute health effects. The Lancet, 345, 176-178, 1995.

Shi, Y., Zhang, J., Reid, J.S., Hyer, E.J., and Hsu, N.C.: Critical evaluation of the MODIS Deep Blue aerosol optical depth product for data assimilation over North Africa. Atmos. Meas. Tech., 6, 949-969, 2013.

Tanskanen, A., Krotkov, N.A., J.R. Herman, and Arola, A.: Surface Ultraviolet Irradiance From OMI. IEEE Trans. Geosci. Remote Sens., Vol. 44, -No. 5, 1267-1271, 2006.

Utell, M.J., and Frampton, M.W.: Acute Health Effects of Ambient Air Pollution: The Ultrafine Particle Hypothesis. J Aerosol Med Pulm Drug Deliv., Vol. 13, 4, 355-359, 2000.

Vakkari, V., Laakso, H., Kulmala, M., Laaksonen, A., Mabaso, D., Molefe M., Kgabi, N., and Laakso, L.: New particle formation events in semi-clean South African savannah. Atmos. Chem. Phys., 11, 3333-3346, 2011.

Vakkari, V., Beukes, J.P., Laakso, H., Mabaso, D., Pienaar, J.J., Kulmala, M., and Laakso, L.: Long-term observations of aerosol size distributions in semi-clean and polluted savannah in South Africa. Atmos. Chem. Phys., 13, 1751–1770, 2013.

Veefkind, J.P., Boersma, K.F., Wang, J., Kurosu, T.P., Krotkov, N., Chance, K., and Levelt, P.F.: Global satellite analysis of the relation between aerosols and short-lived trace gases. Atmos. Chem. Phys., 11, 1255-1267, 2011.

Venter, A.D., Vakkari, V., Beukes, J.P., van Zyl, P.G., Laakso, H., Mabaso, D., Tiitta, P., Josipovic, M., Kulmala, M., Pienaar, J.J., and Laakso. L.,: An air quality assessment in the industrialised western Bushveld Igneous Complex, South Afric S. Afr. J. Sci., 108(9/10), ~~2012.~~ No. 1059, doi:10.4102/sajs.v108i9/10.1059, 2012.

Virkkula, A., Backman, J., Aalto, P.P., Hulkkonen, M., Riuttanen, L., Nieminen, T., dal Maso, M., Sogacheva, L., de Leeuw, G., and Kulmala, M.,: Seasonal cycle, size dependencies, and source analyses of aerosol optical properties at the SMEAR II measurement station in Hyytiälä, Finland. Atmos. Chem. Phys., 11, 4445-4468, 2011.

Winker, D. M., Hunt, W. M., and McGill, M.J.: Initial performance assessment of CALIOP. Geophys. Res. Lett., Vol. 34, L19803, doi:10.1029/2007GL030135, 2007.

York, D., Evensen, N., Martinez, M., and Delgado, J.: Unified equations for the slope, intercept, and standard errors of the best straight line. Am. J. Phys., 72(3), 367–375, 2004.

1

2 Table 1. A summary of the measurements used in this study. Here are listed only
3 measurements between the study period 1.1.2007-31.12.2010.

Instrument	Measurement area/ Location	Measurement period	Measured parameters
Ozone Monitoring instrument OMI (satellite)	25.0-28.0S, 25.5-30.5E (whole study area)	Jan. 2007-Dec. 2010, obs. appr. once/day, only cloud-free obs.	NO_2 and SO_2 column densities, UV-B irradiance
Moderate Imaging Spectroradiometer MODIS (Aqua, satellite)	25.0-28.0S, 25.5-30.5E (whole study area)	Jan. 2007-Dec. 2010, obs. appr. once/day, only cloud-free obs.	Column integrated aerosol optical depth AOD at 550 nm wavelength
Cloud-Aerosol Lidar with Orthogonal Polarization CALIOP (satellite based lidar)	Selected locations within the study area	Selected days between Jan. 2007-Dec. 2010	Vertical profile of aerosol extinction at 532 nm waelength
Aerosol Robotic Network AERONET Sunphotometer (in situ)	Elandsfontein (26.25S, 29.42E)	Mar. – Dec. 2010, only cloudfree obs. during daylight.	Column integrated aerosol optical depth AOD at 500 nm wavelength.
Nephelometer (in situ)	Elandsfontein	Mar. – Dec. 2010	Aerosol scattering coefficient
Differential Mobility Particle Sizer DMPS (in situ)	Marikana (25.70S,27.48E) Botsalano (25.54S, 25.75E) Welgegund (26.57S, 26.94E)	Marikana: Feb 2008-May 2010 Botsalano: Jan. 2007-Feb. 2008 Welgegund: May-Dec. 2010	Particle size distribution, condensation sink, event classification
Scanning Mobility Particle Sizer SMPS (in situ)	Elandsfontein	Mar. – Dec. 2010	Particle size distribution, condensation sink
	All in situ stations	dates/station as above	NO_x, and NO, SO_2, global radiation, T, RH

Table 2. Correlations between in situ- and satellite-based proxies. The number of coincident observations is denoted with "N". Scatter plots for each of the case are provided as a supplementary material.

Station	$(NO_x\text{-}NO)/CS$ vs. NO_2/AOD	SO_2/CS vs. SO_2/AOD	$SO_2 \cdot UV\text{-}B/cs^2$ vs. $SO_2 \cdot UV\text{-}B/AOD^2$	$Glob./CS^2$ vs $UV\text{-}B/AOD^2$
Elandsfontein	~~R~~R^2= 0.~~33 (p=0.026)~~11, N=46	~~R=~~R^2= 0.20 ~~(p=0.218)~~, N=41	~~R= ~~R^2=0.~~19 (p=0.236)~~13, N=39	~~R=~~R^2= 0.~~31 (p=0.051)~~30, N=52
Marikana	~~R~~R^2= 0.~~62 (p<0.001)~~38, N=93	~~R=~~R^2= 0.~~07 (p=0.554)~~005, N=76	~~R= ~~R^2=0.~~04 (p=0.700)~~13, N=76	~~R~~R^2=0.~~41 (p<0.001)~~22, N=117
Botsalano	~~R= ~~R^2=0.~~06 (p=0.821)~~004, N=16	~~R=~~R^2= 0.~~35 (p=0.225)~~12, N=14	~~R= ~~R^2=0.~~29 (p=0.309)~~30, N=14	~~R~~R^2=0.~~62, (p=0.017)~~11, N=18

Table 3. Correlation between the satellite-based proxies and number concentration of nucleation mode particles (D_p<30 nm) at Marikana measurement station. The first column presents the correlations when AOD is used to estimate the sink due to pre-existing aerosols, and the second column represents the correlations when the sink is estimated using the York fit for coincident CS–AOD observations shown in Fig. 5.

Proxy	Sink = AOD	Sink = 0.172·AOD-0.008
NO_2/Sink	R=0.31 (p<0.001)	R=0.34 (p<0.001)
SO_2/ Sink	R=0.11 (p=0.313)	R=0.23 (p=0.046)
SO2·UV-B/ $Sink^{-2}$	R=0.09 (p=0.390)	R=0.21 (p=0.046)
UV-B/ $Sink^{-2}$	R=0.17 (p=0.067)	R=0.25 (p=0.006)

Table 4. A summary of the new particle formation event data at Marikana, Botsalano, and Welgegund stations. The table includes only the measurements that were available between Jan 2007-Dec. 2010. In the brackets it is shown the percentages of days when the satellite proxies could be determined at the station (depending on which proxy was considered). In the most of the cases satellite based proxies could not be determined because of missing AOD due to cloudiness.

Station (measurement period)	Number of measurement days	Number of event days	Number of undefined days	Number of non-event days
Marikana (8.2.2008-17.5.2010)	659	568 (14-21%)	89 (8-7%)	2 (0%)
Botsalano (1.1.2007-5.2.2008)	365	252 (4-5%)	95 (4-5%)	18 (0%)
Welgegund (20.5.-31.12.2010)	118	93 (6-3%)	24 (8-4%)	1 (0%)

Figures

Figure 1. The study area and locations of the in situ measurement stations; BOT = Botsalano, MAR = Marikana, WEL = Welgegund, and ELA = Elandsfontein.

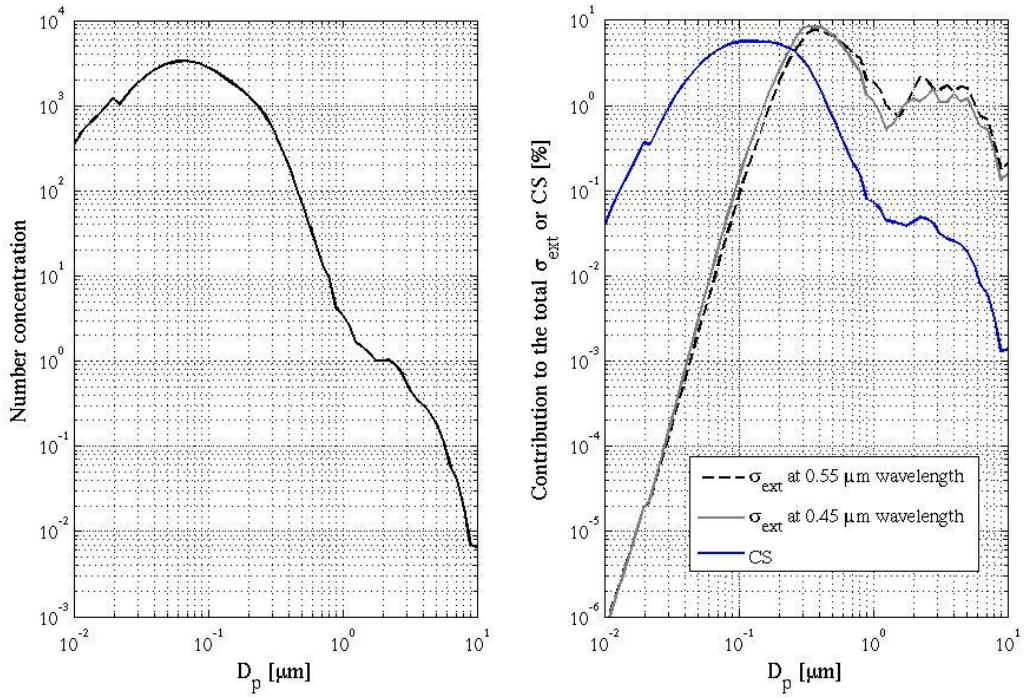


Figure 2. The sensitivity of CS and aerosol extinction coefficient to different particle sizes. In the left panel is shown the aerosol size distribution that is used to calculate CS and ~~rho–~~σ_{ext}~~.~~ is calculated for two wavelengths (0.55 and 0.45 µm) assuming spherical particles with a refractive index of m=1.48+0.001i. In the right panel is shown the contribution of each particle size to the total CS and ~~rho–~~σ_{ext}. The σ_{ext} is calculated for two wavelengths (0.55 and 0.45 µm) assuming spherical particles with a refractive index of m=1.48+0.001i.

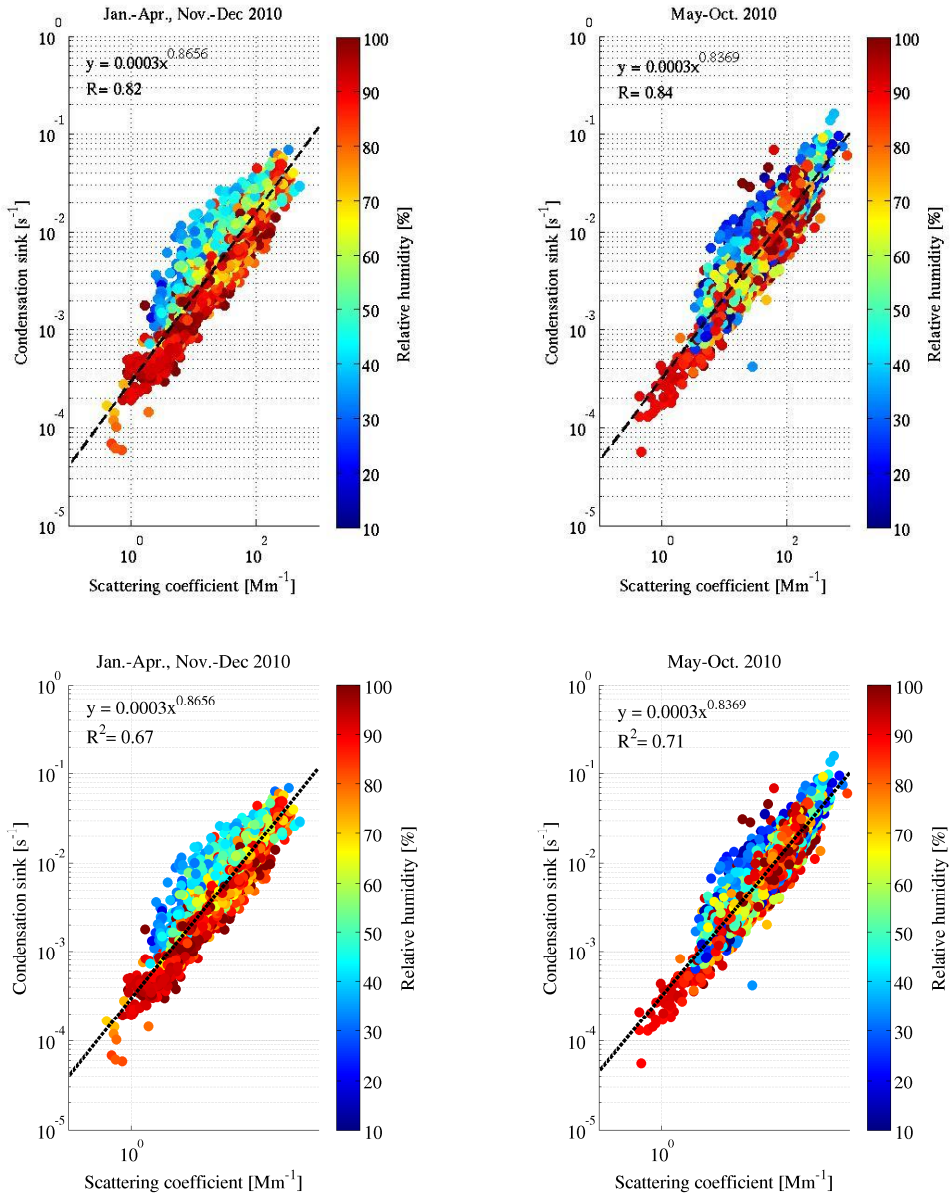


Figure 3. ~~The comparison~~Comparison between ~~in situ~~ condensation ~~sink~~sinks derived from particle size distributions, as described in the text, and nephelometer scattering ~~coefficient~~coefficients measured at Elandsfontein ~~measurement~~ station in 2010 for the warm (Jan-Apr., Nov.-Dec), and the cold ~~season~~ (May-Oct~~.~~).) seasons. CS has been corrected to the ambient relative humidity but the scattering coefficient ~~represents scattering~~was measured from dry particles. The data are colour-coded according to ambient relative humidity (RH) and the strong influence of RH on the relation between CS and scattering coefficient is evident.

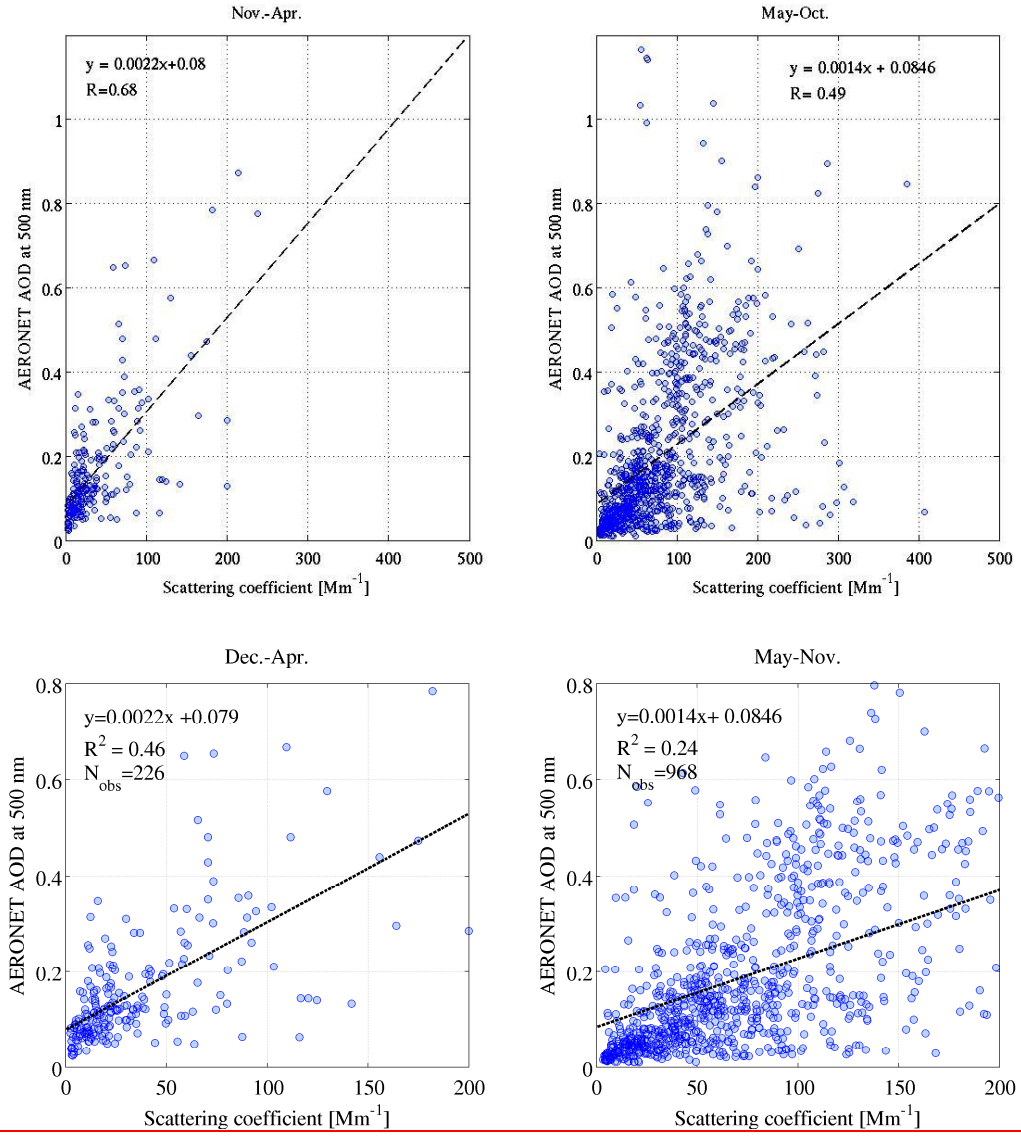


Figure 4. Comparison between ~~the~~ AOD at 500nm available from AERONET ~~AOD~~(see text) and in situ scattering ~~coefficient~~coefficients measured at the Elandsfontein ~~measurement~~ station. The AOD is the column integrated value of aerosol extinction (scattering + absorption) obtained from ~~the~~ sunphotometer measurements. The in situ scattering coefficient is ~~obtained from the~~measured with a nephelometer.

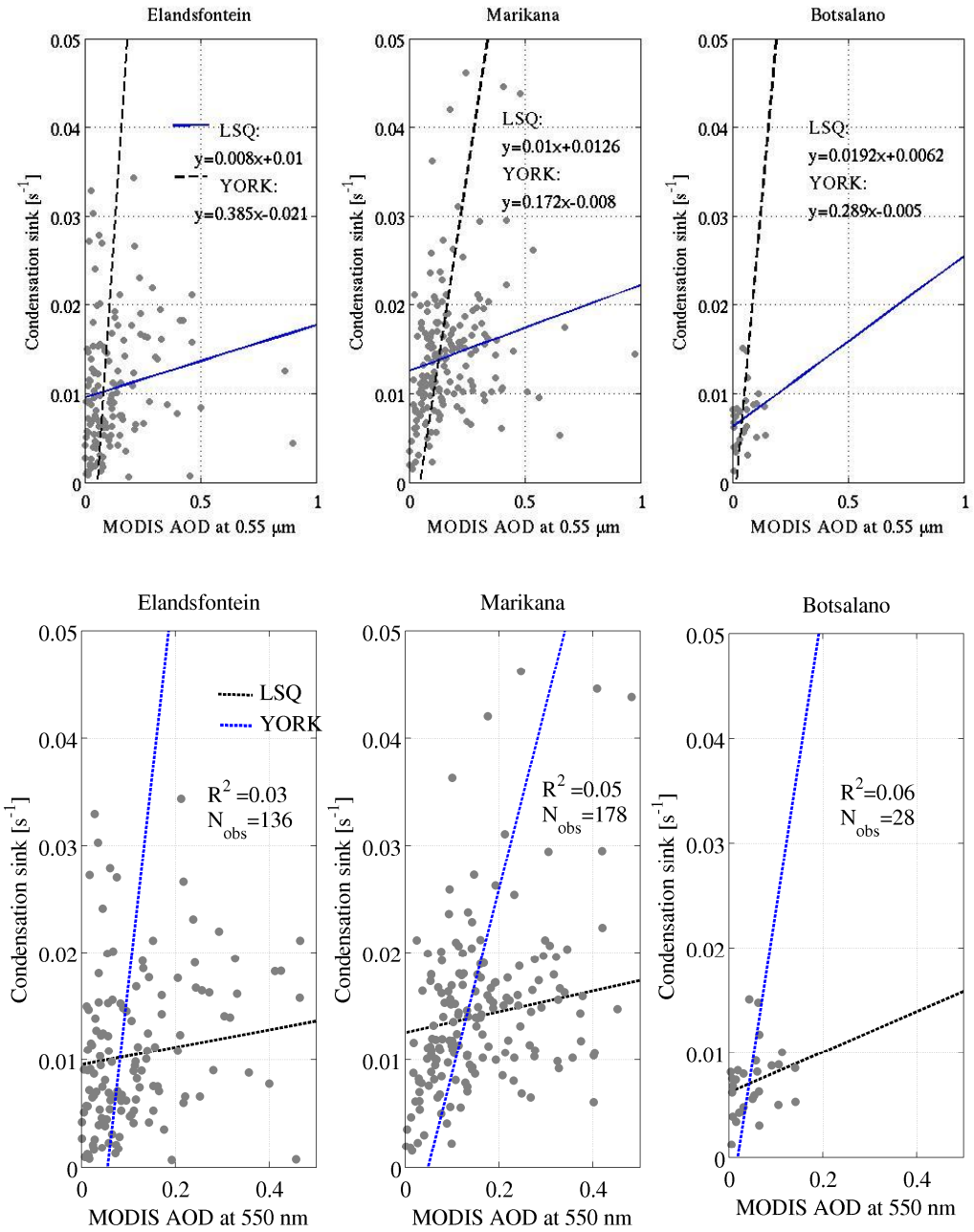
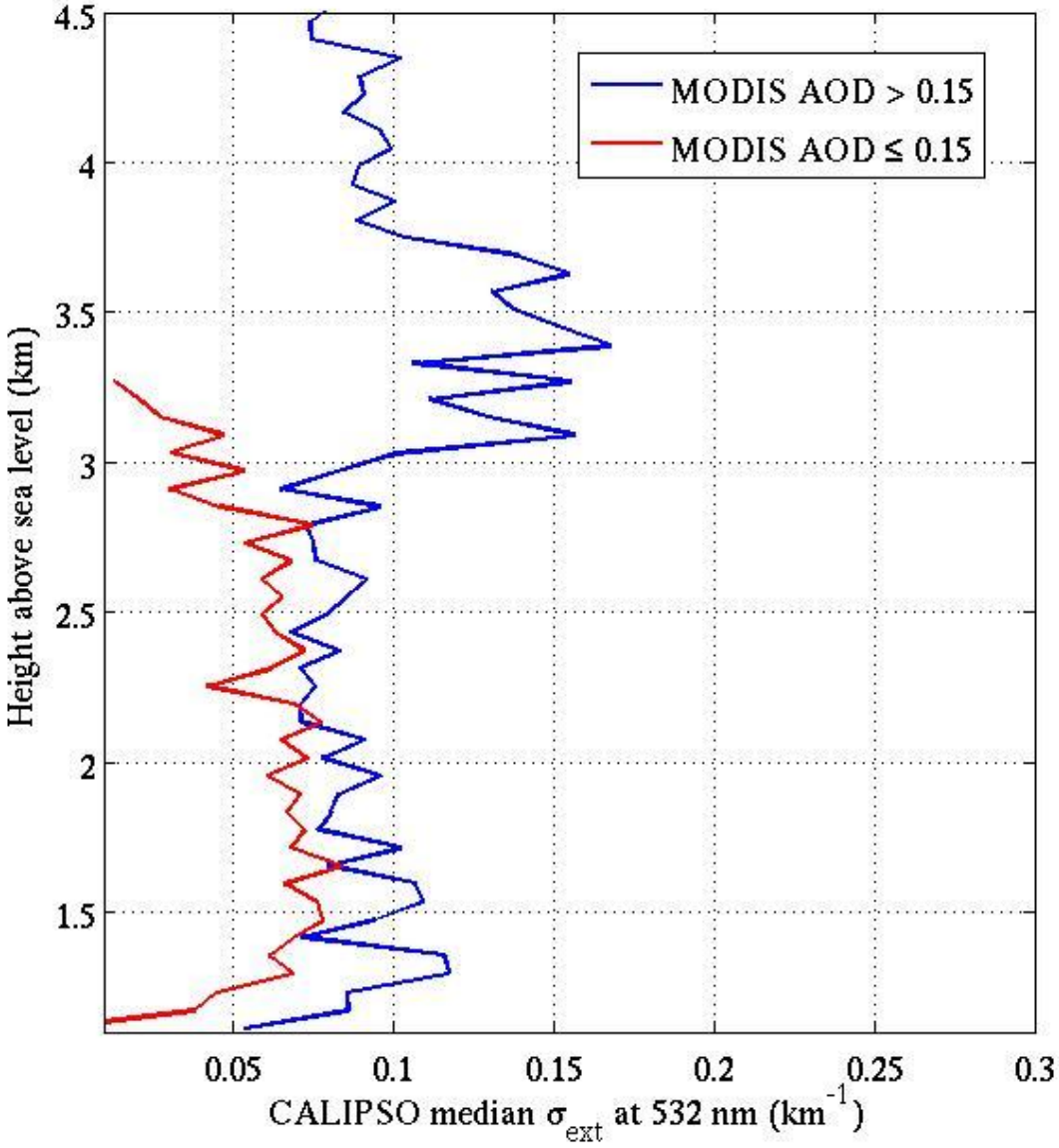


Figure 5. Comparison between MODIS AOD and in situ CS. The MODIS AOD values are spatial averages calculated from the observations within 3 km distance from the measurement station, whereas the CS values are one hour averages (13:00-14:00 LT). The ~~blue~~black lines represent the slope from least squares linear fitting (LSQ). The ~~correlation coefficients for Elandsfontein, Marikana, and Botsalano were R=0.17 (p=0.054), R=0.23, (p=0.002), and R=0.25 (p=0.201), respectively. The dashed black~~blue lines represent the fitting method where the uncertainties related to CS and AOD values have been taken into account (YORK,

York et al. 2004). The uncertainty for CS was set to 10 %, and for AOD to 0.05+15% (Levy et al., 2013).


MODIS AOD > 0.15
MODIS AOD ≤ 0.15
Height above sea level (km)
4.5
4
3.5
3
2.5
2
1.5
0.05
0.1
0.15
0.2
0.25
0.3
CALIPSO median σext at 532 nm (km-1)

1

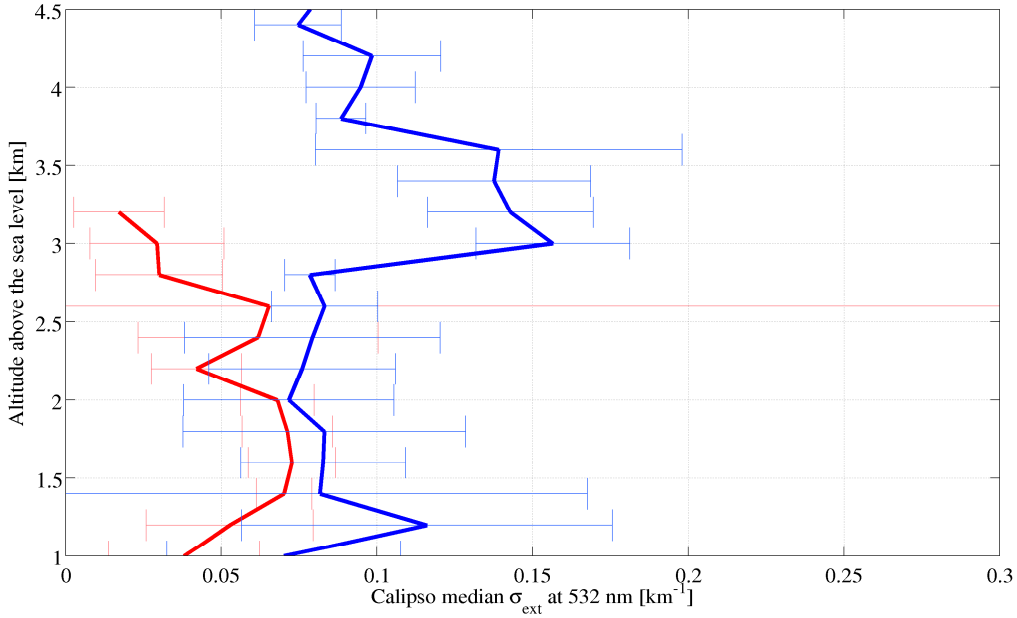
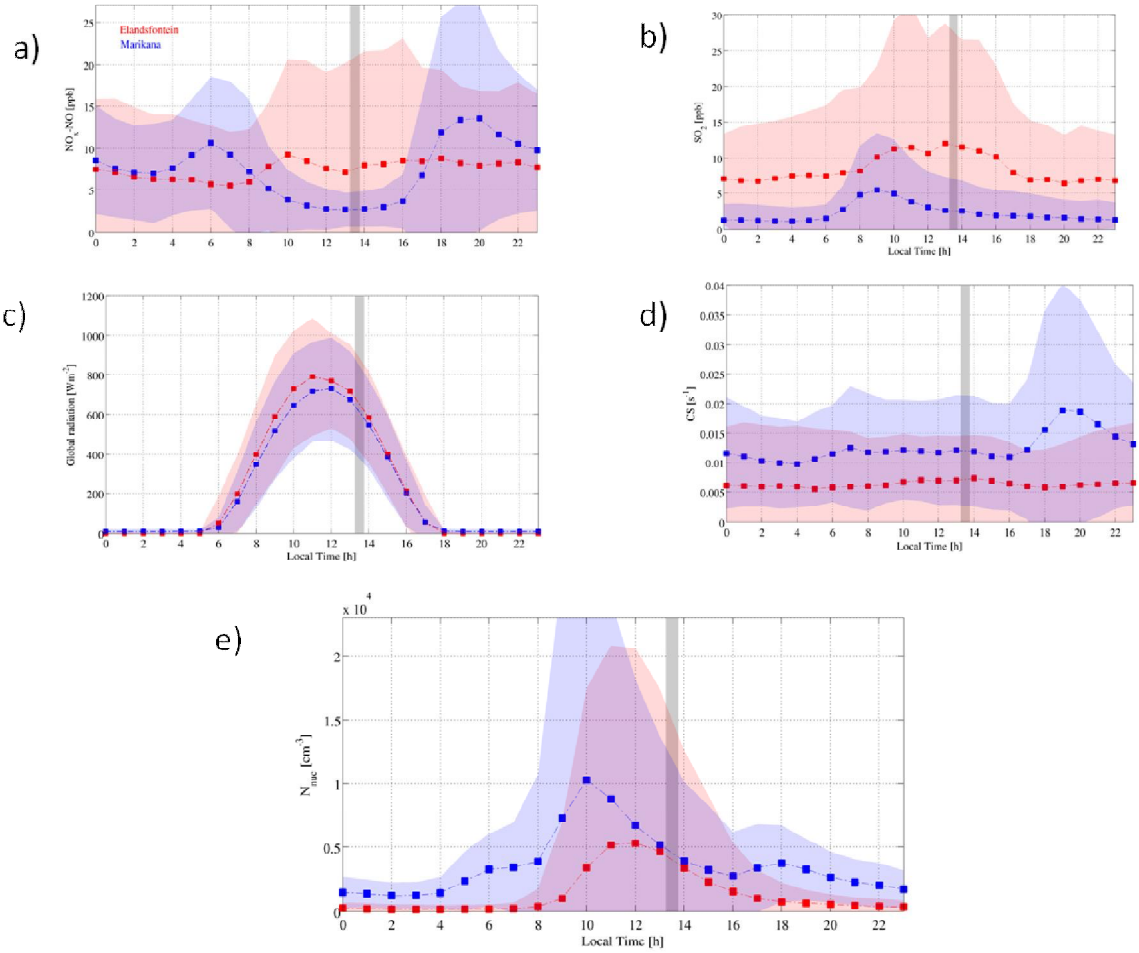
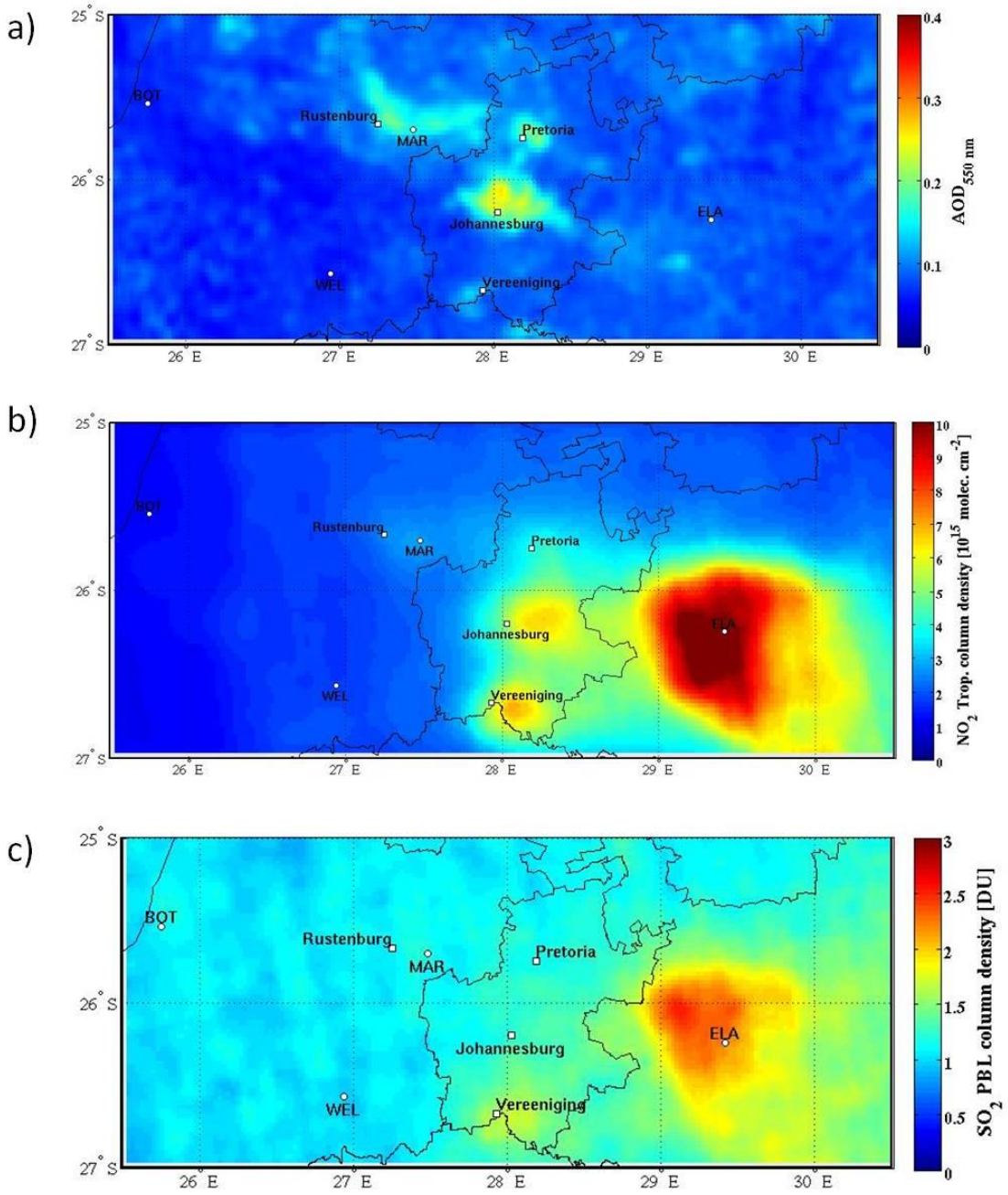


Figure 6. Median CALIPSO extinction profiles for days when MODIS AOD ~~was greater~~>0.15 (blue) and ~~lower~~ AOD<0.15 (red) ~~than the measurement period median (0.15).~~. The CALIPSO profiles are collected within 50 km radius from the Marikana station. The horizontal bars represent the interquartile ranges.


a)
Elandsfontein
Marikana
NO_x-NO [ppb]
b)
SO_2 [ppb]
Local Time [h]
c)
Global radiation [Wm^{-2}]
Local Time [h]
d)
CS [s^{-1}]
Local Time [h]
e)
x 10^4
N_{nuc} [cm^{-3}]
Local Time [h]

1



~~Figure 7. MODIS AOD (a) , OMI NO_2 (b) and SO_2 (c) atmospheric~~Figure 7. Diurnal variation of a) NO_x-NO, b) SO_2, c) global radiation, d) CS, and e) N_{nuc} at Elandsfontein (red) and Marikana (blue) stations. The grey columns represent the time window for the satellite overpass. The blue and red shading denote the 75th and 25th percentiles. It is noted that CS at Elandsfontein is defined with particles D_p<10μm, and at Marikana with particles D_p<1μm. N_{nuc} at Marikana represents particles D_p < 30 nm while at Elandsfontein N_{nuc} represents particles D_p 10-30nm.

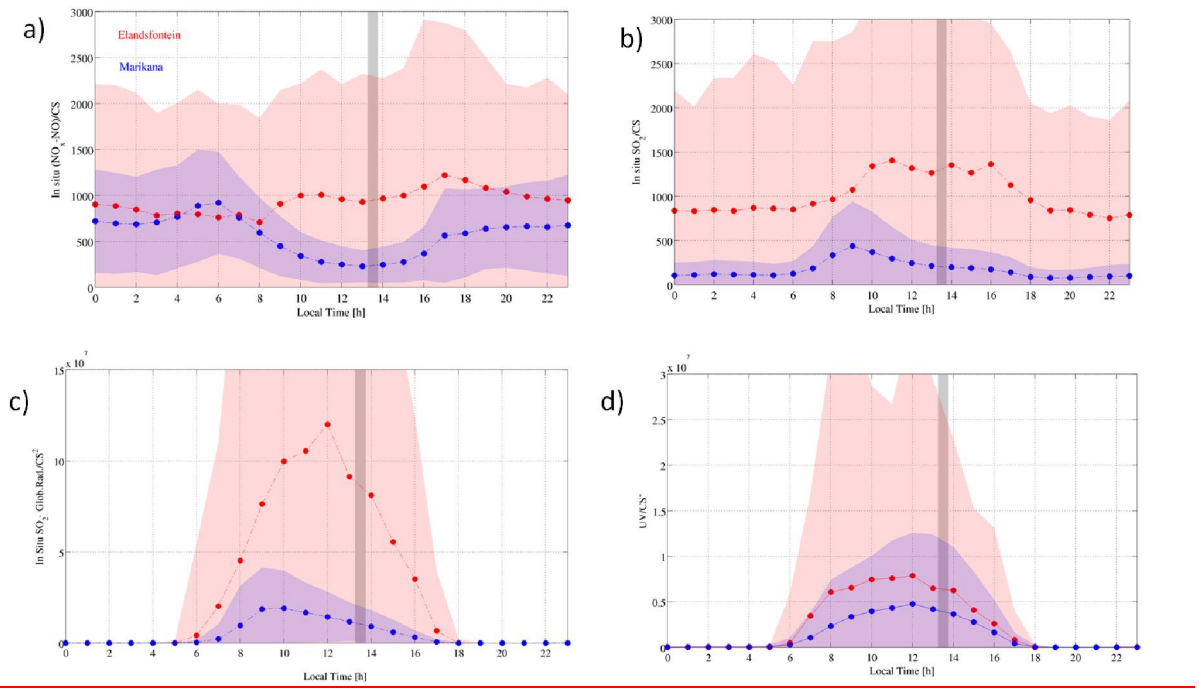


Fig 8. Diurnal variation of the proxies calculated using in situ data at Elandsfontein (red) and at Marikana (blue) stations. The red and blue shaded areas denote the 75th and 25th percentile ranges. The grey column represents the time of the satellite overpass.

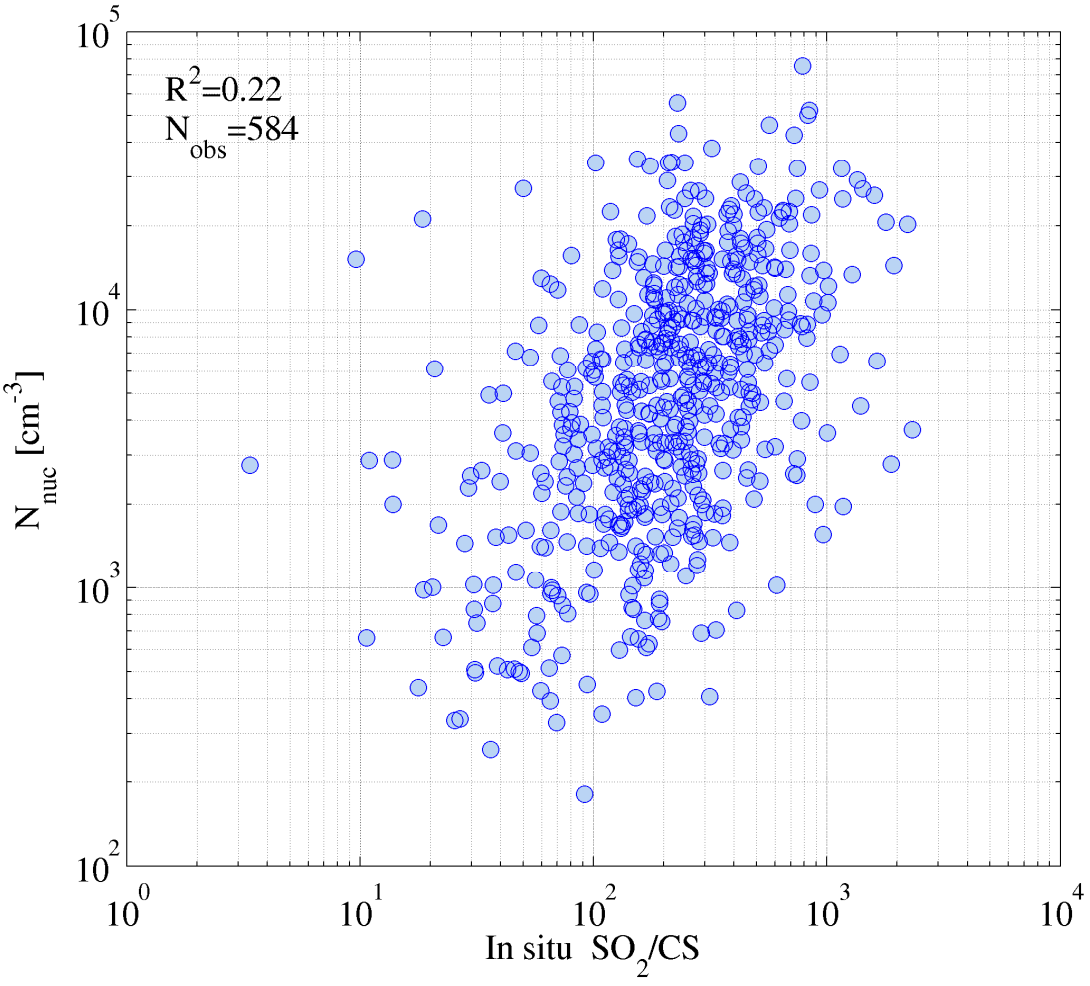


Fig. 9. Correlation between nucleation mode number concentration and SO_2/CS proxy calculated using in situ data at Marikana measurement station at the time of the satellite overpass (13-14 LT).

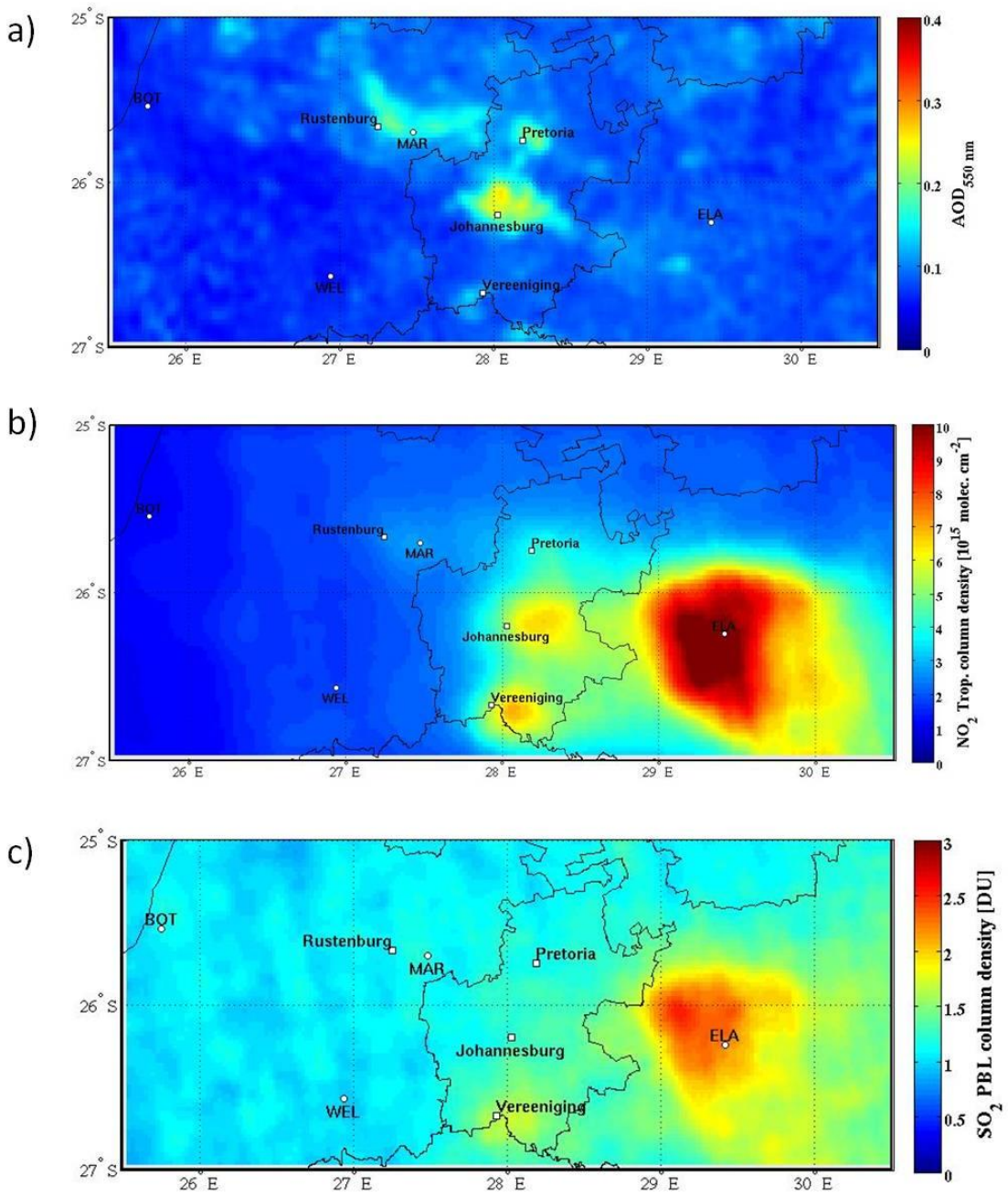
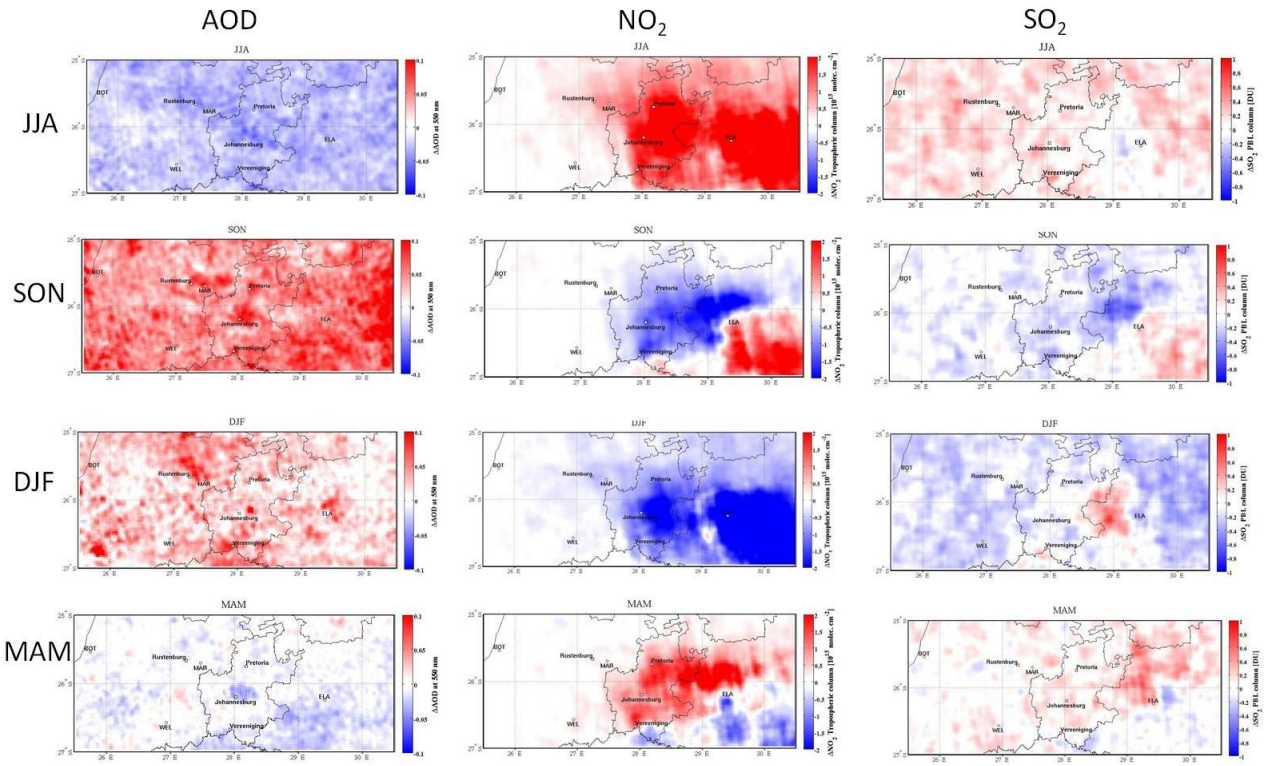


Figure 10. MODIS AOD (a) , OMI NO_2 (b) and SO_2 (c) column density medians for a four year period from Jan. 2007 to Dec. 2010. The locations of the in situ measurement stations (ELA= Elandsfontein, MAR=Marikana, BOT=Botsalano, and WEL=Welgegund) are marked with white dots.



~~Figure 8. Seasonal anomalies of MODIS AOD (left panel), OMI NO_2 (middle panel) and SO_2 (right panel) column densities. The anomalies are calculated using data between Jan 2007-Dec 2010.~~

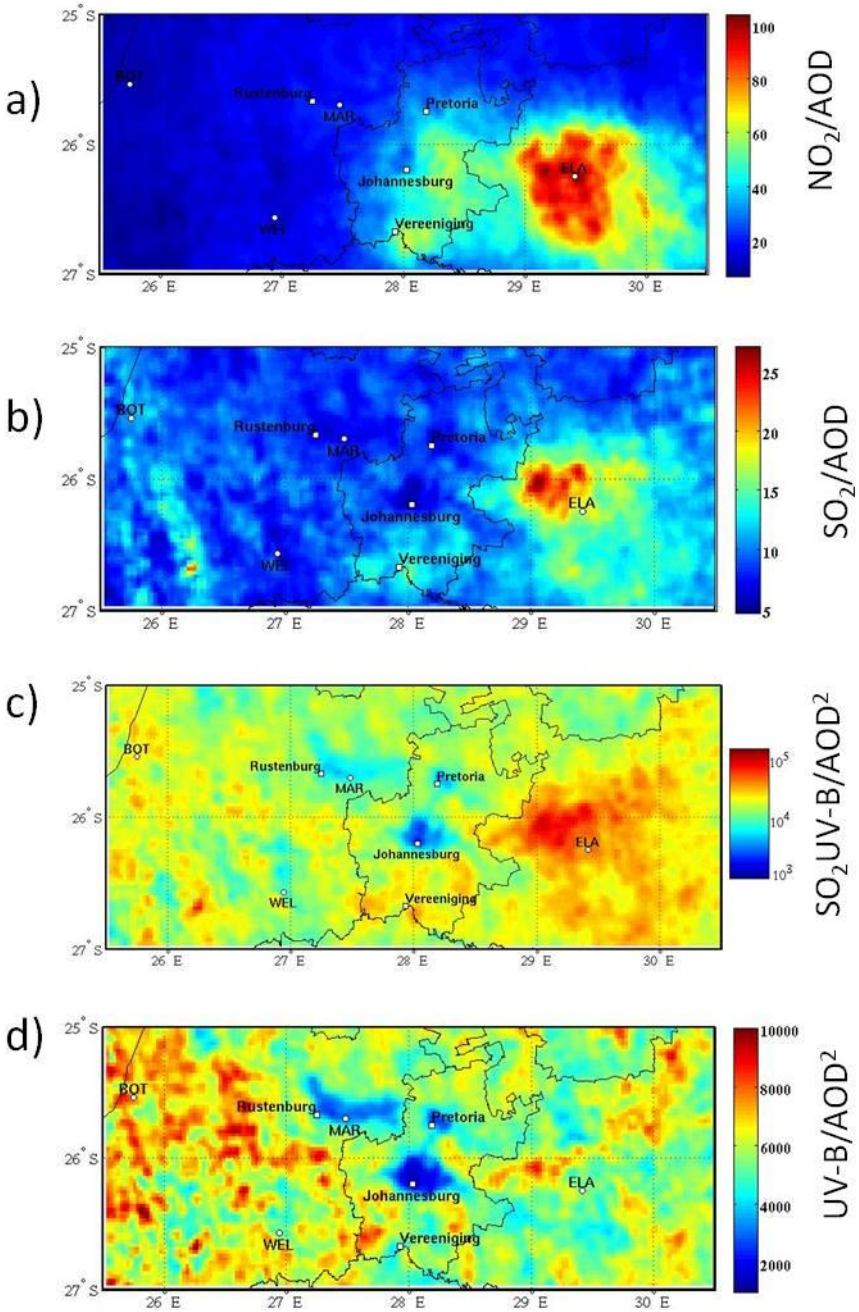


Fig. ~~9. Satellite based~~ 11. Spatial pattern of proxy medians for 2007-2010 ~~for~~calculated using satellite data. The proxies are a) NO_2/AOD, b) SO_2/AOD, c) SO2·UV-B/AOD², and d) UV-B/AOD².

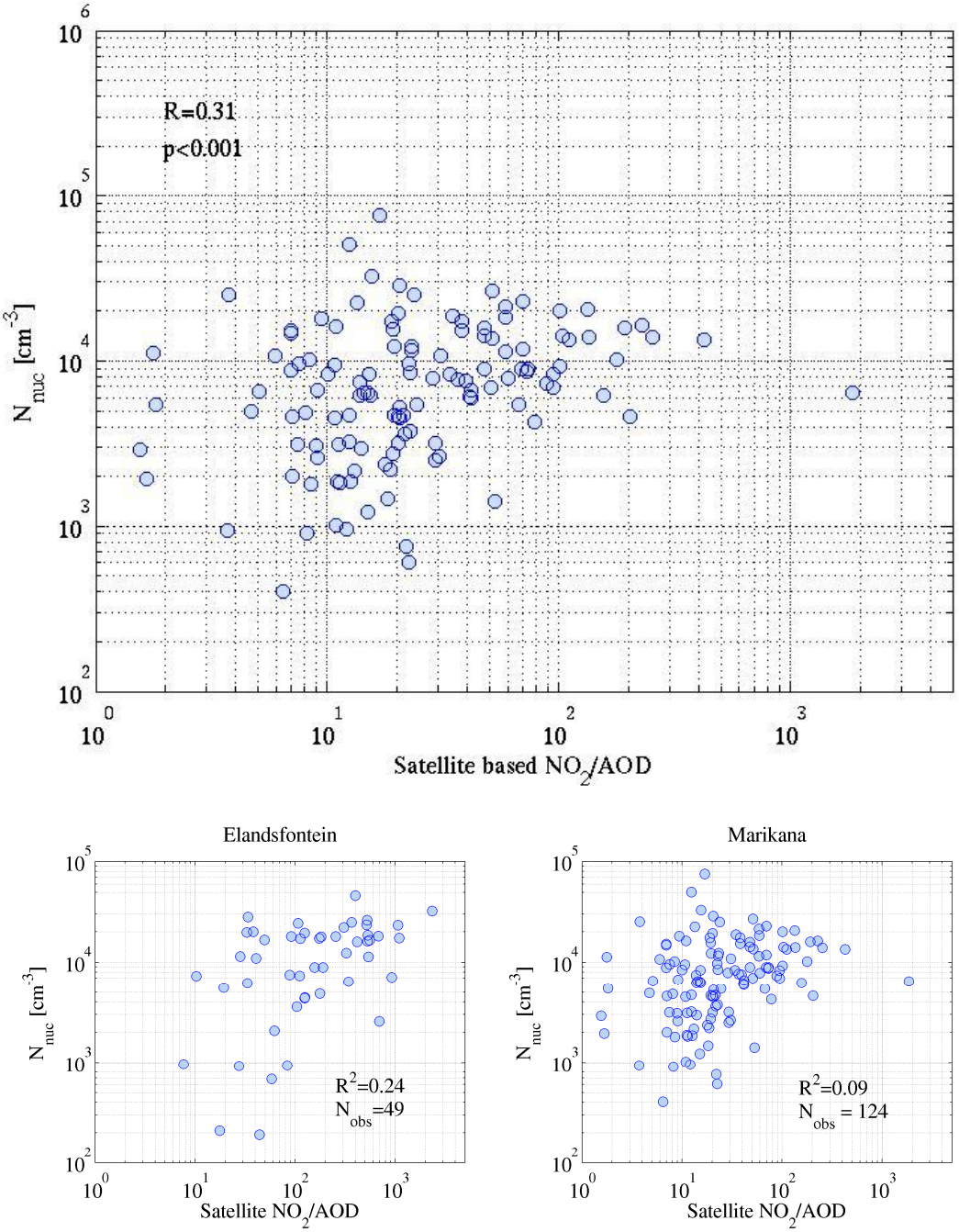
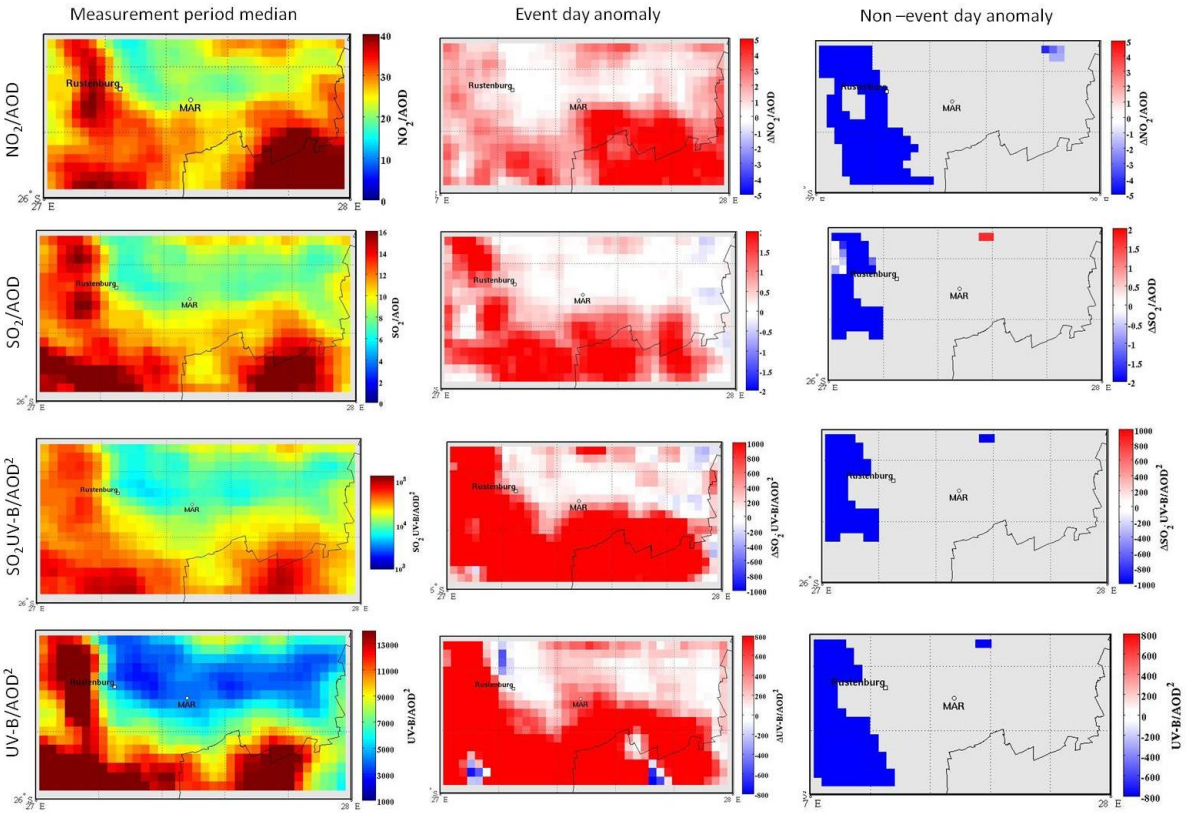


Figure ~~10~~12. The comparison between the number concentration of nucleation mode ~~(D<30 nm)~~ particles and ~~satellite based proxy NO_2/AOD.~~NO_2/AOD calculated from the satellite data at Marikana and at Elandsfontein stations. The number concentrations are one hour averages (13-14 ~~local time), whereas the satellite proxy value is about~~ LT) representative of the satellite overpass ~~at 13:30 local time.~~



1

2

3 ~~Figure 11. Satellite proxy medians, event-~~, time. It is noted that at Elandsfontein Nnuc

4 represents particles with D_p 10-30 nm, and ~~non-event day anomalies calculated from the event~~

5 ~~classification data obtained at Marikana measurement station. The proxies considered were (starting~~

6 ~~from top row) NO_2/AOD, SO_2/AOD, SO_2·UV-B/AOD^2, and UV-B/AOD^2. The grey areas designate~~

7 ~~missing values. For the non-event days the proxy data was missing mainly due to cloudiness.~~

8

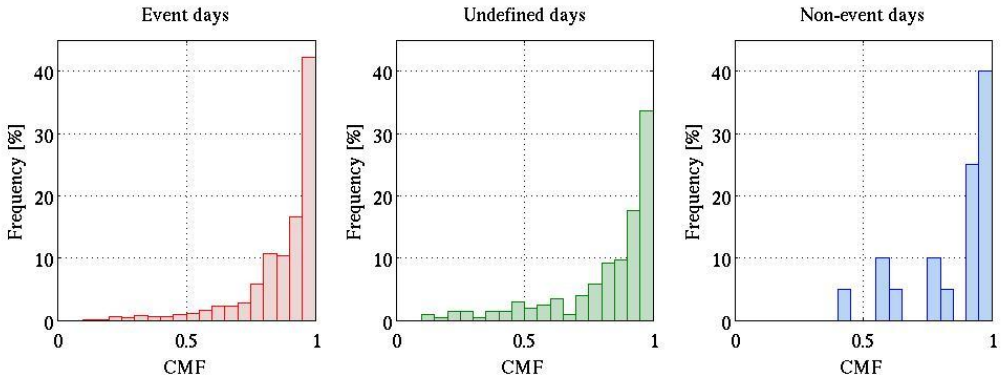


Figure 12. The OMI cloud modification factor (CMF) distribution for event, undefined and non-event days. The distributions are combined data fromat Marikana and Botsalano measurement stations. The particles with D_p<30 nm. N_{obs} denotes the number of event, undefined, and non-event day CMFcoincident observations were 768, 205, and 20 respectively. .