We thank Referee 1 for his/her comments. We have incorporated them into the revised manuscript. Please find our point-by-point answers below.

Overview:

The study provides a comparison of ground-based in-situ measurements (Zeppelin Station) of aerosol light extinction to observations from the satellite CALIOP sensor.

The manuscript details the difficulties in quantitatively comparing satellite and in-situ measurements, including discrepancies in space and time, uncertainty associated with aerosol humidification, and differences in actual measurement techniques. The authors use a complex approach in order to match appropriate CALIOP overpasses with in-situ data involving back trajectory analysis, CALIOP cloud-screening, and use a combination of humidified size distribution and chemical composition measurements to correct in-situ data to ambient humidity. The number of matching cases is extremely low (57 of a total of 2018 CALIOP overpasses), and a comparison of extinction coefficients yields agreement only within a factor of 10 (plus/minus a factor of 5). Additionally, the overpasses closest to the ground-site showed the worst correlation. Since the focus of the manuscript seems to be more about the process of linking the two measurements, rather than on the fairly uncertain results, it is suggested that more work is required to explore which steps are most important and if simplified methods could produce similar results.

The paper is indeed about the process of linking the different observations. We feel that this topic requires some attention as integrated observation systems and multi-platform synergies are considered as the next step in improving our understanding of atmospheric processes. Hence, reconciling such observations will pose a major challenge for the future.

From the reviewer comments we realized that we rushed into describing the detailed comparison approach in Section 3 without proper motivation for that course of action. Consequently, this and two other reviewers are interested in a quantification of the effects of the steps we apply during the comparison procedure. We actually considered the influence of the individual steps when we refined the comparison procedure from the simple closest approach method to what is described in Section 3. The simple comparison could not be used to reconcile the measurements at Zeppelin with CALIOP observations. While the number of comparison cases was much larger than the 57 cases we present in the manuscript, the difference in the extinction coefficients was in the range of three to four orders of magnitude. The increased complexity of the comparison approach decreased the difference in the compared values and led to physically meaningful situations for comparison. To make it easier for the reader to understand the rationale of our approach we added a paragraph to Section 3 that describes the background of why we believe that simplified comparison scenarios, i.e. missing any of the used steps, will lead to physically meaningless results in environments as considered in this study (with low aerosol load, high cloud cover, strong influence of relative humidity).

We started our investigation by applying the closest approach method to link CALIPSO observations in the region of interest to coincident dry in-situ measurements at Zeppelin station. While this course of action led to a high number of matches, it did not enable reasonable case-by-case reconciliation of in-situ and remote-sensing data. Differences in the compared aerosol optical properties ranged between two and three orders of magnitude. Perpetual refinement of the comparison procedure as described below showed that the failure in reconciling the different observations in the initial comparison is due to:

1. *Physically meaningless comparison scenarios in which no connection can be established between the locations of the ground site and the satellite track during heterogeneous aerosol conditions*
2. *The inclusion of apparently unrealistic signal spikes into the CALIOP extinction coefficient in case of fixed or inappropriately selected along-track averaging intervals*
3. *Humidification effects*
4. *The temporal delay in the observations*

The first two points make reasonable comparisons impossible. The latter two can still introduce uncertainties of up to 100%.

Major comments:

I fully appreciate the amount of work that went into this study and the detailed approach was thorough and well-presented, involving humidity correction, spatial scale matching with back-trajectories, and careful cloud-screening. The uncertainties of this process, coupled with the uncertainties associated with CALIOP measurements in the clean Arctic somewhat expectedly lead to non-ideal comparisons between the two measurements. Still, it is unclear how this study does anything more than point out these uncertainties in the form of Figure 4, including what visually looks like a lack of correlation at all.

We show that taking the uncertainties into account can lead to physically meaningful comparison cases. Highly averaged data on the other hand are likely to show agreement for the wrong reason. Consequently, the outcome of any study that attempts to reconcile CALIOP measurements with ground-based observations strongly depends on the comparison approach and data treatment. We see our study also as a critical assessment of the many issues involved with such endeavors.

Results indicating that increased overpass proximity to the ground site leads to decreased accuracy only suggest that the method was fundamentally unnecessary.

It is not clear to us what the reviewer suggests with this comment. We state that overpasses closest to the ground site (increased overpass proximity) generally present the worst comparison cases (decreased accuracy). This means that the presented method was in fact fundamentally necessary (the exact opposite of the reviewer's comment). The closest approach method leads to physically meaningless comparisons ("apples and oranges") for the conditions met in the Arctic. Our complex procedure on the other hand establishes a link between the different observations and reduces the effect of atmospheric variability as best as possible. While this reduces the number of comparison cases, it increases the overall quality of the comparisons. We think that the closest approach method's advantage of having a large amount of comparison cases does not outweigh its implicit drawback of including physically meaningless comparisons.

If accuracy to a factor of 10 is the best possible result, and if presenting this approach is the real result of the manuscript, then I believe a sensitivity study is necessary to assess each step in the process.

When we started our investigation, we intended to perform a systematic comparison of aerosol extinction coefficients as obtained from in-situ and lidar measurements. From this we would have evaluated the representativeness of measurements at Zeppelin for the Arctic and gained additional value to the CALIPSO observations and vice versa. Along the way we realized that such comparisons (even for a high overpass rate as in the Arctic) require significant efforts to ensure acceptable and usable data quality. As the Arctic is one of the hot spots of current research on aerosols and climate change, we believe that it is in the interest of the scientific community to see that reconciling aerosol properties from different platforms is not straightforward. This is of particular importance for data users that don't necessarily have a strong background in (the limitations of) the different measurement techniques.

A factor of ten was the worst agreement we found from our investigations. Most comparison cases were actually within a factor of two. This is orders of magnitude better than what could be obtained when doing the simple closest approach comparison. As suggested by the reviewers, we investigated the potential of using extinction coefficients from humidifying the dry nephelometer measurements with the help of reasonable scattering enhancement factors. This course of action leads to improved agreements as is stated in the revised manuscript.

For example, what does the comparison look like prior to each step in the analysis process? Examples of steps that could be simplified and evaluated for the effect on accuracy and uncertainty of the final in-situ/CALIOP comparison are (but shouldn't be limited to):

As stated in the beginning of our answer to the reviewer's comments, we started with a closest approach comparison which could not even give us the same order of magnitude of the different observations. Considerable refinement to was required to come up with comparison scenarios that are

likely to yield physically meaningful results rather than dealing with an apple-and-oranges situation.

1.) Can a constant humidification factor be used instead of necessitating continuous size distributions and chemical composition data?

For the revised manuscript, we calculated ambient extinction coefficients from the scattering and absorption coefficients measured with the dry nephelometer and PSAP, respectively. For that we used mean, minimum, and maximum scattering enhancement factors obtained by assuming γ-values of 0.57, 0.35, and 0.85, respectively, according to Zieger et al. (2010).

We restructured Section 2 to account for the added information. The procedure of humidifying dry nephelometer measurements is described in new Section 2.2.2 (Nephelometer + PSAP + scattering enhancement factor) as:

The DMPS measurements used in the previous section only cover particles up to a diameter of 790 nm and provide no information on the concentration of larger particles. These coarse-mode particles can have a huge effect on the overall aerosol optical properties as they are much more efficient scatterers of light compared to smaller ones. Hence, missing even low concentrations of coarse particles can cause an underestimation of the aerosol scattering and extinction coefficients by as much as 50% (Zieger et al., 2010, 2013). In addition, it is more straightforward to determine ambient extinction coefficients directly from the nephelometer measurements if the scattering enhancement factor is known or can be estimated within a reasonable range of values.

Therefore, ambient extinction coefficients were also calculated using the dry absorption and scattering coefficients measured with the PSAP and nephelometer, respectively, together with scattering enhancement factors that represent the median, minimum, and maximum effect of hygroscopic growth on light scattering. Values of γ = 0.57, 0.35, and 0.85, respectively, were used to obtain the scattering enhancement factor as $f(RH) = (1 - RH)^{-\gamma}$ (Zieger et al., 2010).

The statistical analysis of the ambient extinction coefficients derived from humidification of the nephelometer measurements were included to revised Figures 1 which is now discussed in new Section 2.2.3 (Dry versus ambient optical properties):

The box plots in Fig. 1 visualize the importance of transforming dry optical properties to ambient conditions. About 75% of the hourly aerosol scattering coefficients at 550 nm measured with the dry nephelometer at Zeppelin station in 2008 are smaller than 5 Mm^{-1}. Humidity correction to ambient extinction coefficients increases the median value for 2008 from 2 to 7 – 10 Mm^{-1}. The differences found in the median values of the ambient extinction coefficients derived according to the two methods described in Sects. 2.2.1 and 2.2.2 is likely to be the effect of coarse-mode particles that are not captured by the DMPS. These particles may contribute to about 20% – 30% of the total extinction coefficient at Zeppelin station (Zieger et al., 2010). The geometric mean has a much lower standard deviation than the arithmetic mean and is similar to the arithmetic median value. Independent of the retrieval method, the ambient extinction coefficient is on average a factor of three to five larger than the dry one when resolved according to different seasons. The Arctic haze period in spring shows the highest median values of the ambient extinction coefficient (17 – 22 Mm^{-1}) followed by winter (8 – 12 Mm^{-1}). Summer and fall are associated with very low median values (3 – 4 and 4 – 6 Mm^{-1}, respectively). Summer is the slightly cleaner season and a larger variation is observed during fall. This is in agreement with previous observations at Zeppelin station (Ström et al., 2003; Zieger et al., 2010; Tunved et al., 2013).

In the following, we use the ambient extinction coefficients derived from the humidified nephelometer measurements. This is because the lower and upper estimate in the γ-value for the determination of the scattering enhancement provides as with an error estimate that is more reliable than what can be obtained using the model approach described in Sect. 2.2.1.

We also revised Figures 3, 4, and 5 as well as Table 1 and their respective captions and discussions according to the new values of the ambient extinction coefficient obtained from the humidification of the nephelometer measurements.

2.) Can back trajectories be avoided by using the overpass point closest to the ground site?

We believe that trajectories should be considered to guarantee that comparisons are physically meaningful. Even for homogeneous aerosol conditions one should use trajectories to ensure that the

closest approach method is a valid simplification and applicable with minor impact on the comparison result. It is actually a message of this study that closest approach is not a suitable choice for the complex aerosol/cloud situation in the Arctic as it leads to physically meaningless solution, i.e. an error of 100%. This is addressed in the discussion of Figure 4.

3.) What are the results if a less rigorous could-screening process is applied?

We performed a signal screening rather than a classic cloud screening. Instead of excluding all CALIOP aerosol profiles that are flagged as cloud-containing, we regarded the aerosol extinction coefficient in our height range of interest of 250 to 750 m above sea level. CALIOP extinction coefficient is not automatically of bad quality if high clouds are present in the lidar profile. Considering the actual extinction coefficients allowed us to assess if these are in a realistic range of values. This investigation showed that CALIOP aerosol profiles – even when flagged as cloud-free – often show signal spikes that are clearly unrealistic, i.e. single height bins with values that are an order of magnitude larger than the adjacent bins. These spikes are an artifact of the low signal-to-noise ratio of the observations ("garbage in, garbage out"), and thus, should not be considered in the comparison. In our case of heterogeneous aerosol/cloud conditions with low signal-to-noise ratios, less rigorous signal screening will lead to physically meaningless comparisons, and hence, an error of 100%.

The beneficial result of a less-rigorous point-matching process is more comparison points and better statistics.

As stated earlier, we believe in quality over quantity. There is no benefit in 'better statistics' if these are skewed towards meaningless comparisons.

The step-by-step evaluation will also be useful for readers without such comprehensive in-situ measurements, and help to justify the benefits of the process.

We added the results of using ambient extinction coefficients derived from dry nephelometer measurements and reasonable assumptions in γ-values for determination of the scattering enhancement factor to the paper. This is a way to simplify the comparison from the side of the in-situ measurements.

Additionally, I would suggest presenting a few case studies that highlight good/bad correlations that may shed light on the underlying issues with the method.

This is already included in the manuscript. We present individual cases in Fig. 3 and discuss why only half of the overpasses in the chosen time window can be used for comparison.

Minor comments:

Page-line

5689-13. remove "among either"

"among other" has been omitted in the sentence

5695-4. Observations from the summer were not used for comparison because of difficulties by CALIOP. If scattering enhancement factors were only derived from July-October, were they used at all in the analysis? If not, it may make sense to remove them.

In the submitted manuscript, we used the measurements with the humidified nephelometer and scattering enhancement factors derived with this instrument to validate the performance of the humidification model that gives us ambient extinction coefficients. Scattering enhancement factors were not used directly in the analysis presented in the submitted manuscript.

As suggested by the reviewer, we added a new part to the revised manuscript in which we investigate if we can use dry nephelometer measurements and a constant scattering enhancement factors to obtain similar results. The scattering enhancement is parameterized with the help of the γ-value which depends on aerosol chemical composition. The used γ-values refer to the median, minimum, and maximum values derived by Zieger et al. (2010). This is described in new Sect. 2.2.2. As seasonal changes in the chemical composition of the aerosols at Zeppelin station are not dramatic (see Fig. 3 of Rastak et al., 2014), we are confident that using this range of γ-values will lead to reasonable ambient extinction coefficients.

5689-15. The four 'issues' you present are certainly pertinent to the study and provide a good review of the difficulties associated with remote/in-situ comparisons. I would suggest providing examples for each, e.g., specifically reference lidar and radiometer techniques under

2.

We thought of providing examples for each point presented in the introduction when we were working on the manuscript. We decided against it to keep these points more general. As these points are universally applicable, we wanted to prevent readers from feeling ignored if we wouldn't address their particular instrument of interest.

5695-9. What variability in the enhancement factor was observed? A factor 3 is very large compared to mid-latitude, continental sampling.

Zieger et al. (2013) present measurements of the scattering enhancement factor at different European sites. Highest f(RH)-values were indeed found in the Arctic at Ny-Ålesund with average values of 3.24 ± 0.63 at RH = 85%. For maritime air masses at Cabauw values could also reach 3 or higher at 85% RH. Even in the mid-latitudes, high values of 2.77 ± 0.37 were measured, e.g. at Melpitz, Germany, and explained by the high inorganic content of the aerosol.

Was the humidified nephelometer system verified with known substances like ammonium sulfate and nitrate?

The wet nephelometer has indeed been characterized and verified with known hygroscopic substances (ammonium sulphate and sodium chloride), as is described in detail by in Fierz-Schmidhauser et al. (2010). See also recommendations given in Zieger et al. (2013).

Fierz-Schmidhauser, R., Zieger, P., Wehrle, G., Jefferson, A., Ogren, J. A., Baltensperger, U., and Weingartner, E.: Measurement of relative humidity dependent light scattering of aerosols, Atmos. Meas. Tech., 3, 39-50, doi:10.5194/amt-3-39-2010, 2010.

Was there any evidence of biomass burning aerosols being transported to the site, which would likely reduce the enhancement factor significantly?

We checked the data of PSAP measurements at Zeppelin station for periods with an increased contribution of absorbing aerosols. We found a mean single-scattering albedo (SSA) of 0.985 ± 0.014 (median of 0.989). The cases we consider in our study did not show SSAs that are lower that these values. Hence, we conclude that the contribution of biomass-burning aerosols is negligible for observations at Zeppelin.

5695-20. How is bimodal (externally mixed) aerosol treated in this scheme? How often were distributions simple and mono-modal? Where changes in the chemical composition consistent with variability in the scattering enhancement factor?

Individual size distributions were monomodal during most of the time. The aerosol was treated as internally mixed with a sectional distribution over 20 size bins. The variability of the enhancement factor was found to be consistent with changes in chemical composition. Further details can be found in Rastal et al. (2014).

5696-1. The average contribution is minor, but were any biomass burning episodes observed which would result in anomalous comparison data points?

As stated in the answer to a previous comment, we are certain that biomass-burning aerosol did not affect the comparison cases conserved in this study. We added the following text to the new Sect. 2.2.1 to elaborate on the issue of absorbing particles:

Note that absorption contributes less than 75% to the ambient dry extinction coefficient of Arctic aerosols (Eleftheriadis et al., 2009; Zieger et al., 2010). This is in agreement with the PSAP measurements at Zeppelin. The effect of light absorption decreases even further when ambient extinction coefficients are considered.

5697-3. The effective radius calculation seems superfluous and seemingly was not used in the analysis. I would suggest removal.

Drying the particles could move them into a size range at which they are no longer efficient light scatterers. Consequently, not accounting for humidity effects will make it impossible to reconcile dry in situ measurements with ambient remote sensing observations. This is one of the major problems of studies like the one presented here. The effective radius is a parameter that is of enormous relevance for optical measurements and is commonly used in the remote-sensing community. Hence, this paragraph is not superfluous at all as we use it to illustrate the problem described above. Nevertheless, we now omit the equation to reduce the paragraph to:

The humidification of the particle number size distribution obtained with the DMPS leads to an increase of the particle effective (surface-weighted) radius from 0.14 ± 0.02 *to* $0.23 \pm 0.04\,\mu m$ *(yearly average, not shown). This moves the aerosol from an optically ineffective state to a size range in which they are more efficient in interacting with visible light. Contributions of particles larger than the maximum DMPS size bin would lead to an overall increase in the effective radius, and thus, further improve the light scattering efficiency of the probed aerosol.*

5701-1. It is hard to believe that distances this large are applicable in most environments. Can you comment on this result, based on your work?

In a previous study (Tesche et al., 2013), we used the trajectory approach (i.e. linking CALIPSO ground tracks to a fixed ground site with the help of forward and backward trajectories) for a validation of CALIOP profiles (particle backscatter coefficient at 532 and 1064 nm, particle extinction coefficient at 532 nm, and particle depolarization ratio at 532 nm) with ground-based lidar measurements at Cape Verde. We could show that successful comparisons of profiles of aerosol optical properties can be performed for distances of as much as 500 km. We reduced the maximum distance for comparison to 300 km in the present study to account for the more heterogeneous meteorology in the Arctic compared to the tropical north Atlantic. It is the advantage of the trajectory approach that overpasses at larger distance compared to the closest approach method can be included in a comparison study. Similar results have been found by studies of Anderson et al. (2003) and Kovacs (2006), which we refer to in the beginning of Section 3 (Comparison Approach) together with Tesche et al. (2013).

5705-13. The dependence on wind direction is weak and only really depends on a few datapoints at high extinction. I would suggest an analysis more quantitative than point-coloring for this figure. Wind-rose plot?

The color coding in this figure refers to the lidar ratio used in the retrieval of the CALIOP extinction coefficient. It shows that the largest absolute differences in extinction coefficients occur for cases with westerly flow and unusual lidar ratios (aerosol type). As the west of Svalbard is ice-free, it is possible for the CALIOP aerosol classification scheme to select from a larger pool of lidar ratios. A respective statement has been added to the discussion of Figure 5:

On the other hand, the CALIOP aerosol classification scheme can choose from a larger pool of lidar ratios for observations over ocean and land compared to those over snow and ice (Omar et al., 2009)

Figure 2. panel a, the colors for the labels ('no features' etc.) are difficult to distinguish.

Sorry for that. We increased the size of the labels to improve readability.

Figure 2. panel b, is 'cloudfree' and 'aerosol only' the same data? If so, please use consistent labels. Likewise for 'cloudy' and 'clouds and aerosols'.

We harmonized the labeling and the colors used in Figure 2: cloud-free was changed to aerosol only and cloudy was changed to clouds + aerosols. We also shortened the figure caption to:

Histograms of the monthly abundance of (a) CALIOP level 2 5-km aerosol profiles and (b) 60-m height-bins with aerosol observations as detected during 2018 CALIPSO overpasses in the region of interest during 2008. The color coding refers to the observed occurrence of atmospheric features (aerosols and/or clouds).

Figure 3. The triangles at the top are difficult to discern, please increase size. Since the colors are the same as other symbols in the figure, it is confusing to interpret. Consider using different shapes?

The colored triangles were replaced by different black symbols. The corresponding part of the caption was changed to:

Symbols and corresponding numbers mark CALIPSO overpasses that could be connected to the ground site for the considered time period: only aerosol features (triangles), aerosol and cloud features (diamond), and no or only cloud features (circles).

Figure 4. Is there any linear correlation between variables? Can a regression line w/confidence limits be added to provide some statistical basis of the correlation? An average CALIOP/in-situ factor of 1.85 and is noted in the text, can these be shown in Figure 4? Are

geometric means more appropriate for log/log plotting like this?

We could not find a linear correlation between the variables and refrain from adding a correlation line with a squared correlation coefficient of 0.16. Note that Fig. 4 now shows the results of humidifying the dry nephelometer measurements and that the discussion of this figure has been revised accordingly.

We thank Referee 2 for his/her comments. We have incorporated them into the revised manuscript. Please find our point-by-point answers below.

Overview:

The manuscript deals with a comparison of extinction coefficients at the Arctic which were derived in situ from nephelometer measurements with subsequent corrections at Zeppelin station and which were taken from remote sensing data (corrected backscatter data measured by CALIOP) in the wider vicinity of Zeppelin station. This is a very challenging task. The problem that the measurements were not taken at the same time and place was tackled by using trajectories in order to assure that at least the same air mass was probed. The authors applied a correction method regarding relative humidity which links the dry nephelometer measurements with the ambient CALIOP measurements.

It seems like the reviewer misunderstood the methodology we describe in the submitted manuscript to retrieve ambient extinction coefficients from the dry in-situ measurements. We did not humidify the dry nephelometer measurements to derive ambient extinction coefficients. The parameter was retrieved using the particle size distribution and Mie-scattering theory. Dry and wet nephelometer measurements were only used to validate the microphysical model that has been used to obtain scattering and extinction coefficients from the dry and humidified size distribution data.

However, we have now added the direct aerosol optical in-situ data to the comparison. We investigated if we can also use measurements of the dry nephelometer together with scattering enhancement factors to derive ambient extinction coefficients. This approach has the advantage of including the contribution of the coarse-mode fraction (that is not included in the size distribution measured with the DMPS) to the extinction coefficient. The scattering enhancement factor was obtained using median, minimum and maximum γ-values of 0.57, 0.35, and 0.85, respectively, according to Zieger et al. (2010). We restructured Section 2.2 accordingly to describe this procedure in Section 2.2.2. The results were added to Figures 1, 3, 4, and 5 as well as to Table 1. The description and discussion of these figures and the table has been revised accordingly. We now used the ambient extinction coefficients obtained from applying the scattering enhancement to dry nephelometer measurements for the comparison to CALIPSO findings presented in Figures 4 and 5.

Finding measurements which are suitable for comparisons required an extensive screening of the data. Besides this huge amount of work, the manuscript lacks an analysis of data quality, especially determination of measurement errors including error propagation through application of the correction method. Without tackling errors, it is difficult interpreting and understanding the results.

The main message of the manuscript is that one has to be very cautious when attempting to integrate spatially non-coinciding observations from different platforms and that the choices for data averaging have huge implications on the results. In particular, the highly averaged data that are often used for such a task can lead to coherence of the observations that is no longer present when comparisons are performed on a case-by-case basis. We refined the comparison methodology to take the influencing factors into account. Each of these factors has the potential to make any meaningful comparison impossible and to cause differences of orders of magnitude ("apples and oranges" comparisons).

The reviewer is correct that our initial approach of determining ambient extinction coefficients from DMPS measurements with the help of hygroscopicity and Mie modelling was lacking a proper representation of the error in the derived parameter. As suggested by other reviewers, we determined ambient extinction coefficients from the nephelometer measurements using scattering enhancement factors derived with γ-values measured at Zeppelin station by Zieger et al. (2010). This procedure is described in new Section 2.2.2 of the revised manuscript. We use a γ of 0.57 to describe the most likely conditions together with minimum and maximum values of 0.35 and 0.85, respectively. The latter are now used as an estimate of the error in the correction method in revised Figures 3 and 4.

Detailed comments

The line and page numbers are taken from the printed version not from the online display.

General

*** I have difficulties in understanding phrases such as "agreement of a factor of 1.85" (page 1), "agreement of a factor of ca. two" (page 4), "factor of five in agreement" (page 12). Does it mean that the data agree with each other or they disagree? I think, data can agree within their error bars and in case there is a factor of something it points towards disagreement.**

We changed the respective formulation to emphasize if we see agreement or rather disagreement.

*** The whole issue of errors, error propagation, error bars is almost ignored. I am also missing comments about calibration of the nephelometer (how often, how old was the latest calibration, the data were reprocessed in order to account for a shift in the calibration constant between the day of calibration), about the detection limit.**

We now provide error bars for the ambient extinction coefficients obtained from the nephelometer measurements as derived from using different estimates in the γ-value for retrieving the scattering enhancement factor (see answer to general comments). The dry nephelometer underwent the usual quality assurance steps (regular CO2 and zero air calibration). The standard nephelometer truncation and illumination correction (Anderson and Ogren, 1998) has been applied as well. The nephelometer was also indirectly validated by using the measured size distribution and Mie theory (see Rastak et al, 2014).

*** It is not clear how the authors tackled the problem of the particle absorption which is not measured by the nephelometer. It seems from the comments on page 6 (lines 193-195) that particle absorption is only taken into account through the refractive index of particles omitting the issue of the mixing state (external/internal mixture).**

In the retrieval of the ambient extinction coefficient from the DMPS measurements, light absorption is accounted for through the refractive index used in the Mie-scattering calculations. In the now added retrieval of the ambient extinction coefficient from nephelometer measurements, we accounted for light absorption with the help of PSAP measurements. The latter do not show increased light absorption for the cases considered in this study. The high single-scattering albedo of Arctic aerosols limits the effect of absorption to a negligible contribution to the ambient extinction coefficient (see answer to next comment).

The analyzed data contain biomass burning episodes? The approach of measuring dry aerosol particle scattering coefficient and calculating ambient aerosol extinction coefficients is valid for such episodes? Furthermore the time resolution for OC/EC ratio is one month (page 6). I doubt that monthly values can be reasonably used for single hourly measurements and correct for absorbing aerosols.

We checked the data of PSAP measurements at Zeppelin station for periods with an increased contribution of absorbing aerosols. We found a mean single-scattering albedo (SSA) of 0.985 ± 0.014 (median of 0.989). The cases we consider in our study did not show SSAs that are lower that these values. The behavior of the scattering enhancement factor with relative humidity dominates the error of the ambient extinction coefficient. Uncertainties in the treatment of humidity effects outweigh the influence of absorption to total light extinction in the Arctic. Hence, we conclude that the contribution of biomass-burning aerosols is negligible for observations at Zeppelin.

*** The authors investigated data for the whole year of 2008. The humidity correction was based on a model taking into account certain chemical components of aerosol particles. This model was evaluated using measurements done between July and October 2008 (see reference Rastak et al., 2014). Is this model for humidity correction suitable for measurements outside the evaluation period? How much did the chemical composition changed over the entire year? Seeing Fig. 3 in Rastak et al. 2014 manuscript, it seems that EC is much less during the evaluation period compared to the months December-May/June.**

We used the chemical data for the respective day (sulphate and sea salt which dominate hygroscopicity) or month (OC/EC which dominates absorption) together with individual size distribution measurements. From the chemical data we derive the kappa values required in the humidification model for the considered measurement days. The increase of the extinction coefficient due to hygroscopic growth is much larger than the contribution of absorption. Hence, even larger errors in the absorption coefficient (as a result of coarse resolution in the OC/EC data) are negligible in com-

parison to the uncertainties that can be introduced by improper description of hygroscopic growth. However, the chemical data that are required for the latter are available with daily resolution.

The model performed well for the evaluation period when using a kappa value based on the chemical information representative for this period. We don't see why it should not be suitable for another part of the year when the underlying particle chemistry (i.e. hygroscopicity) is adapted accordingly.

To assess the validity of our approach we now also present wet extinction coefficients as obtained by applying a constant scattering enhancement factor to the dry nephelometer measurements. The procedure is described in new Section 2.2.2. The results are presented in revised Figures 1, 3, 4, and 5 and in Table 1. The discussion has been revised according to the new findings.

Specific comments

Line 73: "above seal level" Typo; sea instead of seal

corrected

Line 191: "are water-soluble and insoluble organics, sulfate, sea salt, and black carbon." For avoiding misunderstandings, "sulfate" should be changed to "ammonium sulfate" (see Rastak et al., 2014)

changed

Lines 301-302: Which kind of meteorological data were used (GDAS, ReAanalysis. . .)?

We ran the HYSPLIT model with meteorological data from the Global Data Assimilation System (GDAS) archive. The last paragraph of Section 2.4 was extended to now state:

Meteorological parameters from the Global Data Assimilation System (GDAS) are provided along the trajectories and used in this study to estimate RH at the location of the CALIPSO overpass.

Line 388: "the CALIPSO observation is in poor agreement with the result of the in-situ measurement" I would not say it is poor agreement. It seems more like a disagreement.

We changed the wording from "poor agreement" to "disagreement".

Line 389-390: "This emphasized that using a closest approach for comparison of ground-based measurements and CALIPSO observations might not always be the best choice." I disagree with this conclusion. The closest distance approach is related to the idea of spatial homogeneity/inhomogeneity (or representativeness) of a quantity of interest, whereas the approach of this manuscript is related to the idea of probing the same airmass. In the end it is important whether it can be expected that the quantities could be compared or not.

The only thing that is important is whether a comparison of a certain quantity is physically meaningful or not. The closest approach method relies on temporal and spatial homogeneity of a quantity. This condition is often fulfilled if column-integrated properties are considered. This manuscript deals with vertically resolved observations of aerosol optical properties, and thus, an additional level of complexity. The refinement of probing the same air-mass enhances the chance of performing a physically meaningful comparison. It also improves the chances of obtaining quantities that can actually be compared, especially when individual overpasses are considered. None of this is implied in the closest approach method which is furthermore often applied to highly averaged data, i.e. mean values for a certain time period are compared to mean values over a certain area during that period.

Lines 417-419: "It was found that the most characteristic outliers in Figs. 4 and 5a occur for cases that were identified predominantly as polluted dust, polluted continental, and dust in the CALIPSO retrieval. These aerosol types are rather uncommon at 78N and suggest misclassification in the CALIPSO retrieval." Misclassification is a possible cause for explaining the outliers. However it is possible that classification was correct, pointing to an important contribution of aerosol absorption which seemed to be not well taken into account by the correction scheme for the nephelometer? Please note that even uncommon aerosol types could easily be present in single measurements.

The reviewer is correct to point out that single cases could easily represent exotic conditions rather than an erroneous measurement. However, it is more likely that these aerosol types are the result of misclassification – especially when the low signal-to-noise ratio of CALIOP observations in the Arctic during summer is considered. The CALIPSO aerosol model can chose from a wider pool of lidar ratios (aerosol types), if measurements are performed over land and open water (i.e. to the west

of Svalbard) rather than over snow and ice. This leads to a larger variation in the selected aerosol type for westerly flow. We added the following statement to the discussion of Figure 5:

On the other hand, the CALIOP aerosol classification scheme can choose from a larger pool of lidar ratios for observations over ocean and land compared to those over snow and ice (Omar et al., 2009).

Misclassification can occur as a result of signal noise, improper cloud screening, or due to surface effects. Given the structure of the CALIPSO aerosol classification scheme described in Omar et al. (2009), CALIOP observations in the Svalbard region during background conditions (weakly depolarizing and integrated attenuated backscatter coefficient not exceeding the threshold value of 0.0015 at 532 nm) should be classified as clean continental (over land and snow/ice) and clean marine (over ocean).

We thank Referee 3 for his/her comments. We have incorporated them into the revised manuscript. Please find our point-by-point answers below.

Overview:

This study presents a comparison of extinction coefficients as determined from spaceborne lidar measurements and from ground-based in-situ measurements at Zeppelin station during the year 2008. For this, the authors present here a complex procedure to match CALIPSO and ground-based observations based on HYSPLIT back trajectories to ensure the comparison of the same air mass. This procedure leads to only 57 overpasses during 2008 (from over 2000 overpasses in that year). The results obtained by the authors show how difficult is to obtain good results in such comparison.

I would recommend the authors to focus more on the screening and matching of the CALIOP data, analyzing further the associated uncertainties (averaging height range, intervals along the CALIPSO ground track, time, etc.). Although the number of cases analyzed is very low, it can be presented as the first attempt to compare extinction coefficients from spaceborne lidar and ground-based measurements using this approach. However, the authors need to analyze in depth the uncertainty of their approach and the results obtained.

We agree with the reviewer that there are many sources of uncertainties in our approach. We also realized from this and the other reviewer comments that we forgot to mention the history of how we came up with our comparison approach. It is in fact the outcome of a continuous refinement of the simple closest approach method which we found to fail in providing physically meaningful comparison cases. Several orders of magnitude of differences were found between the extinction coefficients from in-situ measurements and CALIOP observations when using the closest approach method. Discrepancies were reduced by trajectory matching, considering time delays, cloud screening, etc. This history of refinement is the reason why we consider a factor of about 2 as a very good comparison result. In the manuscript we name all the sources of uncertainty that we identified along the way. However, it is virtually impossible to quantify the individual errors as they all have the potential to make any meaningful comparison impossible. Accounting for the individual effects as best as possible will not ensure a flawless comparison. However, it will be closer to the truth than using rigid schemes like the closest approach. We added the following text to Section 3 to inform the readers about the background of our comparison approach:

We started our investigation by applying the closest approach method to link CALIPSO observations in the region of interest to coincident dry in-situ measurements at Zeppelin station. While this course of action led to a high number of matches, it did not enable reasonable case-by-case reconciliation of in-situ and remote-sensing data. Differences in the compared aerosol optical properties ranged between two and three orders of magnitude. Perpetual refinement of the comparison procedure as described below showed that the failure in reconciling the different observations in the initial comparison is due to:

1. *Physically meaningless comparison scenarios in which no connection can be established between the locations of the ground site and the satellite track during heterogeneous aerosol conditions*
2. *The inclusion of apparently unrealistic signal spikes into the CALIOP extinction coefficient in case of fixed or inappropriately selected along-track averaging intervals*
3. *Humidification effects*
4. *The temporal delay in the observations*

The first two points make reasonable comparisons impossible. The latter two can still introduce uncertainties of up to 100%.

General comments:

Page 5695, lines 4 - 8: This paragraph repeats the information on Page 5691, lines 28 - 29 and Page 5692, lines 1 - 4. The Zieger et al. (2013) reference is missing here though.

We removed this paragraph from the introduction and left it in the description of the instrumental setup at Zeppelin (Section 2.1).

Page 5696, lines 6 - 13: The hygroscopicity model was validated with data from the period July – October 2008. Can the authors explain further how this is valid for the whole year 2008? How would the annual variation of the aerosol concentration and properties affect this?

The model uses hourly size distribution measurements together with daily or monthly chemical composition data collected during the entire year of 2008 to account for the annual variation in the aerosol conditions at Zeppelin. The model performs satisfactory during the evaluation period. Hence, we assume that it will also do so during the rest of the year 2008 given that the required input parameters are adapted to measurements performed during this time.

Based on the suggestions made by the reviewers we investigated if wet scattering coefficients can also be obtained reliably by using the dry nephelometer and PSAP measurements together with scattering enhancement factors derived for a lower, median, and upper estimate of γ-values. We restructured Section 2.2 accordingly to describe this procedure in Section 2.2.2. The results were added to Figures 1, 3, 4, and 5. The description and discussion of these figures has been revised accordingly. We now used the ambient extinction coefficients obtained from applying the scattering enhancement to dry nephelometer measurements for the comparison to CALIPSO findings presented in Figures 4 and 5.

Page 5696, lines 14 - 15: "Values of $f(\mathbf{RH}) = 4.30 \pm 2.26$ with a range from 1.5 to 12.5 were found for the year 2008." To get these values, the hygroscopicity model by Rastak et al. (2014) was used with measurements of dry aerosol size distribution and aerosol composition. How frequent were these measurements? What is the uncertainty of this model? How would this affect the aerosol extinction coefficient for ambient conditions? And the comparison with CALIPSO?

The details on data availability are provided in Section 2.1:

The aerosol in-situ instruments at Zeppelin station include a differential mobility particle sizer (DMPS) for measuring the particle size distribution in the diameter range from 10 to 790 nm (time resolution of 20 min), a particle soot absorption photometer (PSAP) for measurements of particle light absorption coefficients at 525 nm (time resolution of 60 min) on a filter, and an integrating nephelometer (TSI model 3563) for measurements of particle light scattering coefficients at the wavelengths of 450, 550, and 700 nm (time resolution of 10 min) (Ström et al., 2003; Tunved et al., 2013).

A high-volume sampler with a PM10 inlet was used to obtain the chemical composition of the Arctic aerosol with time resolutions of one day for sulfate and sea salt and one month for OC/EC during 2008.

Details on the model can be found in Rastak et al. (2014), which has been accepted for publication in ACP. A brief summary of the model performance including the validation of the ambient extinction coefficients is provided in Section 2.2.1 of our revised manuscript. However, using the size distribution to only 800 nm is likely to neglect the contribution of the coarse mode. Previous studies showed that large particles can be responsible for up to 30% of the observed extinction coefficients in the Arctic. This would add an uncertainty of a factor of two to the extinction coefficients obtained with the model of Rastak et al. (2014). To assess the actual underestimation of the extinction coefficient due to not accounting for the course-mode contribution we now also derived the ambient extinction coefficient from the dry nephelometer measurements as proposed by Zieger et al. (2010). See also answer to previous comment.

Pages 5700 - 5702: "Comparison approach" The authors should include information about the uncertainties associated to this approach, e.g.,

We refined our comparison procedure from the simple closest approach method to increase the likelihood for meaningful comparison cases. We missed to state that our comparison approach was actually the result of several steps of refinement. A description of this evolution has been added to the beginning of Section 3. See also first answer to this review.

It is futile to quantify the uncertainty associated with the comparison approach as there are too many possibilities that can render a comparison case physically meaningless. We constrained com-

parisons on a case-by-case basis to the best of our knowledge to ensure the highest possible quality in the reconciliation of the different observation. Simpler comparison scenarios will come with a much higher share of "apples and oranges" comparisons caused by insufficiently accounting for, e.g. atmospheric variability or noisy data. In our case, it is unlikely that a less restrictive comparison approach with the resulting higher number of comparison cases ("better statistics") will be of any advantage as most of these additional cases will consist of physically meaningless comparisons scenarios.

"We believe that time rather than distance is a better parameter to assess changes in the aerosol properties in the atmosphere." Why?

Using range as a constraint is the prime limitation of the closest approach method. This method assumes horizontal homogeneity, and thus, limits the number of comparison cases to a certain distance from a site. However, even for the considered cases the method cannot assure that the resulting comparisons are indeed meaningful. For instance, stagnant conditions with low wind speed or atmospheric flow that does not connect the ground site to the spaceborne observation (i.e. along rather than crossing the ground track) would complicate such a procedure. Accounting for such conditions requires the use of backward trajectories as a means of connecting the different locations of observations. Once the connection is established it is the time scale that determines if we can expect conditions for a meaningful comparison of the different observations. See also first answer to this review.

"A change in the along-track average of the CALIOP extinction profile (i.e., from a range related to crossing trajectories with different starting time at the location of the ground site to a fixed interval) can result in large differences of the resulting mean extinction profile." By how much?

To consistently apply the trajectory matching, we used the along-track averaging criterion described in the paper. Accounting for the spread of trajectories is more physically meaningful than using a fixed part of the ground track. Our analysis showed that the spread of the trajectories along the satellite track varies on a case-by-case basis and that using too long track segments increases the risk of incorporating unrealistic or noisy signals. This could be a feature that is typical for the Arctic. Again, it is impossible to quantify the effect of deviating from an approach that is considered to be as physically meaningful as possible.

"Better agreement with the in-situ observation may be obtained for an average over a smaller height range. However, we chose a conservative range that is likely to be suitable for most cases." Please provide level of uncertainty.

The level of uncertainty depends on the individual extinction profile which can change by an order of magnitude over time or with altitude. Instead of speaking of likelihood we changed the statement and now refer to what has been found during this study:

For particular cases, better agreement with the in-situ observation may be obtained for an average over a smaller height range. However, we chose a conservative range that was found to be suitable for the cases considered in this study.

Page 5704, lines 1 - 2: "Using the in-situ measurements at the time of the satellite overpass decreases the agreement of the observations." How much?

The impact of using in situ observations at the time of the CALIPSO overpass depends on the time delay between the satellite observation and the ground-based measurements as determined from the length of the trajectories. Not accounting for the time delay increases the difference between the extinction coefficients. Here is the shift of the ratio in extinction coefficients for the example cases in Figure 2:

Case 1: 7 h delay, 1.08 changed to 1.94
Case 2: 13 h delay, 1.09 changed to 1.41
Case 3: 9 h delay, 1.31 changed to 1.87
Case 6: 1 h delay, 4.79 changed to 5.72
Case 8: 15 h delay, 1.31 changed to 2.25
Case 9: 12 h delay, 1.29 changed to 1.77

We changed the statement to include a quantification of using improper averages of the in-situ measurements to:

Using the in-situ measurements at the time of the satellite overpass increases the ratio of the ambient extinction coefficients from in-situ and CALIOP observations by 30% for the example cases in Fig. 2.

Page 5704, lines 26 - 28: "There is no indication that a closer distance between satellite ground track and in-situ ground site (or a smaller time lag, not shown) would give a better agreement." Please specify or provide examples, references, etc.

This is a conclusion of our investigation for the cases presented in Figure 4. The color coding of the points in this figure refer to the distance of the CALIPSO observation to the ground station. Points coded with cold colors (closer distances) do not accumulate closer to the 1:1 line than those with warm colors (further distances). We changed the statement to clarify that we are still discussion Figure 4:

According to the color coding of the points in Fig. 4, there is no indication that a closer distance between satellite ground track and in-situ ground site (or a smaller time lag, not shown) would lead to a better outcome of the reconciliation procedure.

Page 5705, lines 20 - 21: "These aerosol types are rather uncommon at 78N and suggest misclassification in the CALIPSO retrieval." Has this been proved? What is CALIPSO's ratio of misclassifications/classifications?

Given the structure of the CALIPSO aerosol classification scheme described in Omar et al. (2009), most CALIOP observations in the Svalbard region should be classified as clean continental (weakly depolarizing, not over desert, integrated attenuated backscatter coefficient $\gamma' < 0.0005$ over land or $\gamma' < 0.0015$ over snow/ice) and clean marine (weakly depolarizing, over ocean, $\gamma' < 0.0015$, not in elevated layer). The discrimination between these two aerosol types is influenced by the location of the observation (over land/snow/ice or water) and the threshold in the total attenuated backscatter coefficient. The other aerosol types require elevated aerosol layers (smoke, not observed), increased depolarization ratios (dust and polluted dust), or increased integrated attenuated backscatter coefficients of $\gamma' > 0.0015$ (polluted continental). The latter two can result from improper cloud screening or the presence of diamond dust. Consequently, we conclude that dust, polluted dust, and polluted continental are the result of misclassification. A closer look at the individual cases reveals that they were either observation with a coinciding presence of clouds in the profile (the two dust cases, two polluted dust cases, one polluted continental case) or that several aerosol types were classified in almost equal parts within the respective layers (two polluted dust cases, three polluted continental cases).

It is hard to give a ratio of misclassification for the CALIPSO retrieval as this would require a reliable benchmark that is not available for observation in the Svalbard region. However, one can assess which aerosol types are more prone to misclassification. Dust, polluted dust, and polluted continental are classified according to the exceedance of certain threshold values of the attenuated backscatter coefficient or the approximate depolarization ratio. Improper cloud screening or noisy signals therefore have a stronger effect on these aerosol types than on clean marine or clean continental — especially in our Arctic cases with generally low signal to noise ratio. The latter two are only separated depending on the location of the observation (i.e. over water or not). To elaborate on this background of the misclassification issue, we now write:

On the other hand, the CALIOP aerosol classification scheme can choose from a larger pool of lidar ratios for observations over ocean and land compared to those over snow and ice (Omar et al., 2009).

Misclassification can occur as a result of signal noise, improper cloud screening, or due to surface effects. Given the structure of the CALIPSO aerosol classification scheme described in Omar et al. (2009), CALIOP observations in the Svalbard region during background conditions (weakly depolarizing and integrated attenuated backscatter coefficient not exceeding the threshold value of 0.0015 at 532 nm) should be classified as clean continental (over land and snow/ice) or clean marine

(over ocean).

Page 5705, lines 25 - 26: "It remains unclear, why half of the clean marine cases are within the set of outliers." Why the authors not consider this as misclassifications?

This has been a conclusion of the nature of the CALIPSO aerosol type classification. If a case of clean marine was misclassified, it could only be clean continental instead (as no threshold values are exceeded). The difference in selecting either type is due to the observation being performed over water rather than land or snow/ice. Hence, clean marine should be properly classified.

We now also present ambient extinction coefficients that are obtained from the nephelometer measurements (revised Section 2.2.2 and revised Figures 4 and 5). These values show better agreement with the CALIOP observations and also enable an estimate of an error range (as a result of using a minimum and maximum estimate of the γ-value). Using these new values improves the comparison for cases classified as clean marine in a way that they no longer stick out. Consequently, we dropped the statement in the revised manuscript.

Page 5706, lines 14 - 16: "The RH at the location of the CALIOP observation is taken from the meteorological data provided with the trajectory analysis and thus highly uncertain." Please quantify.

We want to remind the reader that the value is taken from a model field and that relative humidity is one of the most variable atmospheric parameters. The GDAS fields used by HYSPLIT have a horizontal and temporal resolution of 1° by 1° and 6 h, respectively. In addition, lower tropospheric data have a vertical resolution of 25 hPa and 50 hPa below and above 900 hPa, respectively. We believe no error bar is necessary to realize that these data are highly uncertain.

Manuscript prepared for Atmos. Chem. Phys.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 11 June 2014

Reconciling aerosol light extinction measurements from spaceborne lidar observations and in-situ measurements in the Arctic

M. Tesche[1], **P. Zieger**[1], N. Rastak[1], R. J. Charlson[2], P. Glantz[1], ~~P. Zieger~~[1], **P. Tunved**[1], and H.-C. Hansson[1]

[1]Department of Applied Environmental Science, Stockholm University, Stockholm, Sweden.
[2]Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195, USA.

Correspondence to: Matthias Tesche
(matthias.tesche@itm.su.se)

Abstract. In this study we investigate to what degree it is possible to reconcile continuously recorded particle light extinction coefficients derived from dry in-situ measurements at Zeppelin station (78.92°N, 11.85°E, 475 m above sea level) at Ny-Ålesund, Svalbard, that are recalculated to ambient relative humidity, and simultaneous ambient observations with the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite. To our knowledge, this represents the first study that compares spaceborne lidar measurements to optical aerosol properties from short-term in-situ observations (averaged over 5 h) on a case-by-case basis. Finding suitable comparison cases requires an elaborate screening and matching of the CALIOP data with respect to the location of the Zeppelin station as well as in the selection of temporal and spatial averaging intervals for both the ground-based and spaceborne observations. Trustworthy reconciliation of these data cannot be achieved with the closest approach method that is often used in matching CALIOP observations to those taken at ground sites due to the transport pathways of the air parcels that were sampled. The use of trajectories allowed us to establish a connection between spaceborne and ground-based observations for 57 individual overpasses out of a total of 2018 that occurred in our region of interest around Svalbard (0 to 25°E; 75 to 82°N) in the considered year of 2008. Matches could only be established during winter and spring, since the low aerosol load during summer in connection with the strong solar background and the high occurrence rate of clouds strongly influences the performance and reliability of CALIOP observations. Extinction coefficients in the range from ~~1 to 100~~ **2 to 130** Mm^{-1} **at 532 nm** were found for successful matches with a~~n agreement~~ **difference** of a factor of ~~1.85~~ **1.47** (median value for a range from ~~0.38~~ **0.26** to ~~17.9~~ **11.2**) between the findings of in-situ and spaceborne observations (the latter being generally larger than the former). The remaining dif-

ference is likely to be due to the natural variability in aerosol concentration and ambient relative humidity, an insufficient representation of aerosol particle growth ~~in the used hygroscopicity model~~, or a misclassification of aerosol type (i.e., choice of lidar ratio) in the CALIPSO retrieval.

1 Introduction and motivation

Understanding and quantifying the climatic effects of natural and anthropogenic aerosols from direct observations requires a combination of data from a variety of instruments that usually apply very different measurement techniques. For example, ground-based in-situ measurements of aerosol optical, microphysical, and chemical properties (that are usually carried out with very high temporal resolution but only at a limited number of locations) can be combined with satellite observations or aircraft measurements (that generally provide us with better spatial data coverage but are limited in temporal resolution and/or detail). The combination of such data needs to overcome differences in measurement time, location, and measured quantity. It poses ~~among other~~ the fundamental problem of relating point-sampling data to either spatially-resolved data with poor temporal coverage or airborne measurements without profile information. Four issues arise:

1. **Differences in measurement techniques:** Different properties of the aerosols are sensed or observed by the various instruments. Satellite observations usually are based on optical properties, while in-situ measurements can be of optical properties as well as physical and chemical properties that can be transformed via theory or empirical data (i.e., parameterization) to optical properties (and vice versa).

2. **Spatial resolution:** Location and spatial resolution of the aerosol measurements are different. In-situ observations are often point measurements, while the swath width of passive satellite sensors can extend over up to a few thousand kilometers. In addition, active satellite sensors with narrow footprints often do not cover exactly the location where the in-situ observations are performed. It can also happen that clouds obstruct the ~~otherwise~~ wide field-of-view of a spaceborne sensor. If the satellite data were taken at a distance away from the ground site, it is also necessary to consider the time difference as a lead or a lag of timing.

3. **Hygroscopicity:** The thermodynamic state of the air (especially the relative humidity, RH) has a strong effect on the aerosol optical properties (particularly in the lower marine troposphere) and is different for the different observations. Remote sensing of aerosols is normally performed at ambient condition (i.e., within the atmosphere), while most in-situ instruments ~~dry the probed air during the sampling process before the aerosols are characterized~~ **sample the aerosol at dry conditions with RH $<$ 30%–40%** (*WMO*, 2003).

4. **Temporal resolution:** The time periods over which the observations are averaged may be various. Short temporal averages (i.e., few hours) complicate a comparison since such an

effort is only meaningful, when the different sensors actually observe the same air mass. Long-term averages (i.e., monthly means) on the other hand can generate arbitrary coherence of the data—especially when the considered data sets are of different size.

It is necessary to utilize these simultaneous but disparate data to be able to perform a closure study for the validation of remote-sensing data with independent in-situ measurements and vice versa. Such closure studies are not only important for validating the retrievals of aerosol optical thickness (AOT) or the aerosol extinction coefficient but also to investigate how the measured quantities are apportioned to different types of aerosol, e.g., how large the anthropogenic influence is on the optical properties of the atmosphere, and thus, the radiation balance. For this we have to be able to demonstrate that the measurement systems actually are sensing the same entity. The practical reality (i.e., it is not a simple matter to combine the in-situ and satellite data) is made into a doable but challenging task by the recognition at the outset that both the spaceborne and the in-situ instrument are well-tested devices that are operating correctly within the scope of their capabilities. Thus, the effort described here is not the usual *ground truth* sort of activity done in order to constrain measurement uncertainties. We rather intend to devise methods to bring the data sets into concordance.

Here, we consider in-situ measurements performed at the Arctic station at Mt. Zeppelin (78.92°N, 11.85°E, 474 m above seał level), Svalbard, in comparison with data taken simultaneously (or nearly so) with the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO, *Winker et al.* 2009) satellite. CALIPSO is operating in near-polar orbit at an altitude of about 705 km.

In-situ instruments usually measure aerosol properties under dry conditions with a RH of 10-30% in an indoor laboratory, while ambient conditions are usually associated with much higher RH of up to 100%. Hence, in-situ measurements need to be transformed to ambient conditions by means of direct RH-dependent measurements or a microphysical particle model to account for the loss in particle size due to drying the aerosol particles (*Tang and Munkelwitz*, 1994; *Tang*, 1996; *Zieger et al.*, 2013). On the other hand, ambient aerosol extinction coefficients can be measured directly for instance with active optical remote-sensing techniques such as lidar or differential optical absorption spectroscopy (DOAS). Previous closure studies show that reasonable agreement is found between results obtained from remote sensing of aerosols and findings from in-situ observations when the effect of relative humidity has been accounted for (*Hoff et al.*, 1996; *Masonis et al.*, 2002; *Zieger et al.*, 2011, 2012; *Hoffmann et al.*, 2012; *Ziemba et al.*, 2013; *Skupin*, 2014). However, studies in the literature mainly deal with few single cases during intensive field campaigns rather than systematic comparisons of multi-year data sets.

The clean environment of the Arctic is very sensitive to anthropogenic impacts. Arctic aerosol conditions are also strongly influenced by regional meteorology (*Eneroth et al.*, 2003; *Stock*, 2014), which controls the RH of the air. Changes in this parameter have a huge influence on aerosol particle size, and thus, on light scattering (*Zieger et al.*, 2010, 2013) and cloud formation (*Mauritsen et al.*,

2011) in this region. ~~Measurements with a humidified nephelometer were carried out between 15 July 2008 and 12 October 2008 at Zeppelin station (*Zieger et al.*, 2010). A comparison to Zeppelin's dry nephelometer (operating at RH $< 20\%$) showed that the ambient 550-nm scattering coefficients at RH $= 85 \pm 2\%$ were on average about three times higher (scattering enhancement factor $f(\mathrm{RH}) = 3.24 \pm 0.63$) than the ones of the dried aerosol sample (*Zieger et al.*, 2013).~~ Optical properties and concentrations of Arctic aerosols have been measured at Ny-Ålesund, Svalbard, with in-situ instruments (*Covert and Heintzenberg*, 1993; *Ström et al.*, 2003; *Tunved et al.*, 2013) and by means of remote sensing (*Herber et al.*, 2002; *Hoffmann et al.*, 2009, 2012; *Tomasi et al.*, 2007, 2012) for several years.

Hoffmann et al. (2012) performed a combined analysis of ground-based Raman lidar measurement at Ny-Ålesund and in-situ measurements at Zeppelin station. Instead of the aerosol extinction coefficient they compare the particle number concentration as obtained from a microphysical inversion of the lidar data and measured by the in-situ instruments. As the ground-based lidar data cannot be used to derive aerosol optical properties below 750 m height accurately, measurements at Zeppelin station (474 m height) were instead compared to lidar findings obtained at a height of 850 m. Despite the elaborate comparison approach (e.g., different heights, assumptions in the inversion of lidar data) ~~agreement of a factor of ca. two was found in~~ the total aerosol number concentration for the investigated pollution event on 4 April 2009 **could be reconciled to a factor of ca. two** with smaller lidar-derived values compared to the in-situ measurements.

The use of the spaceborne CALIPSO lidar has the potential to overcome the altitude limitations since its observations extend all the way down to the Earth's surface. The high frequency of overpasses at high latitudes makes it attractive to consider the possibility of a combined analysis of ground-based in-situ and spaceborne lidar measurements in the Arctic. In principle, such an analysis connects information on the vertical and horizontal aerosol distribution from the CALIPSO satellite data to the more specific information about aerosol microphysical and chemical properties at the surface. In-situ measurements are quite limited to a few measurement locations, while satellites can (in principle) view the exact same volume of air that is being sampled at the surface. Satellite sensors also have vastly larger fields of view and allow for global or near-global data coverage. Consequently, they have a strong potential to extend the findings of in-situ measurements in space besides giving information on aerosol optical properties. In the same way, findings from detailed in-situ measurements can add further depth to the satellite observations.

Di Pierro et al. (2013) used these advantages to perform a comprehensive study of the spatial and seasonal distribution of Arctic aerosols based on optical properties observed by CALIOP between 2006 and 2012. The authors introduce an empirical correction that accounts for the different measurement sensitivity during day and night—a crucial factor when it comes to summertime CALIOP observations in the Arctic. The authors found CALIOP aerosol extinction in the Arctic to be of the same order of magnitude as nephelometer observations at Barrow and Alert with the latter being

transformed to ambient RH. However, in addition to using highly averaged data (i.e., monthly and seasonal mean values) the averaging methodology of *Di Pierro et al.* (2013) applies a detection frequency that is defined as the ratio of the number of height bins with detected aerosol layers to the total number of height bins in a given area and time period. This procedure is likely to decrease the magnitude of the obtained mean extinction profiles by introducing zero-values to the averaging. In fact, the authors show that the mean CALIOP extinction profile obtained for a comparison to measurements with a high spectral resolution lidar (HSRL) at Eureka yields much smaller values than the ground-based HSRL observations. *Di Pierro et al.* (2013) also provide the readers with the seasonal variation of CALIOP-derived mean extinction coefficients for different atmospheric layers. Their values for the layer from the surface to 2 km height are a maximum at around 10 Mm^{-1} in March for the Atlantic sector that is most representative for the conditions at Svalbard. This relates to a maximum AOT of 0.02 for the polluted spring season if we assume that the majority of aerosols is present within this 2-km deep layer. Such a value is similar to what is observed in the Arctic troposphere around Svalbard during the clean summer season (*Glantz et al.*, 2014). Note that it is more likely that the aerosol-containing planetary boundary layer at Svalbard is between 0.5 and 1.0 km deep—which would decrease the maximum AOT as derived from the values presented in *Di Pierro et al.* (2013) even further. This discrepancy calls for a more detailed investigation of the factors that influence the reconciliation of extinction coefficients from ground-based and spaceborne observations. We will return to this point in the conclusion.

A description of the instrumentation and the data processing used in this study is presented in Sect. 2. Section 3 describes the methodology for relating segments of individual CALIPSO overpasses to in-situ measurements at Zeppelin station. The findings of the comparison for the year 2008 are discussed in Sect. 4. The paper ends with a summary and conclusions in Sect. 5.

2 Instrumentation and methods

2.1 In-situ measurements at Zeppelin station

The aerosol in-situ instruments at Zeppelin station include a differential mobility particle sizer **(DMPS)** for measuring the particle size distribution in the **mobility** diameter range from 10 to 790 nm (time resolution of 20 min)**, a particle soot absorption photometer (PSAP) for measurements of particle light absorption coefficients at 525 nm (time resolution of 60 min) on a filter,** and an integrating nephelometer (TSI model 3563) for measurements of particle light scattering coefficients at the wavelengths of 450, 550, and 700 nm (time resolution of 10 min) (*Ström et al.*, 2003; *Tunved et al.*, 2013). The nephelometer measurements were corrected for the truncation error and lamp non-idealities according to ***Anderson et al.* (1998)** ~~*Anderson et al.* (1996)~~. All in-situ instruments are placed indoors and connected to an inlet without a particle size cut.

The location of the Zeppelin station at 79°N imposes a severe climatic situation with usually low

outside temperature (from -25 to +15 °C) and correspondingly high RH, often near or at 100%. On the other hand, the in-situ instruments in the laboratory are operated at ordinary room temperature of about 20 °C. Hence, sampled air is heated by as much as 40 K during its transit into the laboratory. Continuous aerosol in-situ observations are usually performed at dry conditions with RH $< 30-40\%$ in order to avoid the influence of water uptake on the aerosol optical properties and to keep the measurements at different ambient RH and at different sites comparable (*WMO*, 2003). ~~However, t~~ **T**he humidity effect on the scattering properties of the aerosol has to be accounted for if results are to represent actual atmospheric conditions.

A high-volume sampler with a PM10 inlet was used to obtain the chemical composition of the Arctic aerosol with time resolutions of one day for sulfate and sea salt and one month for OC/EC during 2008.

Measurements with a humidified nephelometer operating at RH between 20% and 95% were carried out between 15 July 2008 and 12 October 2008 at Zeppelin station (*Zieger et al.*, 2010). A comparison to Zeppelin's dry nephelometer (operating at RH $< 20\%$) showed that the ambient scattering coefficients at RH $= 85\%$ were on average about three times higher than the scattering coefficients of the dried aerosol sample (*Zieger et al.*, 2013). Direct measurements of the scattering enhancement factor were only available for ~~4 months~~ **91 days** in 2008.

2.2 Transferring ~~dry~~ measured ~~optical~~ dry parameters to ambient conditions

Hourly measurements of outdoor humidity at Zeppelin station are available to transform the dry in-situ measurements to ambient conditions. This is done following two approaches by using (1) the the chemical composition of the particles in combination with the particle size distribution from the DMPS as input to a hygroscopicity model and (2) the direct measurements of scattering and absorption coefficients from the nephelometer and PSAP in combination with a scattering enhancement factor f(RH). Cases with ambient RH larger than 95% were considered to be measurements within clouds or fog, and thus, excluded from the procedure.

2.2.1 Site specific hygroscopicity model

~~The aerosol extinction coefficient for ambient conditions is obtained with the help of a hygroscopicity model that uses measurements of outdoor humidity (hourly values), the dry particle number size distribution, and hygroscopicity information from measurements of aerosol chemistry. The model is in detail described in *Rastak et al.* (2014) and here only a brief description is being given.~~

Dry size distributions are transformed to ambient conditions and then used as input for a Mie-scattering model to obtain ambient aerosol optical properties. For a detailed description of this procedure we refer to *Rastak et al.* (2014) while a brief summary is given here.

~~The κ-Köhler theory (*Kreidenweis et al.*, 2005; *Kreidenweis et al.*, 2007) is used to account for~~

~~hygroscopicity effects and to transform the dry particle size distribution to ambient conditions.~~ **Hygroscopicity effects are accounted for with the help of κ-Köhler theory (*Kreidenweis et al.*, 2005; *Petters and Kreidenweis*, 2007).** The aerosol growth factor is derived by combining the individual aerosol volume fractions obtained from the analysis of chemical samples collected at Ny-Ålesund with the hygroscopicity parameter κ of the respective components available in the literature. The components considered in this study are water-soluble and insoluble organics, **ammonium** sulfate, sea salt, and black carbon.

A**mbient a**erosol scattering, absorption, and extinction coefficients are obtained from the humidified aerosol size distribution and refractive index by means of Mie-scattering theory. **All optical properties are calculated at a wavelength of 550 nm and with a temporal resolution of 1 h.** Note that absorption contributes less than ~~7~~5% to the ambient extinction coefficient of Arctic aerosols (*Eleftheriadis et al.*, 2009; *Zieger et al.*, 2010). **This is in agreement with the PSAP measurements at Zeppelin. The effect of light absorption decreases even further when ambient extinction coefficients are considered. The uncertainties of a misrepresentation of aerosol light absorption become negligible when put into the context of the challenges imposed by the comparison procedure described in Section 3.** ~~Cases with ambient RH larger than 95% were considered to be measurements within clouds or fog, and thus, excluded from the procedure. All optical properties are calculated at a wavelength of 550 nm and with a temporal resolution of 1 h.~~

A~~n extensive~~ validation of the microphysical model **is presented in *Rastak et al.* (2014)**. ~~also has been performed~~. ~~A comparison of~~ **Dry** aerosol scattering coefficients measured by the ~~dry~~ nephelometer **agree well with those** ~~and~~ calculated from the particle size distributions ~~results in a~~ **(**slope close to unity**,** ~~and a squared correlation coefficient of~~ $R^2 = 0.95$**)** ~~(*Rastak et al.*, 2014)~~. A comparison between humidified scattering coefficients and measurements with the humidified nephelometer during the ~~three months~~ **91 days** of parallel operation (*Zieger et al.*, 2010) showed **a slight tendency of the model to underestimate the measurements** ~~agreement~~ with $R^2 = 0.64$~~, although with a slight tendency of the model to underestimate the measurements~~ (*Rastak et al.*, 2014). The enhancement factor $f(\mathrm{RH})$ is the ratio of ambient to dry extinction coefficients ~~(*Zieger et al.*, 2013)~~. Values of $f(\mathrm{RH}) = 4.30 \pm 2.26$ with a range from 1.5 to 12.5 were found **when relating the results obtained from the humidified size distribution to the dry nepheloemter measurements** for the year 2008. This is in agreement with the findings of *Zieger et al.* (2010) for Arctic aerosols at ambient RH at Zeppelin station.

~~The box plots in Fig. 1 visualize the importance of transforming dry optical properties to ambient conditions. About 75% of the hourly aerosol scattering coefficients at 550 nm measured with the dry nephelometer at Zeppelin station in 2008 are smaller than 5 Mm^{-1}. Humidity correction to ambient extinction coefficients increases the median value for 2008 from 2 to 7 Mm^{-1} (numbers in the upper part of the figure). On average the ambient extinction coefficient is a factor of three to five larger than the dry one when resolved according to different seasons. The Arctic haze~~

~~period in spring shows the highest median values of the ambient extinction coefficient ($17\,\mathrm{Mm}^{-1}$) followed by winter ($8\,\mathrm{Mm}^{-1}$). Summer and fall are associated with very low median values (3 and $4\,\mathrm{Mm}^{-1}$, respectively). Summer is the slightly cleaner season and a larger variation is observed during fall. This is in agreement with previous observations at Zeppelin station (*Ström et al.*, 2003; *Ström et al.*, 2010; *Ström et al.*, 2013).~~

The humidification of the particle number size distribution ~~$n(r)$~~ **obtained with the DMPS** leads to an increase of the particle effective (surface-weighted) radius ~~which is defined as~~

$$r_{\mathrm{eff}} = \frac{\int r^3 n(r) \mathrm{d}r}{\int r^2 n(r) \mathrm{d}r}$$

from 0.14 ± 0.02 to $0.23 \pm 0.04\,\mu$m (yearly average, not shown). This moves ~~much of~~ the aerosol from an optically ineffective state to a size range in which they are ~~very~~ **more** efficient in interacting with visible light. **Contributions of particles larger than the maximum DMPS size bin would lead to an overall increase in the effective radius, and thus, further improve the light scattering efficiency of the probed aerosol.**

2.2.2 Dry aerosol optical measurements and range of observed f(RH)

The DMPS measurements used in the previous section only cover particles up to a diameter of 790 nm and provide no information on the concentration of larger particles. Particles in the coarse mode can have a large effect on the overall extinction coefficient due to their size and increased extinction efficiency, although they appear in a substantially decreased concentration. Hence, missing even low concentrations of coarse particles can cause an underestimation of the aerosol scattering and extinction coefficients by as much as 30%. In addition, it is more straightforward to determine ambient extinction coefficients directly from the nephelometer measurements if the scattering enhancement factor is known or can be estimated within a reasonable range of values.

Therefore, ambient extinction coefficients were also calculated using the dry absorption and scattering coefficients measured with the PSAP and nephelometer, respectively, together with scattering enhancement factors that represent the median, minimum, and maximum effect of hygroscopic growth on light scattering. Values of $\gamma = 0.57$, 0.35, and 0.85, respectively, were used to obtain the scattering enhancement factor for ambient RH as $f(\mathrm{RH}) = (1 - \mathrm{RH})^{-\gamma}$ (*Zieger et al.*, 2010). Absorption coefficients were assumed not to change with increasing RH.

2.2.3 Dry versus ambient optical properties

The box plots in Fig. 1 visualize the importance of transforming dry optical properties to ambient conditions. About 75% of the hourly aerosol scattering coefficients at 550 nm measured with the dry nephelometer at Zeppelin station in 2008 are smaller than $5\,\mathrm{Mm}^{-1}$. Humidity correction to ambient extinction coefficients increases the median value for 2008 from 2 to

7–10 Mm^{-1}. The differences found in the median values of the ambient extinction coefficients derived according to the two methods described in Sects. 2.2.1 and 2.2.2 is likely to be the effect of coarse-mode particles that are not captured by the DMPS. These particles may contribute to about 20% – 30% of the total extinction coefficient at Zeppelin station (*Zieger et al.*, 2010). The geometric mean has a much lower standard deviation than the arithmetic mean and is similar to the arithmetic median value. Independent of the retrieval method, the ambient extinction coefficient is on average a factor of three to five larger than the dry one when resolved according to different seasons. The Arctic haze period in spring shows the highest median values of the ambient extinction coefficient (17–22 Mm^{-1}) followed by winter (8–12 Mm^{-1}). Summer and fall are associated with very low median values (3–4 and 4–6 Mm^{-1}, respectively). Summer is the slightly cleaner season and a larger variation is observed during fall. This is in agreement with previous observations at Zeppelin station (*Ström et al.*, 2003; *Zieger et al.*, 2010; *Tunved et al.*, 2013).

In the following, we use the ambient extinction coefficients derived from the humidified nephelometer measurements. This is because the lower and upper estimate in the γ-value for the determination of the scattering enhancement provides an uncertainty range that is more reliable than what can be obtained using the model approach described in Sect. 2.2.1.

2.3 CALIOP

The CALIOP is an elastic-backscatter lidar that emits linearly polarized laser light at 532 and 1064 nm wavelength and features three measurement channels. It has been operational since June 2006. An overview of the instrument as well as the data retrieval and interpretation algorithms can be found, i.e., in *Winker et al.* (2009), *Young and Vaughan* (2009), and *Omar et al.* (2009).

2.3.1 Data treatment

For the comparison presented here we use level 2 version 3.01 products with a vertical resolution of 60 m (below 20.2 km height) and a horizontal resolution of 5 km. To derive extinction coefficients for comparison we only considered CALIPSO profiles with Atmospheric_Volume_Description bits 1–3 equal to 3 (feature type = aerosol), a CAD_Score below –20 (screen artifacts from data), and an Extinction_QC_Flag_532 of either 0 or 1. A description of the CALIPSO lidar level 2 5-km cloud and aerosol profile and layer products can be found in the *CALIPSO Users Guide* (2012).

Retrieving extinction coefficients from CALIOP observations requires the assumption of an aerosol-type specific extinction-to-backscatter (lidar) ratio (*Müller et al.*, 2007; *Omar et al.*, 2009). The CALIPSO aerosol model separates between six aerosol types that are selected according to the location of the instrument (surface type) and the detected feature (aerosol layer close to surface or elevated), the intensity of the measured signal (integrated attenuated backscatter coefficient), and an approximated value of the aerosol depolarization ratio (*Omar et al.*, 2009). The considered aerosol

types are: clean marine, dust, polluted continental, clean continental, polluted dust, and smoke. The lowest 532-nm lidar ratio of 20 sr is that of clean marine aerosol, while the highest values of 65 and 70 sr are used for polluted mineral dust, polluted continental aerosol, and biomass-burning smoke. Background conditions are described by the clean continental type that features a lidar ration of 35 sr. Lidar ratios of 30 – 40 sr at 532 nm are reported by *Hoffmann et al.* (2012) and *Stock* (2012) for two cases at Ny-Ålesund during spring 2009 and 2008, respectively. Proper aerosol-type identification is crucial for accurate extinction-coefficient retrievals due to the wide range of available lidar ratios (*Müller et al.*, 2007). Details regarding the CALIPSO lidar-ratio selection algorithm are presented in *Omar et al.* (2009).

2.3.2 Representativeness

To assess the representativeness of the CALIOP measurements in our region of interest around Svalbard it is worthwhile to first examine the availability of lidar profiles and the atmospheric conditions (i.e., the abundance of aerosols and clouds) encountered during these observations. Figure 2a shows the number of monthly available lidar profiles subdivided according to what has been detected in the individual profiles: no features (neither clouds nor aerosols), only aerosols (aerosol features but no cloud features in a profile), only clouds (cloud features but no aerosol features in a profile), or clouds and aerosols (both cloud and aerosol features in a profile). For the entire year 2008, only 5.8% of the considered 187711 profiles show conditions of aerosols only (i.e., no disturbance by clouds) that are most favorable for the type of comparison that we pursue in this study. Best conditions are found during March (15.1% cloud-free profiles with aerosols features) while the summer month (May to September) and particularly July (0.6% cloud-free profiles with aerosol features) represent non-ideal conditions for the comparison of surface measurements and spaceborne observations attempted in this study. About 10% of all CALIOP profiles contain neither aerosol nor cloud features with a maximum and minimum occurrence rate of 25% and 4% in July and January, respectively. This effect is due to the weaker signal-to-noise ratio (SNR) of CALIOP measurements during bright daytime conditions (i.e., polar summer) compared to the absence of sunlight during night and the correspondingly higher threshold value that has to be exceeded for feature detection (*Winker et al.*, 2009; *Young and Vaughan*, 2009). Polar summer and winter can be recognized in the occurrence rate of no features (magenta bars) in Fig. 2a. Observation rates of 50% to 85% for clouds only (during March and August, respectively) illustrate that cloudiness is another main obstacle for deriving aerosol information from CALIOP measurements. Most of these clouds are optically thick and lead to significant or full attenuation of the laser light. As long as these clouds form the uppermost feature, no aerosol detection is possible even if cloud and aerosol layers are present at different height levels.

Figure 2b shows the occurrence rate of the number of height bins with aerosol information for profiles that fall into the categories aerosol only and clouds and aerosols (i.e., profiles identified to

contain aerosol information). Note that the information given in Fig. 2a refers to the entire profile while Fig. 2b refers to the height-resolved observation provided by these profiles. Figure 2b shows that the detection rate of aerosol bins (i.e., the amount of aerosol-containing height bins per profile per month) is much higher during winter, when the background of sunlight is absent and clouds are also less frequent ~~(Fig. 2a)~~. During summer, almost no aerosol features are detected. This is probably due to the decreased SNR of the measurement during daytime, the generally cleaner conditions during this time of the year, or a combination of both. It is also apparent from Fig. 2b that most aerosol features are detected in combination with clouds in the same profile (~~red~~**yellow**) rather than during cloud-free conditions (green). A view at the number of detected aerosol layers given in the CALIPSO products reveals that aerosols ~~are restricted to~~ **occur within** a single layer during the majority of observations (not shown). Multiple aerosol layers are ~~restricted to~~ **are only detected during** polar night. The observation of two layers is already rare while the number of cases with four layers is negligible.

Summarizing Fig. 2, we can conclude that obtaining useful results from CALIOP measurements in the Arctic during summer is improbable and that only a very small fraction of all measurements will occur during cloud-free conditions that favor the kind of study we attempt to perform in this paper. Attempts to overcome the limitations of CALIOP observations during Arctic summer as of *Di Pierro et al.* (2013) who introduced a detection rate for correction are likely to produce incorrect data or will at the least overemphasize the few data available during summer. Nevertheless, it is worthwhile to proceed with our study for the limited number of available cases in order to assess the value of the combined data sets.

2.4 HYSPLIT trajectories

We use the HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory, *Draxler and Rolph* 2010) model of NOAA Air Resources Laboratory to study the advection of air parcels to and from the Zeppelin station. Forward and backward trajectories with time intervals of 1 h were calculated starting and arriving every **3 h** ~~three hours~~ at the height and location of the Zeppelin station, respectively.

Meteorological parameters **from the Global Data Assimilation System (GDAS)** are provided along the trajectories and used in this study to estimate RH at the location of the CALIPSO overpass.

3 Comparison approach

Anderson et al. (2003) and *Kovacs* (2006) investigated the regional representativeness of local measurements of atmospheric aerosols by correlating these to the distance at which coincident satellite observations were performed. They concluded that the distance at which two measurements, both at ambient RH, along a trajectory show acceptable correlation to establish a connection are 300 and

500 km for observations over land and sea sites, respectively. **As a result of these earlier studies, we considered a region from 0 to 25°E and from 75 to 82°N for this study. CALIPSO passed 2018 times over this area in 2008. The closest overpass occured only 2 km away from the Zeppelin station, while the furthest one was at a distance of 360 km.**

~~In 2008, CALIPSO passed 2018 times over the area from 0 to 25°E and from 75 to 82°N that is considered in this study. The closest overpass occurs only 2 km away from the Zeppelin station, while the furthest one was at a distance of 360 km.~~

We started our investigation by applying the closest approach method to link CALIPSO observations in the region of interest to coincident dry in-situ measurements at Zeppelin station. While this course of action led to a high number of matches, it did not enable reasonable case-by-case reconciliation of in-situ and remote-sensing data. Differences in the compared aerosol optical properties ranged between two and three orders of magnitude. Perpetual refinement of the comparison procedure as described below showed that the failure in reconciling the different observations in the initial comparison is due to:

1. **Physically meaningless comparison scenarios in which no connection can be established between the locations of the ground site and the satellite track during heterogeneous aerosol conditions**
2. **The inclusion of apparently unrealistic signal spikes into the CALIOP extinction coefficient in case of fixed or inappropriately selected along-track averaging intervals**
3. **Humidification effects**
4. **The temporal delay in the observations**

The first two points make reasonable comparisons impossible. The latter two can still introduce uncertainties of up to 100%.

Differences in exact location of the measurements pose a severe problem, since the humidity and aerosol content of air is highly variable in time and space (horizontally and vertically). Thus, it is essential to select that part of the CALIPSO ground track for which it is most likely that both CALIOP and in-situ instrumentation actually sampled the same air mass. Following the approach presented in *Tesche et al.* (2013), air-mass trajectories are used to connect the in-situ station to the segment of the CALIPSO ground track that is most ~~suitable for~~ **likely to lead to a physically meaningful** comparison. The length of the trajectories between Zeppelin station and the intersection with the CALIPSO ground track provides us with the time lag between fitting observations. **This trajectory matching allows to address items 1 and 4 on the list above.**

Screening of the CALIPSO data is a major effort in obtaining meaningful comparison cases. Our case-by-case investigation shows that profiles fulfilling the quality assurance criteria given in Section 2.3.1 can still contain data points that are obviously unrealistic and could be due to

the low SNR of the observation or improper cloud-screening. Though such points have little impact when comparing highly averaged data, they dominate individual comparisons. ~~In a first step of screening the CALIPSO data~~ **Here**, we selected ~~only those~~ overpasses that **in fact** ~~actually~~ show extinction coefficients (i.e., signals above the CALIOP detection threshold) in a height range from 250 to 730 m that spans around the height of the Zeppelin station. This holds for 24% of all overpasses in the area of interest. **Next, we discarded cases for which trajectories starting every 3 h at Zeppelin for 15 h after and before an overpass, respectively, did not cross the CALIPSO ground track. This left 9% of all 2018 overpasses in 2008.** ~~For these cases, we investigated whether backward and forward trajectories starting every 3 h for 15 h after and before the CALIPSO overpass are actually crossing the ground track (second step). Cases with no such intersections were discarded from the investigation. This left 9% of all 2018 overpasses in 2008.~~ Note that in contrast to ~~the studies by~~ *Anderson et al.* (2003) and *Kovacs* (2006) that referred to the length~~s~~ scale we use a time scale and restrict the comparison to a time delay of 15 h. This corresponds to a maximum distance of 360 km at a mean transport velocity of about 7 m/s. We believe that time rather than distance is a better parameter to assess changes in the aerosol properties in the atmosphere. The majority of the track segments for comparison were located either in the vicinity or to the north (beyond 81°N) of the ground site (not shown).

~~In the third and final step~~ **Finally**, we checked for the availability of (1) CALIOP extinction coefficients at the intersection of satellite ground track and air-mass trajectories and of (2) humidified extinction coefficients at Zeppelin station at the time of the CALIPSO overpass plus/minus the lag provided by the trajectories. That was the case for only 57 individual overpasses (3% of all 2018 overpasses) in 2008, which form the core of this study. The extinction coefficients from CALIOP were averaged in the vicinity of the crossing point of the ground track and the trajectory. The along-track averaging range was determined individually for each overpass according to the spread of the crossing trajectories **with different start times**. A change in the along-track average of the CALIOP extinction profile ~~(i.e., from a range related to crossing trajectories with different starting time at the location of the ground site~~ to a fixed interval~~)~~ can result in large differences of the resulting mean extinction profile **during heterogeneous conditions or physically meaningless comparison scenarios**. Once an extinction profile could be obtained at the proper location for comparison, the values in the height range from 250 to 730 m (eight 60-m height bins) were averaged ~~eight 60-m height bins~~. We chose this height range to account for vertical motions during the transport from the location of the CALIOP observation to Zeppelin station (backward trajectories) or ~~the other way round~~ **vice versa** (forward trajectories). **For particular cases, b**~~B~~etter agreement with the in-situ observation may be obtained for an average over a smaller height range. However, we chose a conservative range that ~~is likely~~ **was found** to be suitable for ~~most~~ the **majority of** cases **considered in this study**. The average and the corresponding standard deviation (as a measure of vertical homogeneity) represent values used in the comparison to the findings of the measurements at Zeppelin

station. To coarsely account for uncertainties in the trajectories, in-situ extinction coefficients were averaged over five hours (five 1-h values) centered around the time during which the in-situ instruments sampled the same air parcels as CALIOP, i.e., time of a CALIPSO overpass plus the time lag determined from the length of the trajectories that connect this overpass to Zeppelin station.

4 Results and discussion

The time period from 22 to 28 January 2008 has been chosen to illustrate the analytical work and some of the results obtained. Figure 3 presents the dry scattering coefficient measured with the nephelometer at Zeppelin station and the ambient extinction coefficient**s derived as described in Sects. 2.2.1 and 2.2.2** ~~calculated from the humidified size distribution~~ during this period. The ambient RH given in the figure reflects the influence of hygroscopicity which causes the huge differences between dry scattering and ambient extinction values. The latter parameter has not been estimated when ambient RH exceeded values of 95%. The time period covered in Fig. 3 shows ten CALIPSO overpasses that were connected to the ground station with the help of trajectories (see ~~colored triangles~~ **symbols** and corresponding numbers at the top of Fig. 3). Extinction coefficients extracted from the CALIPSO observations could be compared to ground-based measurements for six cases (overpasses 1, 2, 3, 6, 8, and 9). Four examples of how trajectories are used to connect the ground site with the proper segment of the CALIPSO track (overpasses 1, 6, 8, and 10) are given in the lower part of Fig. 3. ~~Green t~~**T**riangles mark cases for which aerosol profiles were obtained during cloud-free conditions as indicated by a cloud optical thickness (COT) of zero. The examples of overpasses 1 and 8 show how the trajectories lead to a cloud-free part of the ground track. The different lengths and tracks of the trajectories indicate that time and distance should not be considered as synonyms. The satellite- and ground-based extinction coefficients agree within **their error bars** ~~factors of 1.1 – 1.3~~ for the overpasses on 22 and 27 January 2008 with the shortest time delay of 6 h (201 km distance) and the longest time delay of 15 h (322 km distance). Note that ambient RH was above 90% on 22 January 2008 and that the difference between the dry scattering coefficient and the RH-corrected extinction coefficient is as much as a factor of 10. A much smaller ratio of ambient to dry extinction coefficients can be found for 26 and 27 January 2008, for which RH varies between 65% and 90%. The ~~green~~ cases in Fig. 3 illustrate the importance of accounting for the proper time delay between the measurements of CALIOP and in-situ instrumentation. Using the in-situ measurements at the time of the satellite overpass ~~decreases the agreement of the observations~~ **increases the ratio of the ambient extinction coefficients from in-situ and CALIOP observations by 30% for the example cases in Fig. 3**.

Using the trajectories as described above, a cloudy part of the CALIPSO ground track (COT>0, AOT$=0$) was identified for the overpasses 4, 5, 7, and 10. No comparisons could be performed since there is no aerosol information available for these cases. This kind of situation inhibited comparisons

in 127 cases for the months January to April and October to December 2008. Typical scenarios are: no height bins are marked as containing aerosols at all, all aerosols are located above or below our height range of interest, or the obtained aerosols profile is of unreasonable shape and/or magnitude.

~~However, f~~**F**or overpass 6 ~~(blue triangle~~ in Fig. 3~~)~~ aerosol information was obtained in cloudy environment (COT> 0, AOT> 0). Even though this overpass occurred only 21 km from the ground site, the CALIPSO observation is in ~~poor~~ **dis**agreement with the result of the in-situ measurement. This emphasized that using a closest approach for comparison of ground-based measurements and CALIPSO observations might not always be the best choice. The case also illustrates that even few clouds can disturb aerosol measurements with spaceborne lidar. Note also that trajectories might actually lead to a track segment that is not closest to the ground site, as is the case for overpass 8.

Finally, 57 cases of the 2018 overpasses in 2008 were suitable for comparing extinction coefficients from CALIOP observations and humidified ground-based measurements (Fig. 4). Even though CALIOP extinction coefficients are generally larger than the ones derived from the in-situ measurements, most comparison~~s agree~~ **can be reconciled** to a factor of one to five with a majority not exceeding a factor of two. This is a surprisingly good finding considering the data processing that is necessary to come up with comparable quantities. **According to the color coding of the points in Fig. 4**, ~~T~~**t**here is no indication that a closer distance between satellite ground track and in-situ ground site (or a smaller time lag, not shown) would ~~give~~ **lead to** a better ~~agreement~~ **outcome of the reconciliation procedure**. ~~Suitable agreement~~ **Successful reconciliation** actually occurs for many cases associated with overpasses at larger distances from the ground site. These cases would not have been included in this study if we had chosen a distance in range rather than time for comparison. ~~In fact, Fig. 4 shows that most of the cases exceeding a factor of five in agreement correspond to comparisons at distances of 120 km or less (blue, light blue, and light green dots in the upper left part of the plot).~~ This suggests that the method of comparing local point or column-integrated measurements to the closest-approach observation of CALIPSO is likely to yield misleading results.

We performed a deeper analysis of the factors that could explain why a difference of as large as a factor of five occurs for some of the cases included here. Besides the spatial distance and temporal delay between the observations we considered the relative humidity at Zeppelin station and at the crossing point of the satellite ground track and trajectories, the occurrence of clouds and rain along the trajectory, and the wind direction at the ground site. However, only the latter parameter could be linked to the outliers in Fig. 4. Figure 5a shows that the largest absolute difference in the ambient extinction coefficients from CALIOP and in-situ measurements occur during westerly flow. It could be that aerosol conditions are more stable for air masses approaching Zeppelin station from the north and via ice-covered ocean compared to the open water to the west. **On the other hand, the CALIOP aerosol classification scheme can choose from a larger pool of lidar ratios for observations over ocean and land compared to those over snow and ice (*Omar et al.*, 2009). Hence**~~In addition~~, we investigated the dominant aerosol type selected in the CALIPSO data retrieval for the

individual comparisons. It was found that the most characteristic outliers in Figs. 4 and 5a occur for cases that were identified predominantly as polluted dust~~,~~ **or** polluted continental~~, and dust in the CALIPSO retrieval~~. These aerosol types are rather uncommon at 78°N and suggest misclassification in the CALIPSO retrieval. Misclassification can occur as a result of signal noise**, improper cloud screening,** or due to surface effects. **Given the structure of the CALIPSO aerosol classification scheme described in *Omar et al.* (2009), CALIOP observations in the Svalbard region during background conditions (weakly depolarizing and integrated attenuated backscatter coefficient not exceeding the threshold value of 0.0015 at 532 nm) should be classified as clean continental (over land and snow/ice) and clean marine (over ocean).**

Clean continental aerosol ~~(i.e., background conditions)~~ was classified for most comparison cases (see color coding in Fig. 5a~~) and seems to be the most appropriate choice of aerosol together with clean marine. It remains unclear, why half of the clean marine cases are within the set of outliers~~. In addition, classifying aerosol features as polluted dust or smoke (lidar ratio of 65 – 70 sr) instead of clean continental aerosol (lidar ratio of 35 sr) will only result in a factor of two difference, while the disagreement we obtain in our comparison for cases classified as something other than clean continental **and clean marine** shows factors in the range from ~~6 to 18~~ **0.62 to 11.23 with a median of 4.03**. The range is ~~0.38 to 5~~ **0.26 to 5.72 with a median of 1.36** for cases classified as clean continental **or clean marine**.

Strong variation in RH between the location of the CALIPSO ground track and Zeppelin station could also cause the scatter of values presented in Fig. 4. Such RH differences have a direct effect on the scattering enhancement factor $f(\mathrm{RH})$, and thus, on the difference between dry and ambient extinction coefficients. The scattering enhancement factor was found to be much higher for Arctic aerosol compared to observations at continental, background, or marine sites (*Zieger et al.*, 2013). Consequently, we should expect that even small differences in RH between the measurements at Zeppelin and along the satellite track can lead to high differences in the ambient extinction coefficient. This holds especially for high RH $> 85\%$. We investigated if we can find a connection between the difference in RH (ΔRH) at the two measurement locations (i.e., the CALIOP ground-track segment and Zeppelin station) and the agreement in the comparison of ambient extinction coefficients at those sites. The RH at the location of the CALIOP observation is taken from the meteorological data provided with the trajectory analysis and thus highly uncertain. For the considered 57 cases, the ΔRH showed a mean value of $12 \pm 10\%$ (mean RH of $80 \pm 12\%$ at Zeppelin station) with a maximum value of around 30% (not shown). Though ΔRH was considerable for several cases, we could not establish that this factor or the resulting difference in $f(\mathrm{RH})$ can fully explain the disagreement found in the ambient extinction coefficients. Figure 5b) shows the connection between the relative difference in $f(\mathrm{RH})$ at the locations of CALIOP and in-situ observations and the relative difference in the ambient extinction coefficients obtained from these observations. Values should align along the 1:1 line, if hygroscopic growth was the only factor we would have to consider in our compari-

555 son. Deviations are likely to be related to the observation of different air masses at the two locations ~~(i.e., different aerosol size distributions or chemical composition)~~ or the improper representation of meteorological parameters ~~(i.e., RH)~~ in the trajectory model.

Table 1 gives a detailed overview of the results obtained from the comparison of spaceborne and ground-based observations subdivided according to the months of 2008 and to whether cloud-free
560 or cloudy CALIOP aerosol profiles were used in the comparison. For the 57 considered cases Tab. 1 shows that time delay is rather evenly distributed between 0 and 15 h with a median of 8 h. 39 of the 57 suitable cases occurred during most favorable cloud-free conditions (AOT> 0, COT$= 0$), while the remaining 18 cases represent cloudy comparisons (AOT> 0, COT> 0). Resolving the comparison according to cloudiness in the CALIPSO observations (not shown) leads to ambiguous
565 results: for 7 of the 18 cloudy cases (39%) a difference larger than a factor of two is found between the extinction coefficients from CALIOP and Zeppelin station, while for the cloud-free cases 17 out of 39 (44%) exceed this difference. The average time delay is 9.2 ± 3.8 h for cloud-free cases, while it is only 6.2 ± 3.9 h for cloudy cases. Accordingly, cloud-free cases show a mean distance of 228 ± 100 km and cloudy ones 178 ± 116 km. Extinction coefficients from CALIPSO vary between
570 4.6 and 127.0 Mm^{-1} for cloud-free cases, while the range of values for cloudy profiles is much narrower and only spans from 14.3 to 91.6 Mm^{-1}.

5 Summary and conclusions

This study presents a comparison of extinction coefficients as determined from spaceborne lidar (CALIOP) measurements and from ground-based in-situ measurements at Zeppelin station, Ny-
575 Ålesund, Svalbard, during the year 2008. To obtain meaningful comparison, we had to consider several issues:

1. Neither in-situ instruments nor spaceborne lidar (CALIOP) provide us with direct measurements of the ambient aerosol extinction coefficient.

2. Approved methods were used to obtain ambient extinction coefficients from dry in-situ mea-
580 surements **performed with commonly used instruments** ~~of the particle size distributions in combination with a multi-component growth factor based on κ-Köhler theory and measured chemical aerosol properties~~.

3. Extinction coefficients from the spaceborne sensor were taken from operational CALIPSO products that underwent elaborate calibration and quality assurance.

585 4. Air-mass trajectories were used to ensure that ~~all measurements~~ compar**isons**~~ed~~ were performed ~~within~~ **for** the same air mass. Th**ey allow** ~~is is necessary~~ to establish a connection between the satellite's ground track and Zeppelin station and to adapt the **along-track** averaging intervals ~~along the CALIPSO ground track~~ according to the spatial spread of the crossing

trajectories. The averaging height range of 510 m ~~was~~ centered at the elevation of the ground site ~~and~~ **was** chosen to account for vertical displacement during travel along the trajectories. Temporal averaging of ground-based data of 5 h ~~was introduced to~~ further mitigate**s** imprecision in the trajectory output.

The detailed matching procedure used in this study reduced the number of comparison cases from over 2000 overpasses in 2008 to 57 overpasses during 42 days of that year. Even though it is a costly and elaborate case-by-case comparison it is likely to yield more significant results than what is obtained by comparing monthly means of surface measurements with monthly regional means of CALIOP observations. However, since averaging times of only a few hours were applied in this study, we cannot draw conclusions about what will happen if the length of the temporal averaging window is increased. The median ambient extinction coefficient for the 57 comparison cases was $27.8\,\mathrm{Mm}^{-1}$ for the CALIOP data compared to a value of $14.3\,\mathrm{Mm}^{-1}$ **and $20.7\,\mathrm{Mm}^{-1}$ derived from** ~~for~~ in-situ measurements **of the particle size distribution and dry scattering coefficients, respectively.** ~~that were corrected to ambient conditions.~~ The different humidity during the measurement in the atmosphere and within a laboratory is an omnipresent limitation for studies like the one presented here. The thermodynamic state (e.g., RH) of the samples and the assumptions on the hygroscopic properties for the in-situ measurements are therefore vital factors for a successful comparison of aerosol extinction coefficients. In the case of our study, results are also influenced by the CALIPSO aerosol model that is required for the extinction-coefficient retrieval, the CALIOP feature detection limit, and the criteria that are used to match satellite observations to the measurement at the ground site.

Detailed knowledge of the humidity field is of vital importance when relating in-situ measurements to observations with spaceborne sensors. The effect of relative humidity on the light scattering properties of aerosol particles in the atmosphere is the dominating obstacle for a systematic reconciliation of measurements of the two platforms. Additional disturbing factors in the allocation procedure applied in this study were unfavorable wind direction (no intersection between trajectories and ground track), presence of clouds (RH$>95\%$ at Zeppelin station and/or no aerosol information from CALIOP), no data from Zeppelin station or CALIOP, and the CALIOP detection threshold that prevents reliable aerosol detection in the presence of sunlight. CALIOP detects almost no aerosol features in the Svalbard region during Arctic summer even though the tropospheric median AOT is generally larger than 0.05 at visible wavelengths during May and June (*Tomasi et al.*, 2007, 2012; *Glantz et al.*, 2014). This is in agreement with a study by *Di Pierro et al.* (2013) that investigated the distribution of aerosols in the Arctic from CALIOP measurements. Consequently, CALIOP data have to be treated with great caution when they are used for studies of aerosol occurrence rate, transport patterns, radiative effects, and interactions with clouds under background conditions during polar day.

Based on the study presented here we also conclude that consolidating data sets that are averaged over large areas and/or long time periods can lure us into a feeling of arbitrary confidence, while there may actually be weak or no connection between individual observations. Using highly averaged parameters in the deduction of scientific findings is of particular importance for the validation of model simulations. Consequently, special emphasis should be placed on a proper selection of temporal and spatial averaging intervals when attempting to use spaceborne lidar observations in connection to ground-based measurements and model outputs.

Acknowledgements. CALIOP data used in this study were obtained from the NASA Langley Research Center Atmospheric Science Data Center (http://eosweb.larc.nasa.gov). We thank Birgitta Noone and Tabea Hennig for maintaining the in-situ instruments at Zeppelin station. The Swedish Environmental Agency and Stockholm University are acknowledged for long-term support of the observations at Zeppelin station. **We also acknowledge the support within CLIMSLIP.** Paul Zieger was supported by a PostDoc fellowship of the Swiss National Science Foundation (grant no. P300P2_147776).

References

~~Anderson, T., Covert, D., Marshall, S., Laucks, M., Charlson, R., Waggoner, A., Ogren, J., Caldow, R., Holm, R., Quant, F., Sem, G., Wiedensohler, A., Ahlquist, N., and Bates, T.: Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer, J. Atmos. Ocean. Tech., 13, 967–986, 1996.~~

Anderson, T. and Ogren, J.: Determining aerosol radiative properties using the TSI 3563 integrating nephelometer, Aerosol Sci. Technol., 29, 57–96, 1998.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale variations of tropospheric Aerosols, J. Atmos. Sci., 60, 119–136, 2003.

CALIPSO Users Guide (2012), Lidar Level 2 5 km Cloud and Aerosol Profile Products, available at: http://www-calipso.larc.nasa.gov/resources/calipso_users_guide/data_sum-maries/profile_data.php, last accessed on 6 February 2014.

Covert, D. S. and Heintzenberg, J.: Size distributions and chemical properties of aerosol at Ny-Ålesund, Svalbard, Atmos. Environ., 27, 2989–2997, 1993.

Di Pierro, M., Jaeglé, L., Eloranta, E. W., and Sharma, S.: Spatial and seasonal distribution of Arctic aerosols observed by the CALIOP satellite instrument (2006–2012), Atmos. Chem. Phys., 13, 7075–7095, doi:10.5194/acp-13-7075-2013, 2013.

Draxler, R. R. and Rolph, G. D.: HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (http://ready.arl.noaa.gov/HYSPLIT.php). NOAA Air Resources Laboratory, Silver Spring, MD, 2010.

Eleftheriadis, K., Vratolis, S., and Nyeki, S.: (2009), Aerosol black carbon in the European Arctic: Measurements at Zeppelin station, Ny-Ålesund,Svalbard from 1998–2007, Geophys. Res. Lett., 36, L02809, doi:10.1029/2008GL035741, 2009.

Eneroth, K., Kjellström, E., and Holmén, K.: A trajectory climatology for Svalbard; investigating how atmospheric flow patterns influence observed tracer concentrations, Phys. Chem. Earth, 28, 1191–1203, doi:10.1016/j.pce.2003.08.051, 2003.

Glantz, P., Bourassa, A. E., Herber, A., Iversen, T., Karlsson, J., Kirkevåg, A., Maturilli, M., Seland, Ø., Stebel, K., Struthers, H., Tesche, M., and Thomason, L.: Remote sensing of aerosols in the Arctic for an evaluation of global climate model simulations, J. Geophys. Res., **accepted manuscript** ~~under review~~, 2014.

Herber, A., Thomason, L. W., Gernandt, H., Leiterer, U., Nagel, D., Schulz, K.-H., Kaptur, J., Albrecht, T., and Notholt, J.: Continuous day and night aerosol optical depth observations in the Arctic between 1991 and 1999, J. Geophys. Res., 107, AAC 6-1–ACC 6-13, doi:10.1029/2001JD000536, 2002.

Hoff, R., Guise-Bagley, L., Staebler, R. M., Wiebe, H. A., Brook, J., Georgi, B., and Düsterdiek, T.: Lidar, nephelometer, and in situ aerosol experiments in southern Ontario, J. Geophys. Res., 101, 19199–19209, doi:10.1029/95JD03228, 1996.

Hoffmann, A., Ritter, C., Stock, M., Shiobara, M., Lampert, A., Maturilli, M., Orgis, T., Neuber, R., and Herber, A.: Ground-based lidar measurements from Ny-Ålesund during ASTAR 2007, Atmos. Chem. Phys., 9, 9059–9081, doi:10.5194/acp-9-9059-2009, 2009.

Hoffmann, A., Osterloh, L., Stone, R., Lampert, A., Ritter, C., Stock, M., Tunved, P., Hennig, T., Böckmann, C., Li, S.-M., Eleftheriadis, K., Maturilli, M., Orgis, T., Herber, A., Neuber, R., and Dethloff, K.: Remote

sensing and in-situ measurements of tropospheric aerosol, a PAMARCMiP case study, Atmos. Environ., 52, 56–66, doi:10.1016/j.atmosenv.2011.11.027, 2012.

Kovacs, T.: Comparing MODIS and AERONET aerosol optical depth at varying separation distances to assess ground-based validation strategies for spaceborne lidar, J. Geophys. Res., 111, D24203, doi:10.1029/2006JD007349, 2006.

Kreidenweis, S. M., Koehler, K., DeMott, P. J., Prenni, A. J., Carrico, C., and Ervens, B.: Water activity and activation diameters from hygroscopicity data - Part I: Theory and application to inorganic salts, Atmos. Chem. Phys., 5, 1357–1370, doi:10.5194/acp-5-1357-2005, 2005.

Mauritsen, T., Sedlar, J., Tjernström, M., Leck, C., Martin, M., Shupe, M., Sjogren, S., Sierau, B., Persson, P. O. G., Brooks, I. M., Swietlicki, E.: An Arctic CCN-limited cloud-aerosol regime, Atmos. Chem. Phys., 11, 165–173, doi:10.5194/acp-11-165-2011, 2011.

Masonis, S., Ansmann, A., Müller, D., Althausen, D., Ogren, J. A., Jefferson, A., and Sheridan, P. J.: An intercomparison of aerosol light extinction and 180° backscatter as derived using in situ instruments and Raman lidar during the INDOEX field campaign, J. Geophs. Res., 107, INX2 13-1–INX2 13-21, doi:10.1029/2000JD000035, 2002.

Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., and Pisani, G.: Aerosol-type-dependent lidar ratios observed with Raman lidar, J. Geophys. Res., 112, D16202, doi:10.1029/2006JD008292, 2007.

Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., Lee, K.-P., and Hostetler, C. A.: The CALIPSO automated aerosol classification and lidar ratio selection algorithm, J. Atmos. Oceanic Technol., 26, 1994–2014, doi:10.1175/2009JTECHA1231.1, 2009.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961–1971, doi:10.5194/acp-7-1961-2007, 2007.

Rastak, N., Silvergren, S., Zieger, P., Wideqvist, U., Ström, J., Svenningsson, B., Maturilli, M., Tesche, M., Ekman, A. M. L., Tunved, P., and Riipinen, I.: Seasonal variation of aerosol water uptake and its impact on the direct radiative effect at Ny-Ålesund, Svalbard, ~~manuscript submitted to~~ Atmos. Chem. Phys.~~Discuss.~~, **14, accepted manuscript, doi:10.5194/acpd-14-7067-2014**, 2014.

Skupin, A., Ansmann, A., Engelmann, R., and Baars, H.: Spectral Aerosol Extinction Monitoring System (SÆMS): setup, observational products, and comparisons, Atmos. Meas. Tech.~~ Discuss.~~, ~~6~~**7**, 701-712, ~~10.5194/amtd-6-8647-2013~~ doi:10.5194/amt-7-701-2014, 2014.

Stock, M., Ritter, C., Herber, A., von Hoyningen-Huene, W., Baibakov, K., Gräser, J., Orgis, T., Treffeisen, R., Zinoviev, N., Makshtas, A., and Dethloff, K.: Springtime Arctic aerosol: Smoke versus haze, a case study for March 2008, Atmos. Environ., 52, 48–55, doi:10.1016/j.atmosenv.2011.06.051, 2012.

Stock, M., Ritter, C., Aaltonen, V., Aas, W., Handorff, D., Herber, A., Treffeisen, R., and Dethloff, K.: Where does the optically detectable aerosol in the European Arctic come from?, Tellus B, 66, 21450, doi:10.3402/tellusb.v66.21450, 2014.

Ström, J., Umegård, J., Tørseth, K., Tunved, P., Hansson, H.-C., Holmén, K., Wismann, V., Herber, A., and König-Langlo, G.: One year of particle size distribution and aerosol chemical composition measurements at the Zeppelin station, Svalbard, March 2000 – March 2001, Phys. Chem. Earth, 28, 1181–1190, doi:10.1016/j.pce.2003.08.058, 2003.

Tang, I. N. and Munkelwitz, R. H.: Aerosol Phase Transformation and Growth in the Atmosphere, J. Appl. Meteor., 33, 791–796, 1994.

Tang, I. N.:, Chemical and size effects of hygroscopic aerosols on light scattering coefficients, J. Geophys. Res., 101, 19245–19250, doi:10.1029/96JD03003, 1996.

Tesche, M., Wandinger, U., Ansmann, A., Müller, D., Althausen, D., and Omar, A. H.: Ground-based validation of CALIPSO observations of dust and smoke in the Cape Verde region, J. Geophys. Res., 118, 2889–2902, doi:10.1029/2012JD018877, 2013.

Tomasi, C., Vitale, V., Lupi, A., Di Carmine, C., Campanelli, M., Herber, A., Treffeisen, R., Stone, R. S., Andrews, E., Sharma, S., Radionov, V., von Hoyningen-Huene, W., Stebel, K., Hansen, G. H., Myhre, C. L., Wehrli, C., Aaltonen, V., Lihavainen, H., Virkkula, A., Hillamo, R., Ström, J., Toledano, C., Cachorro, V. E., Ortiz, P., de Frutos, A. M., Blindheim, S., Frioud, M., Gausa, M., Zielinski, T., Petelski, T., Yamanouchi, T.: Aerosols in polar regions: A historical overview based on optical depth and in situ observations, J. Geophys. Res., 112, D16205, doi:10.1029/2007JD008432, 2007.

Tomasi C., Lupi, A., Mazzola, M., Stone, R. S., Dutton, E. G., Herber, A., Radionov, V. F., Holben, B. N., Sorokin, M. G., Sakerin, S. M., Terpugova, S. A., Sobolewski, P. S., Lanconelli, C., Petkov, B. H., Busetto, M., and Vitale, V.: An update on polar aerosol optical properties using POLAR-AOD and other measurements performed during the International Polar Year, Atmos. Environ., 52, 29–47, doi:10.1016/j.atmosenv.2012.02.055, 2012.

Tunved, P., Ström, J., and R. Krejci, R.: Arctic aerosol life cycle: Linking aerosol size distributions observed between 2000 and 2010 with air-mass transport and precipitation at Zeppelin station, Ny-Ålesund, Svalbard, Atmos. Chem. Phys., 13, 3643–3660, doi:10.5194/acp-13-3643-2013, 2013.

Winker, D. M., Vaughan, M. A., Omar, A. H., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, J. Atmos. Oceanic Technol., 26, 2310–2323, doi:10.1175/2009JTECHA1281.1, 2009.

World Meteorological Organization: Aerosol measurement procedures guidelines and recommendations, GAW Report No. 153, World Meteorological Organization Global Atmosphere Watch, Geneva, Switzerland, 2003.

Young, S. A. and Vaughan, M. A.: The retrieval of profiles of particulate extinction from Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, J. Atmos. Oceanic Technol., 26, 1105–1119, doi:10.1175/2008JTECHA1221.1, 2009.

Zieger, P., Fierz-Schmidhauser, R., Gysel, M., Ström, J., Henne, S., Yttri, K. E., Baltensperger, U., and Weingartner, E.: Effects of relative humidity on aerosol light scattering in the Arctic, Atmos. Chem. Phys., 10, 3875–3890, doi:10.5194/acp-10-3875-2010, 2010.

Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M., Petäjä, T., Clémer, K., van Roozendael, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T., Shaiganfar, R., Beirle, S., Apituley, A., Wilson, K., and Baltensperger, U.: Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw, Atmos. Chem. Phys., 11, 2603–2624, doi:10.5194/acp-11-2603-2011, 2011.

Zieger, P., Kienast-Sjögren, E., Starace, M., von Bismarck, J., Bukowiecki, N., Baltensperger, U., Wienhold, F. G., Peter, T., Ruhtz, T., Collaud Coen, M., Vuilleumier, L., Maier, O., Emili, E., Popp, C., and Weingartner, E.: Spatial variation of aerosol optical properties around the high-alpine site Jungfraujoch (3580 a.s.l.),

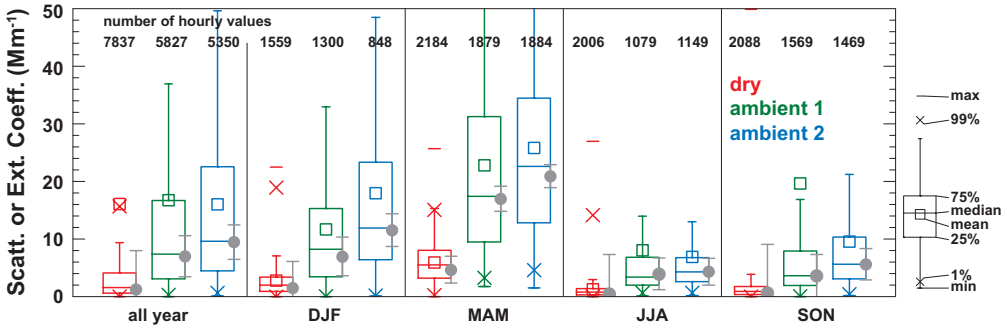

Fig. 1. Statistical overview of the dry scattering (red) and ambient extinction ~~(green)~~ coefficients at 550 nm based on hourly measurements at Zeppelin station in 2008 according to the entire year and the different seasons winter (DJF), spring (MAM), summer (JJA), and autumn (SON). **The ambient extinction coefficients refer to the results obtained by using humidified size distributions from DMPS measurements in combination with Mie-scattering theory (ambient 1, green) and the dry nephelometer and PSAP measurements in combination with a scattering enhancement factor derived for a mean γ of 0.57 (ambient 2, blue).** The numbers in the top of the figure mark the number of available hourly measurements**.**~~, mean, standard deviation, and median values for an arithmetic mean of the data, as well as the mean and standard deviation for a geometric mean of the data (gray circles) for the different time periods. The geometric mean has a much lower standard deviation than the arithmetic mean and is similar to the arithmetic median value.~~ The difference in data availability for dry scattering and ambient extinction coefficients is the consequence of cloud screening and an absence of input data required for humidity correction.

Atmos. Chem. Phys., 12, 7231–7249, doi:10.5194/acp-12-7231-2012, 2012.

Zieger, P., Fierz-Schmidhauser, R., Baltensperger, U., and Weingartner, E.: Effects of relative humidity on aerosol light scattering: Results from different European sites, Atmos. Chem. Phys., 13, 10609–10631, doi:10.5194/acp-13-10609-2013, 2013.

Ziemba, L. D, Thornhill, K. L., Ferrare, R., Barrick, J., Beyersdorf, A. J., Chen, G., Crumeyrolle, S. N., Hair, J., Hostetler, C., Hudgins, C., Obland, M., Rogers, R., Scarino, A. J., Winstead, E. L., Anderson, B. E.: Airborne observations of aerosol extinction by in situ and remote-sensing techniques: Evaluation of particle hygroscopicity, Geophys. Res. Lett., 40, 417–422, doi:10.1029/2012GL054428, 2013.

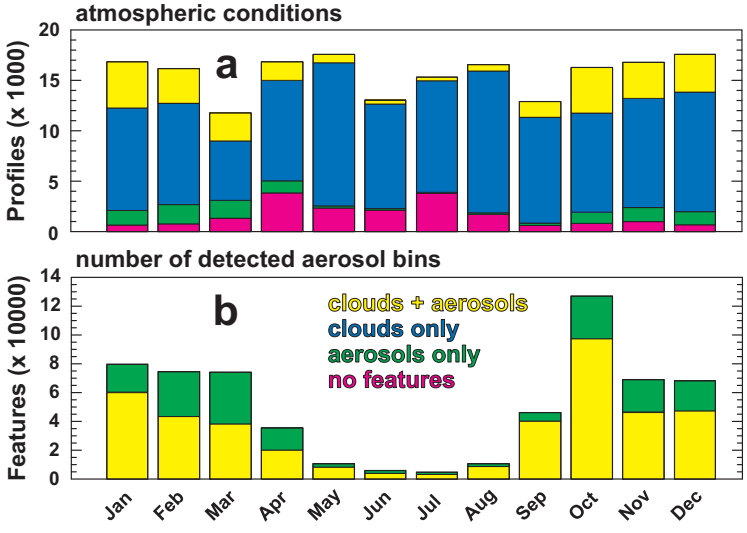

Fig. 2. Histograms of the monthly abundance of (a) CALIOP level 2 5-km aerosol profiles and (b) 60-m height-bins with aerosol observations as detected during 2018 CALIPSO overpasses in the region of interest during 2008. The color coding ~~in (a)~~ refers to the observed occurrence of atmospheric features (aerosols and/or clouds). ~~The number of detected aerosol-containing height bins per month in (b) is subdivided according to whether clouds were absent (green) or present (red) in individual lidar profiles marked as aerosol only or clouds and aerosols in (a).~~

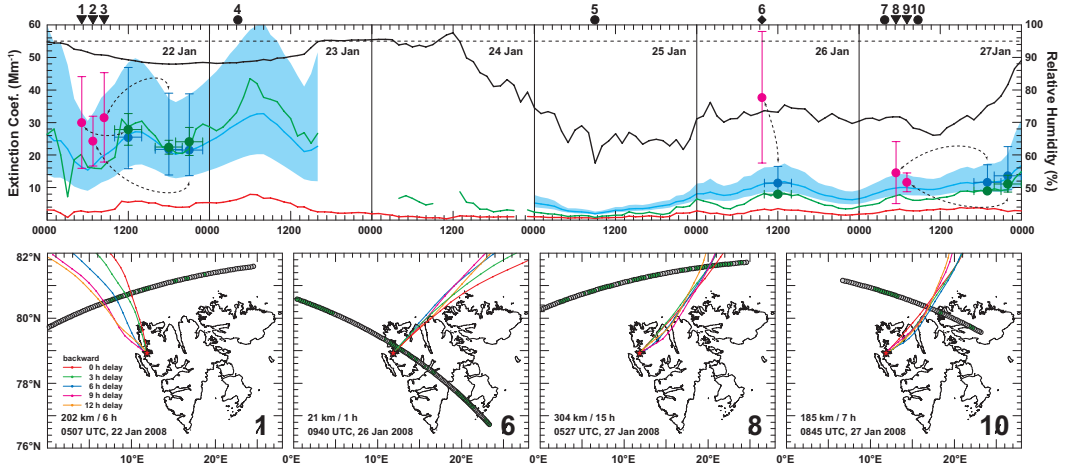

Fig. 3. Upper panel: CALIPSO extinction coefficient (532 nm, ~~green~~ **magenta** circles) compared to in-situ measurements of the dry scattering coefficient (550 nm, red line~~ and dots~~) and the ambient extinction coefficient **determined from the measurements of DMPS** (550 nm, green line ~~and dots~~) **and nephelometer plus PSAP (550 nm, blue line)** for the time period from 22 to 27 January 2008. **The blue shaded area marks the region of possible values based on the minimum and maximum estimates of the γ value. Green and blue** ~~Red~~ circles mark 5-h averages of the ambient extinction coefficient**s from the in-situ observations**. Arrows show which values are compared. Ambient RH is given **in black** ~~as blue line and squares~~. Values above RH$>95\%$ were disregarded (dashed ~~blue~~ **black** line). ~~Colored triangles~~ **Symbols** and corresponding numbers mark CALIPSO overpasses that could be connected to the ground site for the considered time period~~. The color refers to the availability of aerosol information in CALIOP profiles in the track segment chosen for comparison~~: only aerosol features (~~green~~ **triangles**), aerosol and cloud features (~~blue~~ **diamond**), and no or only cloud features (~~red~~ **circles**). Lower panel: presentation of the use of trajectories to connect the in-situ site to the spaceborne measurements for four selected cases (marked as 1, 6, 8, and 10 in the upper plot). The CALIPSO ground track is marked by gray (no aerosol data available) and green (aerosol data available) circles that refer to individual 5-km aerosol profiles. Colored dots and lines mark backward trajectories starting close to the CALIPSO overpass (red) as well as 3 h (green), 6 h (blue), 9 h (magenta), and 12 h (orange) after the overpass. The time of overpass is given in the respective plots. The red star marks the location of the Zeppelin station.

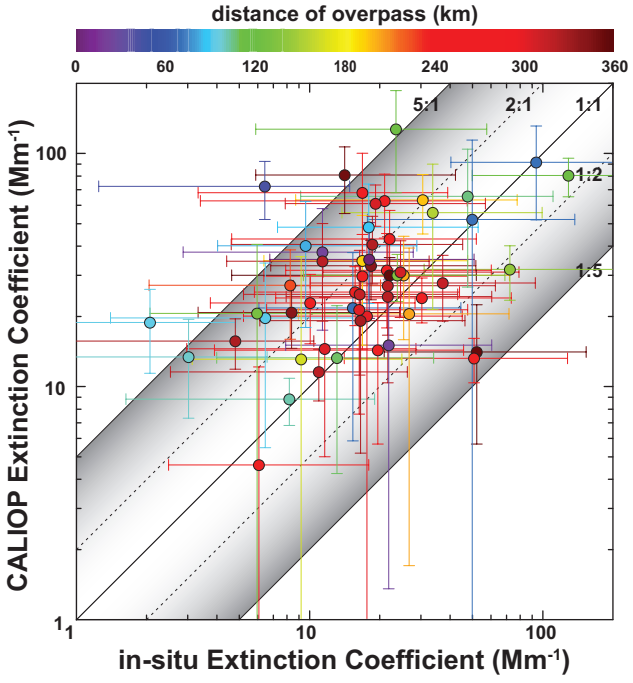

Fig. 4. Comparison of the ~~humidified~~ **ambient** 550-nm extinction coefficient from **humidification of nephelometer and PSAP** ~~the in-situ~~ measurements (see Sect. 2.2.2) versus the ambient 532-nm extinction coefficient extracted from CALIPSO overpasses for 57 suitable cases. The color coding describes the distance of the CALIPSO observation from the ground site. Error bars **refer to the results of using the lower and upper estimate in the γ value for humidification** ~~represent the standard deviation from averaging over five values of hourly humidified in-situ measurements~~ and **the standard deviation from averaging over** nine 60-m CALIPSO height bins between 250 and 730 m, respectively. Ratios of 1:1, 1:2, and 1:5 are marked by solid and dashed lines and the shaded area.

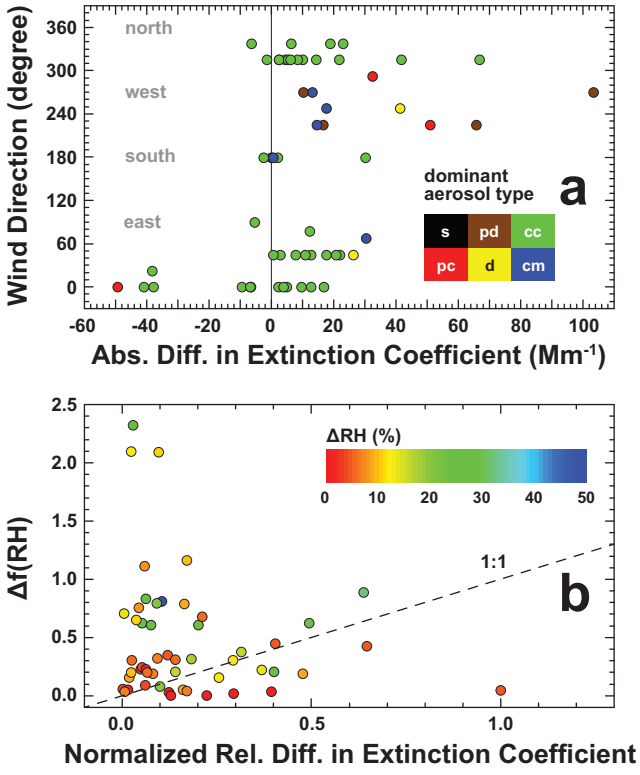

Fig. 5. Detailed view of (a) the effect of wind direction on the absolute difference in the ambient extinction coefficients derived from observations at Zeppelin and by CALIOP and (b) the connection between the relative difference $\Delta f(\mathrm{RH})$ of the scattering enhancement factors at Zeppelin station and at the intersection of trajectories and CALIPSO ground track and the relative difference in the ambient extinction coefficients observed at the two locations. The color coding refers to the dominant aerosol type identified in the CALIOP observations (cm - clean marine; d - dust; pc - polluted continental; cc - clean continental; pd - polluted dust; s - smoke, not observed) and the difference of RH observed at Zeppelin station and taken from the trajectory calculations at the location of the CALIPSO overpass, respectively. ~~Relative values are normalized to the observation at Zeppelin station.~~ The dashed line marks the 1:1 line.

Table 1. Results of the comparison of CALIPSO observations and in-situ measurements at Zeppelin station **(ambient 1 and 2 as in Fig. 1)** subdivided according to months of the year 2008 and to cloud-free and cloudy conditions in the CALIPSO aerosol profiles. The first line (columns 3-7) refers to mean values and standard deviation, while the second line refers to median and range of values.

month	**number**	**distance**	**delay**	**Extinction Coefficient (Mm^{-1})**		
	of cases	**(km)**	**(h)**	**ambient 1**	**ambient 2**	**CALIPSO**
January	14	223±112	8.9±4.5	15.6±7.5	15.5±10.2	27.4±10.5
		271, 21 – 343	9.5, 1.0 – 15.0	17.6, 1.2 – 27.8	17.0, 2.1 – 25.4	26.9, 11.5 – 48.2
February	11	251±110	7.8±4.3	12.0±4.4	19.2±12.6	23.5±15.9
		288, 2 – 357	9.0, 0.0 – 15.0	12.8, 3.0 – 18.0	16.6, 4.8 – 52.2	19.1, 4.6 – 63.2
March	10	223±111	9.8±4.1	21.2±12.1	30.1±22.3	42.1±21.2
		252, 44 – 360	9.0, 3.0 – 15.0	17.4, 4.0 – 43.01	26.3, 6.4 – 51.0	36.5, 13.2 – 72.3
April	10	216±104	8.5±3.6	35.9±27.5	34.8±20.1	59.9±32.6
		203, 69 – 352	9.0, 3.0 – 13.0	27.7, 13.8 – 95.9	23.7, 14.2 – 94.0	58.3, 27.1 – 127.0
October	2	292±40	7.5±0.7	6.0±1.7	10.7±6.6	28.6±8.3
		292, 263 – 320	7.5, 7.0 – 8.0	6.0, 4.8 – 7.2	10.7, 10.1 – 11.4	28.6, 22.8 – 34.5
November	8	128±66	5.9±3.2	16.3±24.6	26.7±17.3	24.8±23.2
		107, 23 – 226	5.5, 1.0 – 12.0	7.1, 1.6 – 75.7	8.8, 3.0 – 130.0	17.4, 8.8 – 80.3
December	2	106±33	6.5±3.5	4.2±1.6	7.8±4.7	30.3±13.8
		106, 82 – 129	6.5, 4.0 – 9.0	4.2, 3.1 – 5.3	7.8, 6.0 – 9.6	30.3, 20.5 – 40.1
all year	57	212±107	8.2±4.0	18.8±17.8	23.3±15.4	34.7±23.7
		242, 2 – 360	8.0, 0.0 – 15.0	14.3, 1.2 – 95.9	20.7, 2.1 – 130.0	27.8, 4.6 – 127.0
cloudy	18	178±116	6.2±3.9	23.0±25.2	28.6±17.9	35.4±20.9
		169, 2 – 323	5.5, 0.0 – 13.0	14.7, 1.2 – 95.9	17.9, 2.1 – 130.0	30.5, 14.3 – 91.6
cloudfree	39	228±100	9.2±3.7	16.9±13.0	20.8±14.2	34.4±25.2
		247, 23 – 360	9.0, 1.0 – 15.0	14.3, 1.6 – 74.9	18.0, 3.0 – 72.6	27.2, 4.6 – 127.0