Responses to Referee #2's Comments

Referee's comment: *The manuscript aims at applying a classification scheme to discriminate aerosol types over the Anmyon site using column-integrated optical properties derived from AERONET observations. The same scheme is also applied to other well-characterized AERONET sites. This reviewer considers that the paper is not adequately exposed and that it may require substantial revision before it can be accepted for publication in ACP. General and specific comments follow here.*

Response: We thank the referee for carefully reviewing the manuscript and providing valuable comments. Since your comments are serious and significant, we have carefully prepared the following responses.

To begin with, we'd like to mention that we will modify the title as "Identification of column-integrated dominant aerosols <u>under high aerosol optical depth conditions using the data set from a single AERONET site</u>". We will add two phrases, "under high aerosol optical depth conditions" and "using the data set from a single AERONET site". The first phrase is to indicate that we identified column-integrated dominant aerosols when AOD $\geq$ 0.4. We provided the total occurrence rate, including the low AOD conditions (AOD < 0.4) to compare it with the occurrence rates of dominant aerosols. However, there seems to be confusion, and thus we decided to clarify in the title that we used the data of AOD $\geq$ 0.4 for dominant aerosols.

The situation is similar for the second phrase. One of our main objectives is to demonstrate that we can obtain good information on dominant aerosols using the AERONET data set just from a single site. As in other works, we also analyzed worldwide AERONET sites which have distinct source characteristics, but our purpose was to evaluate the validity of the results from the Anmyon site.

The following are our responses to your comments. Your comments are shown in italics as seen above, and specific comments as well as general comments were numbered as appropriate.

<u>General comments</u>

1a. *The classification method is not sufficiently described, in particular the steps regarding the cluster analysis. The authors claim that the proper number of clusters is decided in the clustering process (page 26633, line 25). However the aerosol types (MD, OC, BC, etc.) were established in advance. I think this crucial point is not clear at all.*

As you mentioned, the numbers of clusters were decided in the clustering processes. However, the aerosol types were not established in advance. Instead, they were determined by comparing their properties with those from previous works. We agree that our description on the clustering processes (P26633, L21-28) was insufficient, and thus it will be extended as follows:

"(5) Types of absorbing aerosols in fine and coarse modes were distinguished, respectively, using the K-means clustering method with parameters of AOD_{440}, SSA_{440}, $AAE_{440-1020}$, and FMVF (Table 1). Note that AOD was used for classifying aerosol types although it depends on aerosol amount rather than type. This is different from other works in Table 1 except Omar et al. (2005) who used coarse and fine mode volume concentrations. We used AOD for a cluster analysis since we obtained the most plausible results with AOD, which means that aerosol types are not completely separable with their amounts.

K-means clustering is a method to partition a data set into the prescribed number, *K* groups (Jain, 2010). To determine the best *K*, we coupled the K-means clustering with discriminant analysis which classifies the data into the given clusters (Romesburg, 2004; Aczel and Sounderpandian, 2009). We performed the K-means clustering with different values of *K* and chose the best *K* by examining the hit ratio, which is a measure of how correctly the discriminant analysis classifies the data set into the same groups given by the cluster analysis. When the hit ratio is 100%, the classification by the discriminant analysis completely coincides with that from the clustering method.

In this work, the commercial software code SPSS (Version 12.0; http://www.spss.com) was used for clustering and discriminant analysis. For the coarse mode, the hit ratios were 98.4%, 99.5%, 95.9% and 96.4% when *K* was varied from 2 to 5. As a result, we selected *K* = 3 and the clusters were designated as MD, MD mixed with carbon, and mixed coarse particles by comparing their properties with those from the previous works, which will be discussed later. For the fine mode, the hit ratios were 100%, 98.9%, and 97.9% when *K* was varied from 2 to 4. We selected *K* = 2 and the clusters were designated as BC and OC. The discriminant analysis was also used to classify dominant aerosols at worldwide AERONET sites into the aerosol types obtained from the Anmyon site."

Aczel. A. D., Sounderpandian, J.: Complete Business Statistics, 7/e, McGraw-Hill, ISBN: 0073373605, 2009.

Jain, A.: Data clustering: 50 years beyond K-means, Pattern Recognition Letters, 31, 651-666, 2010.

1b. *Are the values of AOD, FMF, SSA and AAE in Table 2 used for clustering or are they the average values for each cluster?*

As mentioned in the response to your comment 1a, we used AOD_{440}, SSA_{440}, $AAE_{440\text{-}1020}$, and FMVF for clustering aerosol types. Please note that we changed FMF with FMVF (fine mode volume fraction) since FMF has been generally used to indicate the find mode fraction of AOD.

The values in Table 2 are means and standard deviations of each cluster as was given in the caption of Table 2

2. *The need of AOD(440nm)>0.4 for level 2.0 optical properties is a strong limitation of the method. In practice the method excludes a huge portion of observations in all investigated sites. "Low AOD" does not mean that no classification can be attempted, even though it cannot be as specific as in the higher AOD cases. For instance, low AOD and low Angstrom exponent may easily lead to marine aerosol type identification (Smirnov et al., 2002). A simple scheme based on AOD, Angstrom exponent and fine mode fraction of the AOD (see climatology by Holben et al., 2001), or Angstrom exponent derivatives (Gobbi et al.) can give good insight on the aerosol type and is clearly missing in this work.*

As you also stated, we did not attempt to distinguish the low AOD cases because *it cannot be as specific as in the higher AOD cases.* If our objective is to obtain an insight or just 'good' information on the aerosol types, the approaches adopted by the works you mentioned will be sufficient. However, we'd like to have more specific information on the aerosol types such as mineral dust, BC, OC, and secondary ions, which has been available from intricate physical and chemical measurements.

The climatology by Holben et al. (2001) is both comprehensive and complete but represents the best that can be obtained from ground-based optical measurements. We have also noted

Gobbi et al. (2007) and Basart et al. (2009) mainly because they use a rich number of direct-sun data compared with diffuse radiation data, not because their approach is more rigorous than other works. You also stated that low AOD and low AE may easily lead to marine aerosol type information. What if the target area is far inland of the Amazon forest or the Chinese desert? We think this statement is true only where the marine influence is important.

We completely agree that AOD ≥ 0.4 is a strong limitation. However, most previous works, including those in Table 1, use SSA or other inversion products which are only available when AOD ≥ 0.4. In our works, we'd like to draw attention to the fact that optical properties widely used to study the physical and chemical characteristics of aerosols during the past decade are based on very limited conditions of AOD ≥ 0.4.

Basart, S., Pérez, C., Cuevas, E., Baldasano, J. M., and Gobbi, G. P.: Aerosol characterization in Northern Africa, Northeastern Atlantic, Mediterranean Basin and Middle East from direct-sun AERONET observations, Atmos. Chem. Phys., 9, 8265-8282, 10.5194/acp-9-8265-2009, 2009.

Gobbi, G. P., Kaufman, Y. J., Koren, I., and Eck, T. F.: Classification of aerosol properties derived from AERONET direct sun data, Atmos. Chem. Phys., 7, 453-458, 10.5194/acp-7-453-2007, 2007.

3. *The results section looks at average optical properties of the investigated types. This is a wrong approach because the properties are conditioned by the classification scheme. The authors know it (p. 26640, line 22) but still used the ill-posed approach throughout the paper. In my view, the correct approach is investigating the presence rate of each aerosol type in the investigated sites (as it is also done by the authors).*

In our view, we always look at average optical properties of the investigated types. Differences are what the investigated sites are. Most previous works distinguished aerosol types by analyzing the data set from distinctive (worldwide) sites. On the other hand, we did that by analyzing the data set from a single site and tried to demonstrate our results are not much different from those of previous works.

In order to clarify our objective we will add a phrase 'from a single AERONET site' to the title as mentioned at the beginning. We presume that you already understand what we did because you said in the comment 15 that you respect our choice.

4. *Section 3.1 is a sort of continuation of the methodology section, with references used to support the classification scheme and very few results. I suggest that the few results are merged with section 3.2 and that all methodology and references are merged with section 2.*

Section 3.1 is prepared to demonstrate that our results are in similar ranges to those of the previous works despite using the data set from a single site. However, the first paragraph (P26634, L3-8) will be moved to P26633, L7 in section 2.

5a. *Two very confusing concepts are used in the paper. First: "the dominant aerosols" is used to define situations with AOD(440nm)>0.4, in which the classification scheme can be applied to discriminate the predominant aerosol type. This must be reformulated.*

Dominant aerosols in our work indicate that those types of aerosols were dominant at a certain event regardless of how frequent that kind of event occurred.

5b. *Second: the way used to calculate the "occurrence rate of dominant aerosols", based on number of sun hours and number of level 2.0 almucantars (if I understood correctly) leads to strange interpretation of the data coverage, of just few percents. Given the typical temporal* *variability of the atmospheric aerosol, I would rather analyze in terms of days instead of hours.*

We recalculated the occurrence rate by introducing the concept of the instrument working hours as follows:

"The occurrence rate was calculated by dividing the occurrence number of the aerosol type by the total number of raw data for diffuse radiation. Here, raw data for diffuse radiation indicate the almucantar raw data. We used those for SZA between 50° and 80° from which inversion products were obtained (García et al., 2008). This means that the occurrence rate represents how often dominant aerosols occur during the instrument working hours when the instrument measures diffuse radiation that can be used for inversion products."

García, O. E., Díaz, A. M., Expósito, F. J., Díaz, J. P., Dubovik, O., Dubuisson, P., Roger, J. C., Eck, T. F., Sinyuk, A., Derimian, Y., Dutton, E. G., Schafer, J. S., Holben, B. N., and García, C. A.: Validation of AERONET estimates of atmospheric solar fluxes and aerosol radiative forcing by ground-based broadband measurements, Journal of Geophysical Research: Atmospheres, 113, D21207, 10.1029/2008JD010211, 2008.

The above description will substitute that from P26636, L25 to P26637, L5. Because of this change the occurrence rates (%) at the study sites will increase as follows:

	Original Works		Revised Works	
	Total	Dominant Aerosols	Total	Dominant Aerosols
Anmyon	4.7	1.2	21.1	5.4
Beijing	10.6	6.3	25.7	15.2
Mexico City	2.9	1.1	8.5	3.1
GSFC	11.6	0.9	25.7	2.0
Mongu	8.7	2.2	23.1	5.8
Alta Floresta	2.8	0.8	10.6	2.9
Cape Verde	6.0	2.2	14.4	5.3

<u>Specific comments</u>

6. *P26632, L5: improve definition of single scattering albedo.*

The definition will be added at the end of the sentence: "SSA is an indicator of aerosol scattering, <u>defined by the ratio of scattering to total extinction</u>."

7. *P26632, L19: "inflection point"*

The term "stationary point" denotes a minimum between fine and coarse modes in the bimodal distribution. It will be changed to "local minimum" as follows to make it more understandable:

"FMVF is the fraction of the volume concentration below the <u>local minimum</u> between fine and coarse modes, the point being in the range of 0.44–0.99 μm (Dubovik et al., 2002; Schafer et al., 2008; Prats et al., 2011)."

8. *P26633, L10: "absorbing aerosols" instead of "absorption aerosol".*

All "absorption aerosols" will be changed to "absorbing aerosols".

9. *P26633, L26: explain SPSS 12.0 and give citation to Romesburg, 2004 in line 22.*

We are not sure what you mean by "give citation to Romesburg, 2004" because Romseburg

(2004) is a citation. You can refer to the response to comment 1a for the whole related to SPSS 12.0 and the citation of Romesburg (2004), but the relevant parts will be changed as follows:

"To determine the best *K*, we coupled the K-means clustering with discriminant analysis which classifies unknown samples into known clusters (Romesburg, 2004; Aczel and Sounderpandian, 2009)" and

"In this work, the commercial software code SPSS (Version 12.0; http://www.spss.com) was used for clustering and discriminant analysis."

10. *P26634, L3: provide number of days instead of number of measurements (similarly to general comment 5).*

We counted the number of individual measurement data and recalculated the occurrence rate on the basis of almucantar raw data in the revised work as mentioned in the response to comment 5b. We think this approach is clearer than a daily-based one because the number of measurements in a day is highly variable.

11. *P26634, L19: specify wavelength for SSA, at least the first time.*

We will specify the wavelengths for AOD, SSA, and AAE when they first appear in the methods section (section 2) to be AOD_{440}, SSA_{440}, $AAE_{440\text{-}1020}$.

12. *P26635, L1: if a mixture of dust and pollutants was measured, you cannot say "dust AAE measured at Gosan was. . ." because they did not measure dust alone. This is repeated several times in the text. The authors can try to be more precise.*

We agree that it is best to measure dust alone. However, in reality, it is difficult to separately measure "pure dust" because it is easily mixed with pollutants even from source regions (Kim et al., 2011). This was first reported by analyzing AERONET data because AERONET is installed at remote sites and has operated continuously. For example, Eck et al. (2005) found that AE at Dalazadgad in the Gobi desert was high in fall and winter due to coal combustion for cooking and heating and also due to inflow of pollutants from gas flaring in Siberia to the north and northwest. Xu et al. (2004) measured the chemical, physical, and optical properties of aerosols at Yulin in the north of Loess Plateau during ACE-Asia. They indicated that scattering and absorption coefficients were elevated in the early morning. At that time, dominant composition of $PM_{2.5}$ was organic matter (41%) followed by crustal material (29%), sulfate (17%), and EC (13%).

In the manuscript we mentioned that MD in Table 2 is close to pure dust. Please note that we can discriminate this from 350 measurement data, not on daily basis (see the response to comment 10). The paragraph between P26634, L20 and P26635, L2 explains how MD in Table 2 is close to pure dust in comparison with others which are mixed with pollutants.

Xu, J., Bergin, M. H., Greenwald, R., Schauer, J. J., Shafer, M. M., Jaffrezo, J. L., and Aymoz, G.: Aerosol chemical, physical, and radiative characteristics near a desert source region of northwest China during ACE-Asia, J. Geophys. Res., 109, D19S03, doi:10.1029/2003JD004239, 2004.

13. *P26636, L8: water vapor is not derived from inversions, it is retrieved from the direct Sun observations.*

Thank you for your suggestion. We will correct the sentence to state "Both water vapor and effective radius were obtained with other inversion products, although water vapor is retrieved from direct Sun measurements."

14. *P26636, L9, In Fig. 3 there is apparently some dependency of fine mode effective radius with respect to water vapor. However quite constant fine mode effective radius was observed by Gonzi et al. (2002) for many AERONET sites. Are you sure that the observed dependency is due to hygroscopic growth?*

We checked the volume median radius for the fine mode for Anmyon. Mean ± standard deviation were 0.191 ± 0.044, and both are higher than the values given in Table 10 of Gonzi et al. (2002), 0.14-0.18 and 0.01-0.03, respectively. However, their values were calculated from monthly means while our values were from individual measurement data.

In addition, we will replace Fig. 3 with the following plots:

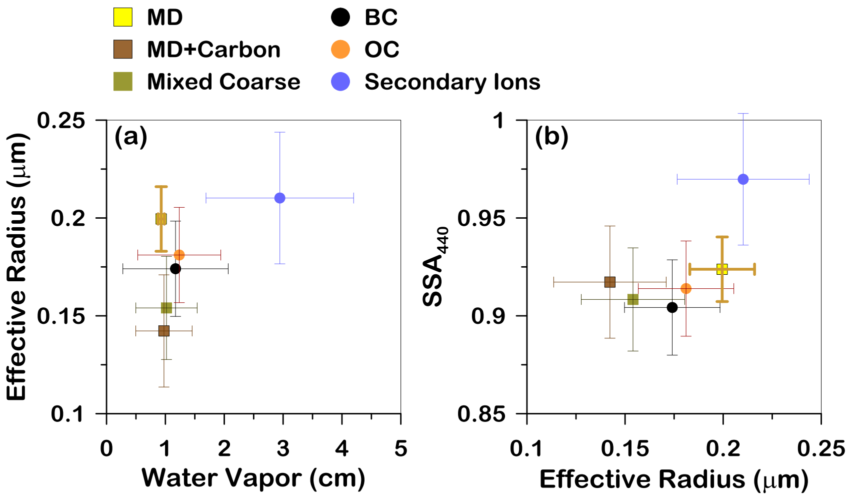

Fig. 3. Plots of (a) fine-mode effective radius vs. column water vapor and (b) SSA_{440} vs. fine-mode effective radius. Symbols and error bars represent means and standard deviations, respectively.

Fig. 3a more clearly shows that the effective radius of secondary ions is larger in association with higher water vapor. We can reasonably interpret that this was due to hygroscopic growth of secondary ions.

Gonzi, S., Baumgartner, D., and Putz, E.: Aerosol Climatology and optical properties of key aerosol types observed in Europe, IGAM/UG Technical Report for EU Np.1/2002, 2002.

15. *P26638, L12: I would find more logical applying the classification scheme first to well characterized AERONET sites; and then to your Anmyon site. However that is just a point of view and I must respect the authors' choice.*

Please refer to the response at the beginning that states we will modify the title and the response to comment 3.

16. *P26642, L5-11: these last sentences are too ambiguous and unnecessary.*

We will revise the last sentences as follows: "One probable reason is that dominant aerosols, which have been physically and chemically characterized, are distinguished using the dataset of properties from optical measurements. However, it could be also because dominant aerosols are column-integrated and because mean properties from a number of occurrences are compared. Close investigation of the meaning and effectiveness of identifying column-integrated dominant aerosols using the dataset of properties from optical measurements are warranted."

With this revision, I hope you can agree why we added these at the end of the manuscript.

17. *P26642, L12: the necessary acknowledgment to AERONET program and station PI's is*

missing.

Thank you for your suggestion. We will add the following to the acknowledgment: "We are grateful to the following principal investigators for establishing and maintaining AERONET sites, H.-B. Chen and P. Goloub of Beijing, D. Tanré of Cape Verde, and A. L. Contreras of Mexico City."

18a. *P26650, Table 2: such high AOD for dust is hard to believe as representative or average value.*

We'd like to point out that MD in Table 2 is identified for dominant aerosols from individual measurement data. This could be different from representative or average values that have been reported. For clarity, the following will be added on P26634, L16:

"However, it is noteworthy that the properties of MD in Table 2, close to pure dust, are partially because MDs were identified from individual measurement data while they have been mostly from mean values for a day or the period (Dubovik et al., 2002; Wang et al., 2004; Eck et al., 2008)."

Eck, T. F., Holben, B. N., Reid, J. S., Sinyuk, A., Dubovik, O., Smirnov, A., Giles, D., O'Neill, N. T., Tsay, S.-C., Ji, Q., Mandoos, A. A., Khan, M. R., Reid, E. A., Schafer, J. S., Sorokine, M., Newcomb, W., and Slutsker, I.: Spatial and temporal variability of column-integrated aerosol optical properties in the southern Arabian Gulf and United Arab Emirates in summer, J. Geophys. Res., 113, D01204, doi:10.1029/2007JD008944, 2008.

Wang, J., Xia, X., Wang, P., and Christopher, S. A.: Diurnal variability of dust aerosol optical thickness and Angström exponent over dust source regions in China, Geophys. Res. Lett., 31, L08107, doi:10.1029/2004GL019580, 2004.

18b. *Furthermore, an extensive property like AOD is not the best choice for type identification.*

We tried the cluster analysis without AOD but could not obtain a reasonable result. As can be seen the response to comment 1a, we will explain it in the revised manuscript as follows:

"Note that AOD was used for classifying aerosol types although it depends on aerosol amount rather than type. This is different from other works in Table 1 except Omar et al. (2005) who used coarse and fine mode volume concentrations. We used AOD for a cluster analysis since we obtained the most plausible results with AOD, which means that aerosol types are not completely separable with their amounts."

19. *P26652, Fig. 2: the second filter regarding imaginary part of the refractive index should take into consideration the estimated uncertainty for this parameter. If the value at 440nm is lower than the value at longer wavelengths but still within uncertainty, I consider that they should not be removed.*

It is true that the estimated uncertainty of imaginary part of the refractive index is quite large as indicated by Dubovik et al. (2000, 2002). However, we used this filter as was done in Arola et al. (2011) and Bergstrom et al. (2007) because the classification results were better with this filter.

20. *P26653, Fig. 3. This kind of relationships can be of great interest but the authors do not investigate or analyze them in depth. I encourage them to analyze it further and find other ones that can be very illustrative to the aerosol type interpretation.*

We prepared this figure to explain both high SSA of secondary ions and high SSA in summer.

We are able to confirm this more clearly with a new figure shown in the response to comment 14. However, if possible, we’d like to analyze these phenomena further in our next study.