Author's response including:
- **Response to Anonymous Referee #1**
- **Response to Anonymous Referee #3**
- **Revised manuscript with tracked changes**

Response to Anonymous Referee #1

Thank you, anonymous referee #1, for your thorough and constructive comments. The manuscript is much improved after your input.

Our response is structured as follows: original comments from reviewer #1 are bolded, our responses are in italics, and the revised portions of the manuscript follow in quotation marks with specific changes/additions in red.

General Comments

The paper addresses the important topic of aerosol type classification through aerosol spectral optical properties. Authors explore how existing schemes reported in the literature may be applied and possibly improved. The paper demonstrates that there still are several uncertainties in all classification schemes proposed here, and discusses major ambiguities and limitations. No significant improvement is, however, proposed. The topic is fully within the scope of ACP, and I recommend this paper to be published in ACP after the following issues are addressed.

Specific Comments

1. **The meaning of AAE as indicator of aerosol properties still contains uncertainties that have not been formally assessed in the scientific community.**

   i. **AAE = 1 as indicator of BC. Although AAE = 1 has often been attributed to BC, several studies have demonstrated that AAE depends on both aerosol composition (and in particular, to the BC-to-OA ratio more than to the BC mass concentration alone), and size (e.g., Lack and Langridge, 2013; Saleh et al., 2014; Costabile et al., 2017). Caution should therefore be taken in interpreting AAE=1 as indicator of BC only. As an example, if AAE = 1 indicates BC only, cluster 1 will indicate BC (pag. 9 lines 35-38); however, authors acknowledge that "the generally low scattering coefficients at these sites indicate the presence of anthropogenic aerosols is infrequent" (pag. 10, lines 6-8).**

   Thank you for the comment. It is certainly important to mention the uncertainties associated with the AAE =1 indicating black carbon, and an expanded discussion of AAE uncertainties as related to BC has been added to the manuscript, on Pag. 2 and in the discussion sect (new Sect. 7 – see author response to reviewer comment 2(i) below).

   Pag. 2, line 34-Pag. 3, line 2: "Black carbon (BC), for example, has a theoretical AAE value around 1, while dust aerosol typically has AAE values greater than 2 (Bergstrom et al., 2002, 2007; Kirchstetter et al., 2004), though AAE of ambient aerosol is likely to evolve with atmospheric processing and will depend strongly on composition (e.g., black carbon-to-organic aerosol ratio), coating and size (Saleh et al., 2014; Costabile et al., 2017; Moosmüller et al., 2011)."

   ii. **AAE > 2 as indicator of brown carbon. Although several studies have indicated that AAE values from 2 to 6 may indicate brown carbon (e.g., Andreae and Gelencsér, 2006; Moosmüller et al., 2011), authors do not consider brown carbon at all in this paper. In fact, the contribution of BrC aerosol with large AAE values (2-6) may be less significant at the**

1

regional remote/rural sites analysed here, as discussed at pag. 18, line 17. Therefore, the paper should mention that the proposed analysis may fail at urban sites and more polluted sites. Indeed, authors acknowledge (pag. 8, line 28) that Cazorla et al.'s scheme may fail for sites in proximity to aerosol sources.

We agree that brown carbon should not be neglected, and did not intend to do so. The authors implicitly assumed that brown carbon was encompassed by the 'OC' designation from the Cazorla et al. (2013) matrix (Cappa et al., 2016 also made this assumption). However, we realize we should specify that BrC is the light-absorbing portion of OC. We have changed the aerosol classification matrix to that of Cappa et al. (2016), which includes BrC (in other words, light-absorbing OC). This change is reflected in Fig. 2(b) and in the text of the results section.

Nonetheless, none of the sites in this analysis show AAE values greater than 2, so this does not affect the results of the analysis. CPR does have a median AAE value of 2, but the supporting evidence is strong that this is likely dust from long range transport, mixed with BrC (see Pag. 11, lines 11-20, and references Prospero et al., 2014; Denjean et al., 2016; Kalashnikova and Kahn, 2008; and Reid et al., 2003). There is a possibility that BrC is not identified by AAE values greater than 2 due to limitations from the instruments measuring aerosol light absorption. Since the AAE values derived from PSAP and CLAP measurements for this analysis are for the 450/700nm wavelength pair, and since BrC absorbs most strongly in UV (which is not measured by the PSAP and CLAP), it is possible that this method would not positively identify BrC even if it was present at the measurement sites. There is no way for us to confirm this given the data that are available to us, but it is a potential limitation of the data that is worth mentioning.

It is not clear why reviewer #1 suggests that the proposed analysis may fail at urban sites and more polluted sites. In fact, there are many urban and/or polluted sites that are included in this analysis for which the classification scheme works well (e.g., GSN, AMY, FKB).
Even polluted sites dominated by BrC would be well classified with the existing schemes, as they would specifically identify OC/BrC when AAE > 2, assuming the potential instrumental limitations mentioned above are not an issue.

iii. AAE of carbonaceous aerosols. The paper mentions that carbonaceous aerosols have AAE ~ 1.2 (e.g., pag. 10, line 10). However, carbonaceous aerosols include both BC, OA, and BrC, and their relative contribution in particle with different size can cause a large variability in AAE values.

More careful use of the term 'carbonaceous aerosol' has been used throughout, and the term has been replaced by more specific reference to BC, OC or BrC as needed. The authors agree that in the example on previous version pag. 10, line 10, the term 'carbonaceous aerosol' was misused in conjunction with the designation of AAE~1.2. This has now been changed, see below.

Now in supplemental materials, page 2: "Cluster 2 includes AMY and GSN, the two coastal stations located in South Korea, and is characterized by high aerosol loadings (high σ_{sp}), small aerosol particles (SAE~1.5) and BC dominated aerosols (AAE~1.2)."

2. Overall presentation and structure
Results section (Sect. 5) is too long and a little bit confusing. In particular:
    i. Cluster analysis results (pag. 9, 10, 11) are somehow ambiguous, as authors acknowledge (pag. 11, lines 20-24)

2

Discussion of individual cluster results from the multivariate cluster analysis has been moved from the manuscript to supplemental materials in order to shorten the results section and highlight the important takeaways from the analysis. Although the results of the cluster analysis are indeed somewhat ambiguous, this is one of the important results from this aerosol classification technique. Namely, a cluster analysis does not eliminate uncertainties in aerosol classification schemes, though it does help give a more complete picture of aerosol conditions at the measurement sites. This is reiterated in the discussion section.

Pag. 17, lines 4-8: "An anticipated advantage to the multivariate cluster analysis was that it would help to reduce ambiguity in results of aerosol typing schemes, though this was not the case with every cluster. Rather than falling more surely within the optical property thresholds of one aerosol type, the median optical properties of a few clusters still fell on the cusp of two or more aerosol type thresholds. This left the aerosol type of some clusters uncertain, particularly for clusters with coastal and/or remote sites."

ii. Description of multivariate cluster analysis (pag. 9, lines 1-25) should be moved to a separate paragraph (e.g., in a data analysis section).

A new data analysis section has been added (Sect. 5), and description of data analysis methods for each of the classification techniques has been added and/or moved to that section to help shorten the results section (now Sect. 6). See Sect. 5 in newest version of manuscript.

Discussion section (Sect. 6) is the most interesting part of the paper, but it should be better clarified in some parts. In particular, I recommend to address:
i. Uncertainties in the AAE attribution method mentioned above

A paragraph expanding on uncertainties in the AAE attribution method (particularly for BC) has been added to the discussion section (now Sect. 7).

Pag. 17, line 32- Pag. 18, line 2: "It should be mentioned that the success of aerosol classification schemes is largely dependent on uncertainties in AAE attribution. The scientific community has yet to fully assess AAE as an indicator of aerosol composition. Although AAE=1 is often taken to indicate black carbon, some studies show that this largely depends on aerosol composition and size, as well as the age of the particle and atmospheric processing that it endures (Lack and Langridge, 2013; Saleh et al., 2014; Costabile et al., 2017; Moosmüller et al., 2011). Furthermore, the accuracy of these aerosol classification methods are only as accurate as the AAE value is an indication of the aerosol composition. As the scientific community advances our understanding of AAE and its relationship to aerosol composition and size, these aerosol classification schemes should be refined."

ii. Differences between columnar and in situ surface spectral optical properties. Authors apply thresholds from Cazorla et al.'s scheme – obtained by columnar measurements (AERONET) – to interpret in situ surface data. Although results may be consistent, in principle that is not completely correct. In fact, aerosol spectral optical properties measured by columnar and in-situ ground instruments may in principle differ significantly. Relevant thresholds seem, however, to be consistent in Tab. 1. Please, add a more comprehensive discussion (this is only sketched in the discussion section, pag. 18, lines 1-10).

It is true that aerosol spectral optical properties measured by columnar and in-situ ground instruments may in principle differ. Since the Cazorla et al. (2013) aerosol classification matrix

3

was based on ambient AERONET measurements, it could include confounding effects of water uptake by aerosols (depending on the aerosol hygroscopicity). Without knowing the ambient RH of the environment during the Cazorla et al. (2013) measurements, we have no way of knowing how the matrix would differ if the data was unaffected by potential water uptake. However, the matrix is just a guiding classification scheme- not a definitive answer for aerosol type. Furthermore, the classification matrix still seems reasonable given what we know about aerosol optical properties, and it does seem to work reasonably well for our data. Moreover, Cappa et al. (2016) use in-situ data to retool the Cazorla et al. (2013) classification matrix, and though their categories are slightly more detailed, the big picture categories align very well with the Cazorla et al. (2013) classifications.

In order to address the concerns of the reviewer that we used a matrix developed with AERONET columnar measurements and applied it to in-situ data, we have changed the matrix to that of Cappa et al. (2016), which was developed based on the Cazorla et al. (2013) matrix, but modified using their results from in situ data. This does not substantially change any results from our analysis, since the matrices are very similar (see comparison of the matrices and added text describing the similarities below).

Cazorla et al. (2013) matrix is on the left side, Cappa et al. (2016) matrix is on the right.
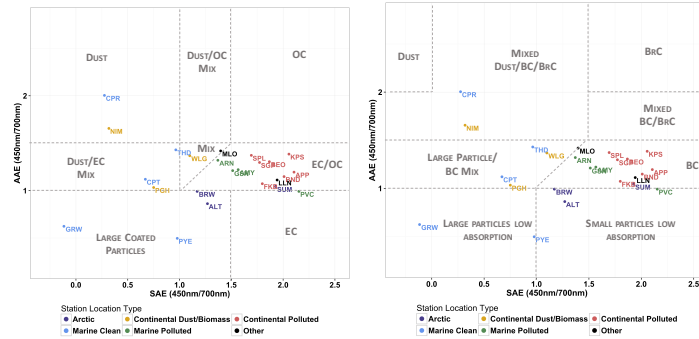
Pag. 8, lines 13-23: "It should be noted that the Cappa et al. (2016) and Cazorla et al. (2013) matrices are very similar. Both designate high SAE and high AAE values as BrC or mixed BC/BrC (though Cazorla et al. (2013) refers to BrC as OC). Both designate low SAE values and high AAE values as dust or dust mixed with BC and BrC, and both suggest an AAE value around 1, accompanied by higher SAE values indicates aerosol populations dominated by BC. Three main differences between the matrices can be identified. The Cappa et al. (2016) matrix makes more specific designations of aerosol mixtures (e.g., adds 'mixed dust/BC/BrC' and 'large particle/BC mix'). The Cappa et al. (2016) matrix also replaces the Cazorla et al. (2013) matrix designation of 'large coated particles' with 'large particle/low absorption mix or large black particles'. Finally, the Cappa et al. (2016) matrix replaces the Cazorla et al. (2013) matrix designation of 'EC' with 'small particle/low absorption mix'. We chose to primarily use the Cappa et al. (2016) matrix since it is based on in situ data (Cazorla et al. (2013) is based on AERONET data), and since the aerosol designations seemed to align most closely with our data. Results are presented in Sect. 6.1."

iii. **Classification of measurement site. Authors demonstrate that none of the classification schemes applied here can classify measurement sites. None of the sites has indeed only a single dominant aerosol type (pag. 12, lines 6-8). Only continental polluted sites are well classified. This is a reasonable conclusion, as these classification schemes should classify**

4

aerosol type, not measurement site. Please, discuss this point more clearly (this is roughly discussed in Sect. 6).

The classification schemes applied here are not meant to classify measurement sites. Rather, the use of measurement site classifications here is meant to help validate aerosol type inferred from the aerosol classification techniques, since chemical measurements are not available at these sites to validate aerosol types. This is an imperfect method- not only because sites may have more than one dominant aerosol (e.g., a polluted marine site that measures both anthropogenic pollution aerosols and natural sea salt aerosols), but also because sites may measure aerosols that are unexpected given the site classification due, for example, to long range transport (e.g., a clean marine site measuring pollution aerosol that was transported a long distance). However, given the remaining ambiguities, the site classifications paired with previous analyses of these sites and the aerosols they measure do provide useful information that shows when/where the aerosol classification techniques are most useful, and when/where they tend to fail.

3. **Sampling line description is not clear enough**
   i. **Are all sites but SUM equipped with PM_{10} sampling heads?**

   All measurements analyzed here are collected from 10 µm size cuts, with the exception of those from SUM, which utilizes a 2.5 µm inlet (see Pag. 7, lines 17-19).

   ii. **Heated inlets might cause losses in organic aerosol and volatile compounds. This can influence aerosol spectral optical properties. Is heating performed at all sites? Please, add more details.**

   Thank you for encouraging us to add this information to the manuscript. The following paragraph has been added:

   Pag. 6, lines 4-11: "In order to minimize aerosol hygroscopic effects, measurements at all stations (except SUM and SPL) are made at a reduced relative humidity (RH < 40%) by heating the inlet air or by diluting with filtered, dry air. The inlets at most sites are either gently heated (heating does not exceed 40°C) with a stack heater or a small heater by the impactor, and are only utilized if the relative humidity exceeds 40%. Although heating the sampling inlet can cause loss of organic and volatile aerosol material, which can alter the aerosol spectral optical properties, this is not expected to substantially impact results here. Studies that analyze the amount of volatile components removed at 40°C (by a thermal denuder) is less than 10% (Mendes et al., 2016; Huffman et al., 2009). For this particular study, we do not have the data necessary to evaluate the extent to which aerosol optical properties are affected by the heating, but evidence from other studies suggests the effect is likely small."

<u>**Technical Corrections**</u>
-Pag. 8, line 2: is

This suggested correction was not made since we intended to write 'data are' (data is a plural of datum).

-Tab. 4: check log(sigma)

Thank you for catching this. We changed $Iog(\sigma_{sp})$ to $log(\sigma_{sp})$ in Tab. 4.

-Pag. 12, line 20: "is"

5

We could not identify what needed to be changed to 'is' on Pag. 12, line 20 in the old version of the manuscript.

5 **References**
Andreae, M.O., and Gelencsér, A.: Black carbon or brown carbon? The nature of light-absorbing carbonaceous aerosol, Amtos. Chem. Phys., 6, 3131-3148, doi: 10.5194/acep-6-3131-2006, 2006.

10 Costabile, F., Gilardoni, S., Barnaba, F., Di Ianni, A., Di Liberto, L., Dionisi, D., Manigrasso, M., Paglione, M., Poluzzi, V., Rinaldi, M., Facchini, M.C., and Gobbi, G.P.: Characteristics of brown carbon in the urban Po Valley atmosphere, Atmos. Chem. Phys., 17, 313-326, doi: 10.5194/acep-17-313-2017, 2017.

Lack, D.A. and Langridge, J.M.: On the attribution of black and brown carbon light absorption using the Angstrom exponent, Atmos. Chem. Phys., 13, 10535-10543, doi: 10.5194/acp-13-10535-2013, 2013.

15

Moosmuller, H. Chakrabarty, R.K., Ehlers, K.M., and Arnott, W.P.: Absorption Angstrom coefficient, brown carbon, and aerosols: basic concepts, bulk matter, and spherical particles, Atmos. Chem. Phys., 11, 1217-1225, doi:10.5194/acp-11-1217-2011, 2011.

20 Saleh, R., Robinson, E.S., Tkacik, D.S., Ahern, A.T., Liu, S., Aiken, A.C., Sullivan, R.C., Presto, A.A., Dubey, M., Yokelson, R.J., Donahue, N.M., and Robinson, A.L.: Brownness of organics in aerosols from biomass burning linked to their black carbon content, Nat. Geosci., 7, 647-650, 2014.

Thank you for the useful references. We have incorporated these into the manuscript as needed, and have added
25 *them into the manuscript reference list.*

Response to Anonymous Referee #3

The authors thank anonymous referee #3 for their useful comments. The manuscript is better after considering
30 referee #3's suggestions.

Our response is structured as follows: original comments from reviewer #1 are bolded, the author's responses are in italics, and the revised portions of the manuscript follow in quotation marks and red letters.

35 **General Comments**
The authors present an important topic and explore a variety of aerosol classification schemes using an extensive data set with different aerosol types. This analysis allows authors to discuss about the differences between classification schemes and one of the main conclusions is that all classification schemes fail at some point, and some aerosol types are missing. This is, there is no combination of extensive and/or
40 **intensive optical properties that allow the perfect classification of aerosol types and it looks like knowing the measurement sites is needed, or helps discriminate the aerosol type. This should be included in the conclusions. Also the conclusions should explicitly indicate the specific ideas for future work needed (last paragraph on conclusion section).**

45 **The paper is within the scope of ACP and presents an interesting analysis that show the goodness and weakness of the different aerosol classification schemes presented and propose other approaches for the classification of aerosol. I recommend the publication of this paper.**

Thank you for the comment; it is a good suggestion to bolster the conclusion by stating specifically that there is
50 *no 'silver bullet' classification scheme that can deduce aerosol type from only a combination of intensive and extensive aerosol optical properties. The authors also added specific ideas for future work needed into the conclusion. Please see changes to manuscript below.*

6

Pag. 20, lines 9-17: "This study has further assessed existing aerosol typing schemes, provided additional methods that can be implemented to reduce ambiguity in typing schemes, elucidate aerosol conditions that accompany aerosol type, and allow for identification of multiple aerosol types at one site. A major conclusion from the analysis, however, is that there is no combination of extensive and/or intensive optical properties that allow for a perfect classification of aerosol types. Prior knowledge of the measurement site can help inform aerosol classification schemes, but obscurity remains in these techniques. Furthermore, this paper highlighted the need for further analyses and suggests specific ideas for future work needed to progress and refine aerosol typing schemes that infer aerosol type from optical properties. Namely, repeating this analysis with concurrent aerosol chemical and optical measurements to verify aerosol classification thresholds will be essential to expand and improve aerosol classification schemes."

Revised manuscript with tracked changes:

Classifying aerosol type using in situ surface spectral aerosol optical properties

L. Schmeisser*[1,2], E. Andrews[1,2], J.A. Ogren[1], P. Sheridan[1], A. Jefferson[1,2], S. Sharma[3], J.E. Kim[4], J.P. Sherman[5], M. Sorribas[6], I. Kalapov[7], T. Arsov[7], C. Angelov[7], O.L. Mayol-Bracero[8], C. Labuschagne[9,10], S.-W. Kim[11], A. Hoffer[12], N.-H. Lin[13], H.-P. Chia[13], M. Bergin[14], J.Y. Sun[15], P. Liu[16], H. Wu[16]

[1]National Oceanic and Atmospheric Administration, Earth Systems Research Laboratory, Boulder, CO, USA
[2]University of Colorado at Boulder, CIRES, Boulder, CO, USA
[3]Environment and Climate Change Canada, Science and Technology Branch, Ontario, Canada
[4]Environmental Meteorology Research Division, National Institute of Meteorological Sciences
[5]Appalachain State University, Boone, NC, USA
[6]Atmospheric Sounding Station-El Arenosillo, Atmospheric Research and Instrumentation Branch, INTA, 21130, Mazagón, Huelva, Spain
[7]Institute for Nuclear Research and Nuclear Energy by the Bulgarian Academy of Sciences, Sofia, Bulgaria
[8]University of Puerto Rico, Department of Environmental Science, San Juan, PR, USA
[9]South African Weather Service, Stellenbosch, South Africa
[10]Unit for Environmental Sciences and Management, North-West University, Potchefstroom Campus, South Africa
[11]Seoul National University, Seoul 08826, Korea
[12]MTA-PE Air Chemistry Research Group, Veszprém, P.O. Box 158, H-8201, Hungary
[13]National Central University, Department of Atmospheric Sciences, Chung-LI, Taoyuan City, Taiwan
[14]Duke University, Department of Civil & Environmental Engineering, Durham, NC
[15]State Key Laboratory of Severe Weather & Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing 100081, China
[16]China GAW Baseline Observatory, Qinghai Meteorological Bureau, Xining 810001, China
Correspondence to: Lauren Schmeisser (lauren.schmeisser@gmail.com)

*Corresponding author now at: University of Washington, Department of Atmospheric Sciences, Seattle, WA, USA

Abstract

Knowledge of aerosol size and composition is important for determining radiative forcing effects of aerosols, identifying aerosol sources, and improving aerosol satellite retrieval algorithms. The ability to extrapolate aerosol size and composition, or type, from intensive aerosol optical properties can help expand the current knowledge of spatio-temporal variability of aerosol type globally, particularly where chemical composition measurements do not exist concurrently with optical property measurements. This study uses medians of scattering Ångström exponent (SAE), absorption Ångström exponent (AAE) and

7

single scattering albedo (SSA) from 24 stations within the NOAA federated aerosol network to infer aerosol type using previously published aerosol classification schemes.

Three methods are implemented to obtain a best estimate of dominant aerosol type at each station using aerosol optical properties. The first method plots station medians into an AAE vs. SAE plot space, so that a unique combination of intensive properties corresponds with an aerosol type. The second typing method expands on the first by introducing a multivariate cluster analysis, which aims to group stations with similar optical characteristics, and thus similar dominant aerosol type. The third and final classification method pairs 3-day backward air mass trajectories with median aerosol optical properties to explore the relationship between trajectory origin (proxy for likely aerosol type) and aerosol intensive parameters, while allowing for multiple dominant aerosol types at each station.

The three aerosol classification methods have some common, and thus robust, results. In general, estimating dominant aerosol type using optical properties is best suited for site locations with a stable and homogenous aerosol population, particularly continental polluted (carbonaceous aerosol), marine polluted (carbonaceous aerosol mixed with sea salt), and continental dust/biomass sites (dust and carbonaceous aerosol); however, current classification schemes perform poorly when predicting dominant aerosol type at remote marine and Arctic sites, and at stations with more complex locations and topography where variable aerosol populations are not well represented by median optical properties. Although the aerosol classification methods presented here provide new ways to reduce ambiguity in typing schemes, there is more work needed to find aerosol typing methods that are useful for a larger range of geographic locations and aerosol populations.

1. Introduction

Although it is well established that aerosol particles affect the radiative forcing of climate both directly by scattering and absorbing sunlight and indirectly by influencing cloud formation and precipitation, aerosols still remain a primary source of uncertainty in assessing the Earth's radiative budget (Boucher et al., 2013). This uncertainty arises from a large range of aerosol chemical and physical properties, as well as from the high spatio-temporal variability of aerosol particles. In order to help reduce this uncertainty and be able to better predict climatic effects of aerosols, there is a need for long-term global monitoring of aerosols (Hansen et al., 1996), compiling records not only of aerosol loading but also of aerosol characteristics and type.

Determination of aerosol type (e.g., black carbon, sea salt, dust, etc.), which is defined by the size and composition of an aerosol, is important in characterizing the role of aerosols in atmospheric processes and feedbacks, since different aerosol types have different radiative forcing effects and atmospheric behavior. Additionally, knowledge of aerosol type helps identify aerosol source, which can be useful in implementing controls or policies to reduce aerosols that negatively influence air quality and public health, and also to better understand atmospheric dynamics and long-range transport. Constraining aerosol type is also needed for improving aerosol satellite retrieval algorithms and for validating climate models (Russell et al., 2014).

Recent studies, discussed below, present classification schemes to infer aerosol type from intensive optical properties, which are calculated from ratios of extensive properties and thus not directly dependent on the aerosol amount. Successful application of this method could allow for access to aerosol composition information from remote or in situ optical property measurements that do not otherwise provide an indication of aerosol type.

8

2. Background

Three optical properties that hold information on aerosol type include scattering Ångström exponent (SAE), absorption Ångström exponent (AAE) and single scattering albedo (SSA). SAE represents the wavelength dependence of scattering, and varies inversely with particle size, so that small values of SAE indicate larger aerosol particles (e.g., dust and sea salt), and large values of SAE indicate relatively smaller aerosol particles (Schuster et al., 2006; Bergin et al., 2000 and references therein). AAE represents the wavelength dependence of absorption and depends on the composition of absorbing aerosols, such that aerosol materials have a unique range of AAE values (Russell et al., 2010; Bergstrom et al., 2002, 2007). Black carbon (BC), for example, has a theoretical AAE value around 1, while dust aerosol typically has AAE values greater than 2 (Bergstrom et al., 2002, 2007; Kirchstetter et al., 2004), though AAE of ambient aerosol will likely evolve with atmospheric processing and depend strongly on composition (BC-to-OA (organic aerosol) ratio), coating and size (Saleh et al., 2014; Costabile et al., 2017; Moosmüller et al., 2011). SSA is the ratio of scattering to extinction (absorption + scattering) and provides information on aerosol darkness and composition, and may determine the net sign of an aerosol's radiative forcing (Hansen et al., 1997). High SSA values near 1 indicate low- or non-absorbing "white" aerosols, while low SSA values (below 0.85) indicate "darker" highly-absorbing aerosols, and thus an SSA value can be used to characterize the aerosol type (Bergstrom et al., 2002; Russell et al., 2010; Gyawali et al., 2012). Equations for calculating these properties from extensive optical parameters are found in Sect. 4. Many studies have used the information inherent in these optical properties to predict aerosol type; Table 1 provides a review of previous studies that have utilized intensive optical property thresholds to identify aerosol type.

The studies listed in Table 1 all take slightly different approaches to show that intensive aerosol optical properties (SAE, AAE, and SSA) can be utilized to classify aerosol type. Bahadur et al. (2012) determine a scheme to partition various absorbing aerosol types based on absorbing aerosol optical depth measurements from numerous AERONET sites that represent a single absorbing aerosol, and test the proposed scheme using California AERONET sites with mixed aerosols. Cazorla et al. (2013) also make use of California AERONET sites by combining the measured aerosol optical properties with in situ aerosol chemical composition measurements from an aircraft campaign to create a matrix that delineates aerosol type in an AAE vs. SAE plot space. Eleven AERONET sites from around the globe are used in the study by Russell et al. (2010) to show that AAE values from full column measurements are highly correlated with aerosol type, in general agreement with the two previously mentioned AERONET aerosol typing schemes that suggest AAE values near 1 indicate fossil fuel burning aerosol, higher AAE values indicate absorbing organic carbon (OC)/biomass burning aerosols, and the highest AAE values indicate dust aerosols.

In situ measurements have also been used for aerosol classification schemes. In situ optical measurements from the INTEX-NA aircraft campaign are used by Clark et al. (2007) to separate biomass burning from pollution plumes. Costabile et al. (2013) propose a scheme to classify aerosols based on absorption and scattering values, using 2 years of in situ urban data from Rome, Italy coupled with numerical simulations to create a paradigm linking key aerosol populations to their unique aerosol optical properties. Six months of optical property measurements from the in situ monitoring site in Gosan, South Korea are used by Lee et al. (2012) and categorized by air mass type (either pollution or dust) using chemical composition, back trajectories and meteorological conditions, and SAE and AAE values are analyzed, yielding results that show dust air masses have the highest AAE values, with OC polluted air masses showing the next highest AAE values. Cappa et al. (2016)

9

utilized surface in situ measurements from the CARES field campaign in California to categorize aerosol they observed and to suggest some modifications to the Cazorla et al. (2013) aerosol classification scheme. Finally, Yang et al. (2009) used the distinct SSA, AAE and SAE values of different air plumes in the EAST-AIRE campaign to identify absorption contributions from desert dust, biomass burning, industrial plumes, and clean air in Beijing, China. It is worth mentioning that some studies

5   take into account the spectral dependence of SSA in aerosol classification schemes (Li et al, 2015; Russell et al., 2010). This parameter was calculated for the monitoring stations in this study, but was not useful in classifying aerosol type compared to the other optical properties discussed; therefore, the spectral dependence of SSA is not discussed here.

Care must be taken in comparing thresholds from all aforementioned studies, as differences are likely between column-average, ambient AERONET measurements and low-RH, surface in situ measurements. Furthermore, different wavelength

10 pairs are used to calculate AAE and SAE depending on the study. In general, however, all studies suggest similar typing thresholds. Most previous works agree that AAE values around 1 represent BC and/or fossil fuel burning aerosols, and higher AAE values indicate light-absorbing OC (aka brown carbon (BrC)) and/or dust, and that high SAE values are associated with small anthropogenic aerosols (e.g., BC, sulfates, or nitrates) and low SAE values are associated with large aerosols like sea salt and dust.

15 This paper aims to assess the applicability of previous typing methods/schemes to data from 24 in situ monitoring sites within the NOAA/ESRL Federated Monitoring Network, and to explore how typing schemes may be improved based on methods using cluster analyses and air mass back trajectories. The following questions are addressed: (1) Are the relationships between SAE and AAE data from 24 stations in the NOAA federated monitoring network consistent with relationships used to identify dominant aerosol type using aerosol classification schemes previously reported in the literature?; (2) Can multivariate cluster

20 analyses on aerosol properties be used to reduce both the ambiguity in inferring likely dominant aerosol type from median aerosol optical properties, and the uncertainty in aerosol type optical property thresholds?; and (3) How can back trajectory clusters and subsequent information on air mass source help elucidate multiple aerosol types at individual sites?

The literature on classifying aerosols has been largely dominated by analysis of ground-based remote sensing or satellite data (Cazorla et al., 2013; Russell et al., 2010; Russell et al., 2014; Omar et al., 2005; Giles et al., 2012; Bergstrom et al., 2007;

25 Bergstrom et al., 2010; Bahadur et al., 2012; Dubovik, 2002), with fewer analyses done using surface in situ aerosol optical property measurements (Cappa et al., 2016; Costabile et al., 2013; Yang et al., 2009; Lee et al., 2012). The analyses in this paper utilize ground-based in situ spectral optical data that afford a unique insight into long-term, quality-assured point observations. Furthermore, since the in situ data sets used in this study are not restricted by AOD thresholds as are AERONET data sets, they offer a more thorough look at regions with relatively clean air.

30 Unlike most previous studies, this study looks at long-term records of aerosol optical properties, and does so at a wide range of geographic locations, including mountaintop, desert, continental and coastal sites. Not only does the study offer a wide range of aerosol types to be analyzed in an individual geographic location, but provides analysis of the same aerosol type in different geographic locations.

3. Site descriptions

35 This study investigates aerosol populations at 24 monitoring stations in the NOAA/ESRL Federated Aerosol Monitoring

10

| Deleted: in this study. |
| Deleted: and/or dust and the highest AAE values indicate |
| Deleted: |

Network. Sites were selected for the study based on availability of data - each site had to meet the following criteria: (1) aerosol optical data available at 3 wavelengths, and (2) long-term (>6 months) continuous measurement records of scattering and absorption coefficients during the two year time period 2012-2013, unless otherwise noted (see Table 2 for time range for each site). The ARM Mobile Facility (AMF) (part of the U.S. Department of Energy's ARM Climate Research Facility) deployments, indicated in bold in Table 2, are typically one to two year deployments. Most of the AMF measurement times do not overlap with the 2012-2013 analysis period, but should nevertheless be comparable to other sites, and are included as a means of broadening the range of geographic locations for the analysis. One advantage of this study is the wide diversity of location types and observed aerosol loadings (which span over 3 orders of magnitude). This study includes sites in both the northern and southern hemispheres, ranging in altitude from sea level to 3800 m above sea level (asl), and with various climate regimes including marine, continental and Arctic. The sites experience different levels of anthropogenic influence ranging from clean remote sites to very polluted urban sites. The 24 stations are described in Table 2, and Fig. 1 shows a map of the stations.

Table 2 presents monitoring site location, latitude, longitude, altitude, scattering and absorption instruments, date range of data utilized, site classification, and site description for 24 monitoring stations in the NOAA/ESRL Federated Aerosol Monitoring Network. Bolded station names in the table indicate sites where the short term AMF was deployed.

Sites are categorized based on the site's geography and surrounding land use. Arctic sites are at latitudes greater than 70 N. Continental polluted sites have influence from urban and industrial pollution. Continental dust/biomass sites are generally more rural with influence from desert dust and/or biomass burning. Marine clean sites are in remote coastal locations, have little influence from pollution sources (except perhaps from long-range transport events), and see an abundance of marine aerosols. Marine polluted sites are also in coastal locations and may measure pollution aerosols (from continental air masses) or marine aerosols (from oceanic air masses) or some combination thereof, depending on the wind direction. Mountaintop classifications indicate sites that are higher than 2800m in elevation; these high altitude monitoring stations sample both free troposphere air and air masses transported from lower elevations due to upslope/downslope flow. Site classification is inherently subjective and not always clear-cut. The authors acknowledge that sites could be considered to have more than one classification and have multiple aerosol types. However, the classifications were designated based on 'best fit' to the site characteristics and are intended to be representative of the dominant aerosol type at each site.

4. Data and Instruments

The data sets used for the analysis are comprised of in situ scattering and absorption coefficients (σ_{sp} and σ_{ap}, respectively), which are quality assured and used to calculate additional parameters (AAE, SAE and SSA) as described in Eqs. (1)-(3). One-hour averaged data are used for the assessment of aerosol classification schemes and the multivariate cluster analysis. However, we use 6-hour averaged optical properties for the back trajectory analysis, since back trajectories are run at 6-hour intervals. Datasets from NOAA and collaborators are publically available from the World Data Center for Aerosols (http://ebas.nilu.no/), with the exception of WLG data, while the AMF datasets are publically available from DOE (http://www.arm.gov/).

Scattering coefficients were obtained with a TSI 3563 integrating nephelometer (TSI Inc.) at all sites, operating at wavelength channels 450, 500 and 700 nm. Absorption coefficients were measured by either a 3-wavelength Particle Soot Absorption

11

Deleted:

Deleted: , size cuts

Photometer (PSAP, Radiance Research), or a 3-wavelength Continuous Light Absorption Photometer (CLAP, NOAA). The PSAP instruments operate at wavelengths 467, 530, and 660 nm, and CLAP instruments operate at wavelengths 467, 528, and 652 nm. In either case, the σ_{ap} values are corrected to 450, 550, and 700 nm (using AAE) so as to match the wavelengths of the σ_{sp} measurements.

Table 2 indicates which instruments operate at each station. At MLO and BND, data from both the PSAP and CLAP were utilized, since at both stations the PSAP was replaced with a CLAP in the middle of the study period. An analysis of concurrent PSAP and CLAP measurements shows that the two instruments produce comparable measurements, and thus combining or directly comparing data from both instruments is not expected to affect results (Ogren et al., 2013).

To ensure datasets are comparable across monitoring stations, all data are quality controlled. In order to minimize aerosol hygroscopic effects, measurements at all stations (except SUM and SPL) are made at a reduced relative humidity (RH < 40%) by heating the inlet air or by diluting with filtered, dry air. The inlets at most sites are either gently heated (heating does not exceed 40°C) with a stack heater or a small heater by the impactor, and are only utilized if the relative humidity exceeds 40%. Although heating the sampling inlet can cause loss of organic and volatile aerosol material, which can alter the aerosol spectral optical properties, this is not expected to substantially impact results here. Studies that analyze the amount of volatile components removed at 40°C (by a thermal denuder) is less than 10% (Mendes et al., 2016; Huffman et al., 2009). For this particular study, we do not have the data necessary to evaluate the extent to which aerosol optical properties are affected by the heating, but evidence from other studies suggests the effect is likely small.

Monitoring station buildings are also temperature controlled, and inlet stacks have protective caps and screens to prevent interference from precipitation, insects or debris. All aerosol scattering coefficient measurements from the TSI nephelometers are corrected for angular non-idealities using corrections from Anderson and Ogren (1998). After the corrections, scattering coefficients measured by the nephelometer have an uncertainty of 9.3% for the 10 μm size cut, based on the analysis by Sherman et al. (2015). The Sherman et al. (2015) calculations represent median continental conditions, and might change at sites with cleaner or more polluted conditions. Aerosol absorption coefficient measurements from PSAP and CLAP instruments are adjusted for flow rate, spot size, and aerosol scattering, using the correction from Bond et al. (1999) and further adjusted for wavelength based on corrections from Ogren (2010). After corrections, absorption coefficients measured by the PSAP or CLAP have an uncertainty of ~20% (Sherman et al., 2015). Finally, all data are passed through a quality assurance/quality control editing process in which measurement records are screened for atypical aerosol parameters (see Delene and Ogren (2002) and Sheridan et al. (2015) for detailed descriptions of quality assurance and quality control procedures). Points that appear anomalous due to local pollution sources (non-representative of regional aerosol), instrument error or excessive noise are not included in this analysis.

The measured scattering and absorption coefficients are extensive aerosol properties because they depend on the amount of aerosol present (Ogren, 1995; Delene & Ogren, 2002). Intensive aerosol optical properties are calculated from ratios of the extensive properties. The aerosol intensive properties including absorption Ångström exponent (AAE), scattering Ångström exponent (SAE), and single scattering albedo (SSA), are of primary interest to this study since they contain information on aerosol size or composition, and are calculated as indicated in the following equations:

12

<div style="margin-left:auto">

Deleted: 700nm

Deleted: A comparative

Moved (insertion) [1]

</div>

$$AAE_{\lambda1/\lambda2} = \frac{-\log(\frac{\sigma_{ap,\lambda1}}{\sigma_{ap,\lambda2}})}{\log(\frac{\lambda_1}{\lambda_2})} \quad (1)$$

$$SAE_{\lambda1/\lambda2} = \frac{-\log(\frac{\sigma_{sp,\lambda1}}{\sigma_{sp,\lambda2}})}{\log(\frac{\lambda_1}{\lambda_2})} \quad (2)$$

$$SSA_{\lambda1} = \frac{\sigma_{sp,\lambda1}}{\sigma_{sp,\lambda1} + \sigma_{ap,\lambda1}} \quad (3)$$

where $\sigma_{ap,\lambda1}$ represents absorption coefficient at wavelength λ_1, and $\sigma_{ap,\lambda2}$ represents absorption coefficient at wavelength λ_2. Similarly, $\sigma_{sp,\lambda1}$ and $\sigma_{sp,\lambda2}$ represent scattering coefficients at wavelengths λ_1 and λ_2, respectively. Unless otherwise indicated, all data presented here refer to the green wavelength channel (550 nm) for SSA, absorption, and scattering coefficient values, or the blue/red wavelength pair (450 nm/700 nm) for the SAE and AAE values. CLAP and PSAP wavelengths were adjusted to match the nephelometer wavelengths to compute the intensive variables.

Only aerosol measurements where $\sigma_{sp} > 1$ Mm^{-1} and $\sigma_{ap} > 0.5$ Mm^{-1} are included in the analyses. Data below these values are less reliable due to instrument noise at low aerosol loading, thus the constraints are meant to act as noise thresholds. This inherently adds bias to the data, as monitoring sites with consistently low absorption and scattering coefficients may end up with limited data points after the thresholds are applied, leaving measurement records with higher loadings that may not be fully representative of typical aerosol populations at the site. This constraint has the greatest effect on clean sites like ALT, BRW, and SUM (which measure Arctic air), BEO and MLO (which sometimes measure free tropospheric air), and CPR, CPT, PVC, PYE, and THD (which sometimes measure clean marine air). The constraints push the extensive scattering and absorption values higher. More details on the effect of the thresholds on the analysis of clean stations can be found in Table S5 in the supplemental materials.

There are some differences in monitoring station data that may affect the results of the following analyses, and are noted here. SUM utilizes a 2.5 μm size cut, while all other stations use a size cut of 1 and 10 μm, but only the 10 μm data are used in this study. This size cut discrepancy will bias SUM data towards higher SAE values than would be found with a larger size cut. Since ARM station data records are typically less than one year in length, while all other station data are 2 years in length, any site-specific seasonal variations may not be captured in the ARM data records. Furthermore, ARM measurement times and CPT times typically do not overlap with the baseline study period of 2012-2013, so any extreme events specific to those years are not reflected in the CPT (data only from years 2010-2011) or ARM (FKB, GRW, NIM, PGH, PVC, PYE) sites measurements.

5. Data Analysis Methods

The aerosol classification analysis presented here proceeds in three steps: (1) Application and assessment of previous aerosol typing schemes: presenting station intensive optical property medinas in an AAE vs. SAE plot space modeled closely after Cappa et al. (2016) in order to link a combination of AAE and SAE values to aerosol type; (2) Multivariate cluster analysis: performing a multivariate cluster analysis to group stations with like optical properties to better infer a common aerosol type;

13

and (3) Back trajectory analysis: combining back trajectories and the land type over which they traveled with aerosol optical properties to better understand the relationship between trajectory origin (proxy for likely aerosol type) and aerosol intensive properties, while allowing more than one dominant aerosol type at each station. The methods for these analysis techniques are described in detail here.

5.1 Methods for application and assessment of previous aerosol typing schemes

Like many previous studies (Cappa et al., 2016; Cazorla et al., 2013; Costabile et al., 2013; Yang et al., 2009; Russell et al., 2010; Lee et al., 2012; Bahadur et al., 2012), an AAE vs. SAE plot space is used here to visualize relationships between aerosol optical properties and likely aerosol type. Since SAE indicates aerosol size and AAE holds information on aerosol composition and size (Costabile et al., 2017), a unique combination of the two, and thus where that combination falls within the AAE vs. SAE plot space, suggests a particular aerosol type. Many previous studies use chemical composition data (Costabile et al., 2013; Lee et al., 2012; Cazorla et al., 2013) or numerical simulations (Costabile et al., 2013) to validate the proposed aerosol classification scheme; however, since neither of those methods are available for this study, thresholds from previous studies are used here to infer likely dominant aerosol type, and results are assessed based on knowledge of the site. For the first iteration of the analysis, long-term optical property medians from multiple stations are presented in one plot space, for a comparative overview of inferred dominant aerosol type at many sites. A variation of the Cappa et al. (2016) classification matrix is used here. The version used here omits 'large black particles' from the lower left plot space designation, as this does not correspond with data presented here.

It should be noted that the Cappa et al. (2016) and Cazorla et al. (2013) matrices are very similar. Both designate high SAE and high AAE values as BrC or mixed BC/BrC (though Cazorla et al. (2013) refers to BrC as OC). Both designate low SAE values and high AAE values as dust or dust mixed with BC and BrC, and both suggest an AAE value around 1, accompanied by higher SAE values indicates aerosol populations dominated by BC. Three main differences between the matrices can be identified. The Cappa et al. (2016) matrix makes more specific designations of aerosol mixtures (e.g., adds 'mixed dust/BC/BrC' and 'large particle/BC mix'). The Cappa et al. (2016) matrix also replaces the Cazorla et al. (2013) matrix designation of 'large coated particles' with 'large particle/low absorption mix or large black particles'. Finally, the Cappa et al. (2016) matrix replaces the Cazorla et al. (2013) matrix designation of 'EC' with 'small particle/low absorption mix'. We chose to primarily use the Cappa et al. (2016) matrix since it is based on in situ data (Cazorla et al. (2013) is based on AERONET data), and since the aerosol designations seemed to align most closely with our data. Results are presented in Sect. 6.1.

5.2 Methods for multivariate cluster analysis

In order to infer a more accurate representation of aerosol type using intensive optical properties as an indication of aerosol size/composition and extensive optical properties as an indication of loading, a multivariate clustering analysis is performed to build on the first classification method. A cluster analysis is the process of statistical grouping that yields 'clusters' with similar characteristics. A few other studies also implement multi-dimensional clustering as a means of solidifying aerosol property thresholds for different aerosol types (Russell et al., 2010; Omar et al., 2005; Levy et al., 2007). In this study, a cluster analysis is used to determine groups of stations with similar aerosol type based on aerosol optical properties. The clusters are then plotted in a 3D parameter space (AAE vs. SAE vs. $\log(\sigma_{sp})$) as a means of visualizing any spatial patterns

14

that emerge.

The k-means clustering algorithm was run using medians of four aerosol optical property parameters – SAE, AAE, SSA, and the log of the scattering coefficient ($\log(\sigma_{sp})$) – from hourly averaged records at each monitoring station. The scattering coefficient, σ_{sp}, is an indication of aerosol loading and is implemented here as an additional parameter to improve the inference of aerosol types. The log of σ_{sp} (in Mm^{-1}) is used rather than the raw σ_{sp} median in order to make the scattering coefficient values more comparable with the magnitude of the optical property values, so the clustering is not dominated by one parameter. While the magnitude of loading (σ_{sp}) alone does not correspond to a specific aerosol type (for example, high loadings can be observed for dust, pollution, or biomass burning events) it may act as a secondary indicator of aerosol conditions (i.e., frequency of aerosol type occurrence, loading) and source contributions, so it is included in the clustering analysis.

To run the clustering algorithm, a number of clusters 'k' is selected. Choosing the 'k' initial seed points is inherently subjective – in this analysis, k needs to be small enough such that the number of stations that fall into each cluster makes for a meaningful grouping, and large enough such that a distinction between station groups is apparent. The algorithm then takes 'k' initial seed points at random and iteratively assigns each point to the nearest cluster centroid taking into account the clustering properties. The next iteration chooses 'k' new seed points and repeats the process until the algorithm converges. In this study, six clusters are selected, creating six unique groups each with similar SAE, AAE, SSA, and $\log(\sigma_{sp})$ characteristics. Each monitoring station was assigned to one of the six clusters produced from the algorithm, and the groupings were used to further analyze aerosol type and conditions. Results are presented in Sect. 6.2.

5.3 Methods for back trajectory analysis

The NOAA Air Resources Laboratory Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Rolph, 2012) was utilized to produce 3-day air mass back trajectories at 6-hour intervals for the entirety of the measurement period at each station. A cluster analysis was performed in HYSPLIT on the back trajectories from individual stations in order to group air masses of similar speed, direction, and altitude. A thorough description of the HYSPLIT cluster analysis methodology can be found in Kelly et al. (2013). The number of back trajectory clusters differs by station, since the selection of cluster numbers is dependent on the individual data set and is somewhat subjective. For this study, and in adherence with typical clustering methodology, a plot of total spatial variance versus number of clusters was used to determine cluster number; the cluster number point just before the total spatial variances increases dramatically is the number of clusters used for analysis at that site. From the cluster analysis, each 6-hour (0, 6, 12, 18 UTC) trajectory was assigned a cluster number and paired with 6-hour averaged aerosol optical property data from the monitoring station for which the back trajectories were produced. For example, the back trajectory at 6 UTC was paired with aerosol optical property data averaged over hours 3-9 UTC. The paired optical property data were then plotted in the AAE vs. SAE plot space and color coded based on back trajectory cluster number, individually for each site. The method described assumes clustered back trajectories may carry similar aerosol type(s) that may be unique from aerosol found in another back trajectory clusters, allowing for temporal variation in aerosols at a site that is dependent on the geography from which the air masses arrived at the station. Results are presented in Sect. 6.3.

Moved (insertion) [3]

15

6. Results

6.1 Application and assessment of previous aerosol typing schemes

The median and interquartile spread of SAE, AAE, SSA, scattering coefficient and absorption coefficient values at each site are presented in Table 3. Additionally, Table 3 indicates the aerosol type as determined by the variation of the Cappa et al. (2016) matrix overlaid on the plot of optical property medians in Fig. 2(b) ('aerosol type before clustering'), as well as the aerosol type determined from a clustering analysis ('aerosol type after clustering'), as described in the next section. Descriptions of the aerosol types can be found in Cazorla et al. (2013) and Cappa et al. (2016).

Median AAE and SAE values for each station are shown in Fig. 2(a) along with bars that represent the interquartile spread (25th to 75th percentiles) of the data. Points are shaded by median SSA value at that station. Medians are used in order to minimize influence from outliers. There are no strong spatial patterns visible in SSA shading within the AAE vs. SAE plot space in Fig. 2(a). Stations with high median SAE (smaller particles) tend to have slightly lower median SSA values (darker particles) than those with low median SAE, and vice versa. However, there are exceptions to this tendency, with NIM having a low median SAE value and relatively low median SSA, and PVC having a high median SAE value and relatively high median SSA. Previous studies established that SSA and the wavelength dependence of SSA can be used to signify aerosol type (Yang et al., 2009; Russell et al., 2010). A three-dimensional plot space helps visualize the relationships amongst SAE, AAE and SSA. This will be further explored in the next section.

Figure 2(a) shows the wide variance of intensive properties at any one site, with values spanning beyond the optical property signatures of a single aerosol type. For example, CPR has interquartile AAE values ranging from 1.16-2.65, a spread that encompasses multiple potential aerosol compositions, as outlined by the thresholds in Table 2 and by the classification matrix in Fig. 2(b). Interquartile ranges conservatively bound the intensive properties and thus represent the dominant aerosol type at each monitoring site. Some, if not all, of the sites could have multiple aerosol types that are not well represented by the medians illustrated in Fig. 2, as discussed in the next section.

Figure 2(b) shows the same optical property medians that are plotted in Fig. 2(a). Station points are colored by station location type (as listed in Table 2), with the aerosol classification matrix from Cappa et al. (2016) overlaid on the plot space. Optical properties from the 24 NOAA Federated Aerosol Network stations were evaluated with multiple existing published aerosol classification schemes; however, given the clear visualization and complete characterization of the parameter space afforded by the Cappa et al. (2016) matrix, that is the only scheme used for a visual comparison in this study. The station location type provides the reader guidance on what aerosol types might be expected at the site.

There is a natural clustering of all continental polluted sites on the right hand side of the plot in Fig. 2(b), in the section Cappa et al. (2016) designated as BC dominated. Median AAE > 1 at these sites is consistent with other studies (Russell et al., 2010; Lee et al., 2012; Yang et al., 2009; Cazorla et al., 2013). Furthermore, both remote/clean marine (e.g., GRW, PYE, THD) sites and dust-influenced sites (e.g., NIM) tend to fall on the left hand side of the plot with low SAE values, indicative of sea salt, highly processed and coated particles, or dust (Cappa et al., 2016; Cazorla et al., 2013; Lee et al., 2012; Clarke et al., 2007; Yang et al., 2009). The largest median AAE values are observed at NIM and CPR, both of which experience Saharan dust events. NIM is located at the southern edge of the Saharan desert. Dust transport to CPR is predominantly from the African

16

Sahel region (Prospero et al., 2014). Although ARN experiences Saharan dust events (Toledano et al., 2007), these events are not frequent enough to substantially influence the median in situ aerosol optical properties. The high AAE values at sites influenced by dust agree with the findings of Russell et al. (2010), Lee et al. (2012), and Yang et al. (2009), which identified dust aerosol as having the largest AAE values of observed aerosol types. These sites also fit in well with the Cappa et al.

5 (2016) and Cazorla et al. (2013) matrices. Aerosol types assigned to the marine THD, ARN, GRW, PYE, CPT, CPR, and PVC sites by the Cappa et al. (2016) and Cazorla et al. (2013) aerosol classification schemes exhibit high variance in their properties, indicating a diverse influence of aerosol. For example, the high SAE values at PVC show the strong influence of transport from nearby urban centers of Boston and Providence as well as pollution from summer traffic on Cape Cod, which dominate the effect of marine aerosol on the site's median SAE value (Titos et al., 2014).

10 Figure 2 illustrates that the Cappa et al. (2016) aerosol classification scheme agrees with the expected dominant aerosol type at continental polluted (BC dominated), marine polluted (BC dominated mixed with sea salt), and continental dust/biomass sites (Mixed dust/BC/BrC). On the other hand, the classification scheme assigns dominant aerosol type at remote marine sites and Arctic sites that differ from what would be expected at these sites, given their location and proximity to aerosol sources. Marine clean sites in this analysis (CPR, CPT, GRW, PYE, THD) have a wide spread of AAE values, and although they are

15 all situated on the left side of the plots in Fig. 2, due to a common low SAE value among the sites, they are not clustered along the AAE axis. All stations in the plot with median SAE values less than or equal to 1.1 are classified as either continental dust/biomass or marine clean, but those classifications cannot be distinguished in the Cazorla et al. (2013) matrix or the modified Cappa et al. (2016) matrix. An improved matrix may include dust, marine aerosol, large coated particles and/or highly processed (aged) particles as possible aerosol types for SAE values less than 1.1. Figure 2(a) shows that marine clean

20 sites exhibit much higher SSA values than the continental dust/biomass sites with similarly low SAE values, which suggests that the addition of more optical parameters, including SSA, into the clustering analysis could yield more optimized aerosol classification results. Consequently, in the next section, results from a multivariate cluster analysis are used to help reduce ambiguity in aerosol classification and further hone potential aerosol type identification.

6.2 Multivariate cluster analysis

25

Figure 3 shows median optical property values, plotted in a 3D AAE vs. SAE vs. $\log(\sigma_{sp})$ parameter space. Station points are color coded by cluster number, and sized by SSA median values. Not only does the 3D parameter space provide a robust visualization of the clustering results, it provides further insight into an aerosol population than the AAE vs. SAE parameter

30 space used previously, since information on loading and SSA are also visible.

Table 4 shows median AAE, SAE, SSA and $\log(\sigma_{sp})$ values along with interquartile values for each cluster, plus aerosol type and condition (where applicable) based on cluster optical property medians, thresholds from previous literature, and previous knowledge of station characteristics at the sites within each cluster.

In the 3D plot seen in Figure 3, stations that fall within the same cluster number also are located near each other in the three-

17

Deleted: Propero

Deleted: Cazorla

Deleted: matrix

Deleted: scheme

Deleted: Cazorla

Deleted: 2013

Deleted: EC/OC aerosol

Deleted: EC/OC aerosol

Deleted: and/or EC/OC

Deleted: 5

Deleted: … [1]

Moved up [2]: A cluster analysis is the process of statistical grouping that yields 'clusters' with similar characteristics. A few other studies also implement multi-dimensional clustering as a means of solidifying aerosol property thresholds for different aerosol types (Russell et al., 2010; Omar et al., 2005; Levy et al., 2007). In this study, a cluster analysis is used to determine groups of stations with similar aerosol type based on aerosol optical properties. The clusters are then plotted in a 3D parameter space (AAE vs. SAE vs. $\log(\sigma_{sp})$) as a means of visualizing any spatial patterns that en … [2]

Formatted: Heading 2, Left

Formatted: Character scale: 0%, Border:: (No border), Pattern: Clear (Black)

dimensional parameter space, making for an effective visualization of the relationship between aerosol population and optical properties. Furthermore, stations in each cluster generally share similar site characteristics and expected aerosol type. Discussion of results for each individual cluster is available in supplemental materials, while more general results are discussed here.

5   The clusters presented in Figure 3 generally group together sites that are expected to have similar aerosols, and the expected aerosol characterizations generally agree with the aerosol type inferred with the aerosol classification schemes. The method does particularly well with identifying aerosol type at stations with a more or less stable, homogeneous aerosol population, including continental stations sampling BC dominated aerosol (i.e., Clusters 2 and 3), as well as the continental stations sampling high loads of dust aerosol (i.e., Cluster 4). The method also does a fair job at identifying remote Arctic or
10   mountaintop sites (i.e., Cluster 1) that sample large processed particles (due to aging during transport) and occasional instances of local pollution. These methods do not do as well at identifying dominant aerosol type at stations with more complex location and topography, where variable aerosol populations that depend on wind direction and/or occasional extreme aerosol events are not well characterized by median optical properties within the parameter space.

An advantage to the incorporation of $\log(\sigma_{sp})$ into the clustering algorithm and the 3D parameter space plot is that it allows
15   for a more complete picture of aerosol type and conditions at the station. For example, even though the Cappa et al. (2016) aerosol typing scheme assigns a BC dominated aerosol to both Clusters 1 (remote Arctic and mountaintop stations: ALT, BRW, SUM, MLO, SPL) and 2 (heavily polluted urban coastal sites: AMY, GSN), Fig. 3 shows that these clusters are clearly different, given that Cluster 1 exhibits much lower aerosol loading than Cluster 2. The stations in these clusters are indistinguishable within just the AAE vs. SAE 2D parameter space. Using σ_{sp} in the analysis gives further insight into the
20   frequency of occurrence and loading of the inferred aerosol (stations in Cluster 1 measure less BC dominated aerosol than stations in Cluster 2).

There are a few weaknesses to the approaches used thus far in typing aerosols using median optical properties and clustering to reduce ambiguity in the aerosol classification. First, knowledge of station location alone cannot accurately determine the type of aerosols found there (Omar et al., 2005). For example, long-range transport or extreme events may result in aerosols
25   being sampled that are not generally representative of the local geographic region. Second, using a climatological mean or median value of an optical property like SAE or AAE can be misleading in the case that two or more differing aerosols are present at different times over the measurement period. For example, a median SAE value of 1 for a site that measures sea salt (low SAE near 0) over half the measurement period and pollution aerosol (high SAE near 2) over the other half of the measurement period, does not provide any real information about the aerosol population, since neither aerosol type has an
30   SAE value of 1. In order to address these concerns, an additional analysis using air mass back trajectories is performed as a means of exploring the spread in optical property data at each site. This analysis also allows for multiple aerosol types to be present at any one location.

6.3 Back trajectory analysis

The preceding results are derived from application of aerosol typing schemes to median optical properties at multiple stations,
35   a method that depends on the assumption that each site has only a single dominant aerosol type. Many of the sites in this analysis, however, are likely to have a heterogeneous aerosol population with various aerosol types. Backward air mass

18

Deleted: Cluster 1 includes MLO, SUM, BRW, ALT and SPL and is characterized by stations with low aerosol loadings (small σ_{sp}), medium to small aerosol particles (1< SAE<2), and AAE values with a range of 0.5 < AAE < 1.5. All of the stations in Cluster 1 are either Arctic stations (BRW, ALT, and SUM) or remote mountaintop stations (MLO and SPL). The median optical properties of the cluster correspond with an aerosol type of 'large coated particles' or 'EC/OC' according to Cazorla et al. (2013). On the one hand, the inferred aerosol of 'large coated particles' type agrees well with what aerosols would be expected at remote locations such as these, since most aerosols to reach these sites are well aged and could be coated due to the aging time (Jacobson, 2001). However, this aerosol type may need to be generalized to 'processed aerosol', to include a larger diversity of aged and processed aerosol that may be sampled at these remote site locations. An aerosol type of 'EC/OC' would also not be surprising at these sites, given the potential for local aerosol source contributions from biomass or fossil fuel burning, though the generally low scattering coefficients at these sites indicate the presence of anthropogenic aerosols is infrequent.

Deleted: Cluster 2 includes AMY and GSN, the two coastal stations located in South Korea, and is characterized by high aerosol loadings (high σ_{sp}), small aerosol particles (SAE~1.5) and carbonaceous aerosols (AAE~1.2). Existing aerosol classification schemes would designate the median optical property values of this cluster as fossil fuel burning aerosols (EC) and/or biomass burning aerosols (OC) (Cazorla et al., 2013; Lee et al., 2012; Russell et al, 2010; Yang

Moved up [1]: et al., 2009).

Deleted: Given the location of these sites within a highly polluted region and occasional local biomass burning, the inferred aerosol types make sense. The relatively low SAE compared to the sites in Cluster 3 can be attributed to the contribution of sea salt aero[... [3]

Deleted: here

Deleted: EC/OC

Deleted: Cazorla

Deleted: 2013

Deleted: mixed EC/OC

Deleted: EC/OC

Deleted: 5

trajectories are incorporated into the analysis here as a means of both (1) allowing for the consideration of multiple dominant aerosol types at one station and (2) allowing for attribution of likely aerosol source, which can help confirm the practicality of using optical properties to infer aerosol type.

6.3.1 Case studies

Due to a need for brevity, the back trajectory analyses for all 24 stations cannot be presented, so we selected four monitoring stations to present here: Mt. Waliguan, China (WLG), Cape Cod, Massachusetts, USA (PVC), Niamey, Niger (NIM) and Heselbach, Germany (FKB). The 4 sites presented here were chosen to represent cases both where back trajectories helped identify aerosol types and where back trajectories did not elucidate information beyond the initial aerosol classification analysis using median optical properties. Shown for each of the 4 stations (Figs. 4-7) are a map of mean back trajectory paths for each cluster, a plot of trajectory height vs. backward time (color coded by trajectory cluster number), and a plot of AAE vs. SAE properties for 6-hour averaged optical property data, color coded by paired trajectory cluster number, and overlaid by median optical property values of each cluster in the largest color-coded point. If a station's dominant aerosol type differs with air mass origin, these plots can elucidate a station's various aerosol types.

6.3.1.1 Mt. Waliguan, China

The back trajectories at Mt. Waliguan (WLG) were grouped into 4 clusters in HYSPLIT, as shown in Figure 4. Cluster 1 contains ~33% of the site's back trajectories and has origins to the west of the station near Northern Pakistan and traveling through Western China; Cluster 2 contains ~30% of the site's back trajectories and has origins (on average) to the west of the station in rural China; Cluster 3 in contains ~33% of the site's back trajectories and has origins very near the site itself and slightly to the east; and Cluster 4 contains ~3% of the site's back trajectories and has origins to the far northwest of the station, traveling to the station at high altitudes from rural Russia. AAE values are similar for each trajectory cluster, though SAE values vary. Furthermore, the median aerosol optical property values from each of the trajectory clusters are unique, suggesting a variety of aerosol types using thresholds from previous literature (Cazorla et al., 2013; Costabile et al, 2013). The optical properties from the aerosols in back trajectory Cluster 1 (from deserts in Northern Pakistan and Western China) imply a dust mixture. Lower SAE values mean the aerosols from this trajectory cluster are larger, and AAE values near and above 1.5 likely mean dust and/or carbonaceous aerosol mixture (Cazorla et al., 2013). These results support those of Che et al. (2011) and Kivekäs et al. (2009), which cite deserts as aerosol sources from western wind sectors at WLG. Clusters 1 and 2 are most similar in terms of median SAE and AAE values, though the map shows that Cluster 1 trajectories traveled farther in the 3-day period, and thus had faster wind speeds. Cluster 2 and cluster 3 have mean trajectory paths that are relatively short, and thus associated with low wind speeds. This means that these clusters are likely to be more influenced by local aerosol sources. The optical properties of the aerosols from back trajectory Cluster 3 coming from the east suggest BC given the AAE value near 1. This is in agreement with findings of Kivekäs et al. (2009) that show increased particle concentrations from the east of the WLG station indicated anthropogenic pollution. Cluster 4 looks quite different than the other trajectories, and has median optical properties indicative of dust (Cazorla et al., 2013), which makes sense given the trajectory cluster's origin to the northwest of the site (Che et al., 2011; Kivekäs et al., 2009).

19

6.3.1.2 Niamey, Niger

The back trajectories at Niamey (NIM) were grouped into 3 clusters in HYSPLIT, as shown in Figure 5. Cluster 1 contains slightly over half (~53%) of the back trajectories, with air-mass trajectories reaching the site (on average) from the south/southwest, and traveling at a relatively low altitude over populated regions. Cluster 1 differs from Clusters 2 and 3 in that it has a lower median AAE value and higher median SAE value. Given the optical properties of the trajectory cluster 1, along with the knowledge of anthropogenic activities in the source region, the likely dominant aerosol during those trajectories is a biomass burning/soot aerosol mixture (e.g., Osborne et al., 2008; MacFarlane et al., 2009). Clusters 2 and 3 constitute slightly less than half (~46%) of the back trajectories at NIM and originate (on average) from the north and northeast of the site. In Fig. 5, the median optical property values of Clusters 2 and 3 are nearly indistinguishable. For these two clusters, the small SAE values and AAE values above ~1.5 suggest dust mixtures (Cazorla et al., 2013; Lee et al., 2012; Yang et al., 2009). Previous observations by Osborne et al. (2008) noted dust during northerly flow due to the proximity of the Sahara desert to the north/northeast of the site, as did MacFarlane et al. (2009). NIM provides a good example of trajectory analysis elucidating two dominant aerosol types that were obscured when only the climatological medians of AAE and SAE values were evaluated. However, it should be noted that local sources and meteorological conditions also have a large influence on aerosol at the site, in addition to trajectory sources.

6.3.1.3 Cape Code, Massachusetts, USA

Back trajectories at Cape Cod (PVC) were clustered into 3 groups in HYSPLIT, as shown in Figure 6. Cluster 1 contains almost half (~49%) of the trajectories and originates (on average) to the south and southeast of the Cape Cod site along the heavily populated eastern U.S. seaboard. Cluster 2 contains ~43% of the trajectories and (on average) travels to the monitoring station from the Northwest over eastern Canada. Cluster 3 contains only ~8% of trajectories and comes to the station from over the North Atlantic. Cluster 3 is distinct from Clusters 1 and 2 with the lowest SAE value (largest particles), and given its source region suggests at least partial marine sea salt aerosols. Clusters 1 and 2, on the other hand, with continental source regions and optical properties indicative of elemental and organic carbon suggest anthropogenic aerosols. Due to its proximity to both the ocean and large cities like Boston, it is unsurprising that the site measures both marine and urban aerosols, depending on the wind direction. The pairing of back trajectory analysis with optical property classification gives a more detailed picture of the multiple aerosol populations at PVC, in accord with other aerosol research done at the site (Titos et al., 2014). Since the back trajectories from over the Atlantic make up such a small portion of the air masses that arrive at PVC, this could explain why this station clusters with continental polluted stations instead of marine polluted stations in the first cluster analysis of this study in Sect. 5.1.

6.3.1.4 Heselbach, Black Forest, Germany

Back trajectories at FKB group into 2 clusters (Figure 7), each containing approximately half of the back trajectories. Back trajectories associated with Cluster 1 typically originated from the northwest over the North Atlantic and are associated with higher wind speeds and longer distance transport than those in cluster 2. Cluster 2 tended to travel shorter distances in reaching the site, with mean back trajectories originating from the east of the station in Southern Germany, as shown in Figure 7. Despite the very different geographical origins of the two air mass clusters, and very different wind speeds (on average), both trajectory groups have similar median optical property signatures, and suggest a BC dominated aerosol type (Cazorla et al.,

20

2013; Yang et al., 2012; Costabile et al., 2013; Lee et al., 2009). The similarity of aerosol properties between the two trajectory clusters arriving at FKB suggests that FKB measures aerosols that are regionally representative aerosol of Western Europe. Previous analysis of FKB data shows that the site is dominated by anthropogenic aerosol (Jefferson, 2010). Due to the homogeneity of the aerosol population at the FKB site, back trajectory analysis does not provide any additional information useful for aerosol typing.

6.3.2 All stations

A broader understanding of the link between back trajectory clusters, aerosol optical property measurements and aerosol type can be gained by collectively analyzing all trajectory clusters at all stations, rather than looking at stations individually. Here each trajectory cluster from every site is classified based on where the trajectory originated and the geography over which the air mass traveled, then trajectory clusters from *all* stations are plotted in one AAE vs. SAE plot space. The classifications include continental Arctic, continental dust, continental dust/polluted, continental polluted, polluted marine, and remote marine. A trajectory cluster is classified as continental Arctic if it passes over land north of 60N latitude, continental dust if it passes over remote desert, continental dust/polluted if it passes over populated desert regions with anthropogenic influence, continental polluted if it passes over populated land, polluted marine if it passes over populated coastal regions with anthropogenic influence, and remote marine if it passes over clean, unpopulated ocean regions. Table S6 in supplemental materials details classifications of each trajectory cluster at all stations. There is unavoidable subjectivity in this classification method, for a few reasons. For one, some trajectories travel over geography that falls into one or more of the classifications chosen for the analysis. In these cases, other factors, such as underlying geography and typical site aerosol populations, were considered to make the more nuanced classifications. Back trajectory analysis of aerosol type are needed to account for air mass dispersion, aerosol wet and dry deposition, cloud processing, as well as additional sources added at low altitudes and locally. A good example of this is long-range transport of African dust over the Atlantic Ocean. A 3-day back trajectory may not be sufficient to identify long-range dust transport from the African continent. Here, delineation of dust from marine aerosol is ambiguous. More information on the aerosol composition and hygroscopicity is needed for more conclusive aerosol identification. The authors acknowledge this weakness of the methodology and its inherent uncertainty and subjectivity.

Median values of optical properties from each trajectory cluster at all sites are presented in Figure 8. There are some clear spatial patterns that emerge when visualizing the trajectory cluster classifications and the median optical properties in the AAE vs. SAE plot space. The majority of continental polluted trajectory clusters group tightly in the area of the plot that would be classified as BC dominated by the Cappa et al. (2016) matrix. This is similar to earlier findings in this paper where continental polluted sites were aggregated at higher SAE (smaller size) and at AAE values in the range of ~1-1.5. Trajectories classified as polluted marine show a similar range of AAE values as the continental polluted trajectories, though with lower SAE values, indicative of large sea salt mixed with organic carbon. Trajectory clusters classified as continental dust are best defined by AAE values greater than 1.4, though are poorly defined by SAE values due to the large variance in SAE for those clusters. Continental dust/polluted trajectory clusters are more or less tightly defined by AAE values between 0.9-1.4, and SAE values between 0.5-1.2, though it is hard to draw significant conclusions about this trajectory type, since only three trajectories meet this classification. Trajectories identified as continental Arctic are not well defined in this plot space. Both AAE and SAE values of this trajectory type are variable, though median SSA values for this trajectory class are more similar, and are close to 0.95.

21

The range of Arctic optical properties most likely stems from the seasonal transport of European and Siberian continental aerosol to the sites in the winter and spring, contrasted with sea salt from open water in the summer. Remote marine trajectories are the least well defined of all the trajectory cluster classes, with highly variable optical properties. Remote marine trajectories show AAE values that range anywhere from 0-2.2, with SAE values slightly more defined at a range of -0.4-1.2. Median SSA values are, however, quite similar within more remote marine trajectories, with high values near 0.96 indicating a whiter aerosol such as sea salt.

There are some clear outliers within trajectory classification groups that may be explained by misclassification of trajectories. For example, the points labeled 1 and 2 in Figure 8 are back trajectories from CPR, both with 3-day paths that travel only over the Atlantic Ocean. Although the trajectory classification methodology yielded a class of remote marine for those specific trajectories (the air masses only traveled over unpopulated ocean regions for 3 days before reaching the site), previous studies suggest that these air masses could be heavily influenced by African dust events (Denjean et al., 2016; Kalashnikova and Kahn, 2008; Reid et al., 2003). If indeed the dominant aerosol type in these back trajectories was dust, this would fit in much more neatly to previous dust classification schemes (i.e., Lee et al., 2012; Clarke et al., 2007; Yang et al., 2009) and the Cappa et al. (2016) matrix.

By classifying back trajectory clusters from all station locations, and including them in the optical property plot space, we get a clearer idea of what types of trajectories, and thus likely aerosol type, are well defined by median optical properties, and those that are poorly defined by median optical properties. Continental polluted and marine polluted trajectories have median optical parameters that are well defined, and visually cluster in the plot space. Continental dust and continental dust/biomass are somewhat well defined by optical properties in the plot space. Continental Arctic trajectories appear to be well defined by AAE, with all cluster AAE values around 1, though the trajectories are not well defined by SAE, which shows a larger range. The remote marine trajectory cluster (presumably clean air masses) is poorly defined by optical properties and thus is not easily visualized in the plot space.

To our knowledge, few previous studies have classified remote marine aerosol (only Costabile et al. (2013) classified a coarse marine mode in the suburbs of Rome, Italy), and no previous studies have classified continental Arctic aerosols using an aerosol classification matrix. Our findings show that at these site types, typing schemes that use aerosol optical properties need more detailed analysis that account for seasonal variability and local sources. Using aerosol optical parameters to infer aerosol type works well for certain types of aerosol that fit neatly into matrices like that from Cappa et al. (2016), including BC dominated aerosol and dust mixtures. Marine aerosol, processed aerosol, and highly heterogeneous aerosol populations are much more poorly defined by optical properties, and do not fit cleanly in existing matrices without overlap with different aerosol types.

7. Discussion

Application of previous aerosol classification schemes to the aerosol optical property data from stations in the NOAA Federated Aerosol Network generally yields a dominant aerosol type that would be expected at that site location. The classification schemes do particularly well at inferring aerosol type from optical properties at continental sites that measure EC mixtures, but do not do as well at sites with more complex topography (e.g., mountaintop, coastal) that measure a more heterogeneous aerosol population that changes with wind direction. Including median optical parameters from multiple

22

Deleted: Cazorla

Deleted: 2013

Deleted: Cazorla

Deleted: 2013

Deleted: EC/OC

Deleted: mixed

Deleted: 6

Deleted: carbonaceous aerosols

stations on one AAE vs. SAE plot allows for comparison of dominant aerosol type at many sites, though the use of median optical properties makes the most sense for sites with a homogenous aerosol population. The single AAE vs. SAE plot can provide ambiguous results for sites with a heterogeneous aerosol population.

The two aerosol classification methods (Sect. 6.1 and 6.2) had varying degrees of success. The first method, a multivariate cluster analysis, generated groups of monitoring sites with similar AAE, SAE, SSA and $\log(\sigma_{sp})$ values. The first classification scheme was applied to median optical properties from all station data within each cluster to produce a new aerosol type for stations within that cluster. One advantage to this approach is that the inclusion of $\log(\sigma_{sp})$ in the clustering analysis, and subsequent visualization of station clusters in the AAE v. SAE v. $\log(\sigma_{sp})$ 3D parameter space, provides insight not only into a cluster's aerosol type. This approach also provides insight as to how aerosol loading (and thus site conditions) differs between clusters. Although the AAE and SAE aerosol typing schemes yield similar inferred aerosol type of BC dominated aerosol for both remote Arctic/mountaintop sites and continental sites, the notable difference in $\log(\sigma_{sp})$ values among these dissimilar stations defines the separate clusters. An anticipated advantage to the multivariate cluster analysis was that it would help to reduce ambiguity in results of aerosol typing schemes, though this was not the case with every cluster. Rather than falling more surely within the optical property thresholds of one aerosol type, the median optical properties of a few clusters still fell on the cusp of two or more aerosol type thresholds. This left the aerosol type of some clusters uncertain, particularly for clusters with coastal and/or remote sites.

The third method (Sect. 6.3), pairing 6-h averaged optical properties with corresponding back trajectories, provided more detailed insight into the aerosol population at an individual station. This method allowed for typing of multiple aerosols related to different air masses. At stations where aerosol populations are diverse and varying, such as NIM (dust and biomass burning), WLG (dust, pollution, free troposphere long-range transport aerosol), and PVC (marine aerosol and pollution), the different aerosol types that were previously obscured using the site's median optical properties were more apparent when using the trajectory cluster approach. At stations where aerosol populations are homogeneous (like FKB (regional pollution)), no new information on aerosol type was gained. Consolidating all trajectory clusters and corresponding classifications into one plot space (Sect. 6.3.2) allowed us to see a large variety of back trajectory and likely aerosol type, and confirmed previous findings from the paper- that some trajectory classes (like continental polluted and marine polluted) are well defined by a unique range and combination of optical properties, while other trajectory classes (like remote marine and continental Arctic), have highly variable ranges and combinations of SAE, AAE and SSA, and are thus less likely to be typed by aerosol classification schemes using only optical parameters.

The application of varying classification methods gave satisfactory inferences of some aerosol types, in great part due to the quality of previously developed aerosol classification schemes. Despite the differences in optical property thresholds presented from each scheme, many of the schemes' thresholds do have large overlap, making it easy to affirm inferred aerosol type with multiple schemes. Many typing schemes provided satisfactory aerosol typing results for fossil fuel burning aerosol, biomass burning aerosol and dust (Cappa et al., 2016; Cazorla et al., 2013; Lee et al., 2012; Yang et al., 2009; Bahadur et al., 2012; Russell et al., 2010), though fewer schemes were available to type large coated particles (Cazorla et al., 2013), sea salt (Costabile et al., 2013) and mixed aerosol (Cappa et al., 2016; Cazorla et al., 2013). Perhaps the most useful typing schemes were that of Cazorla et al. (2013) and Cappa et al. (2016), which provided thresholds for typing mixed aerosol and large coated particles or large particle/low absorption mix. The Cazorla et a. (2013) and Cappa et al. (2016) schemes also delineated

23

the entirety of the AAE vs. SAE plot space, leaving no combination of optical property values without a category.

It should be mentioned that the success of aerosol classification schemes is largely dependent on uncertainties in AAE attribution (Cappa et al., 2016). The scientific community has yet to fully assess AAE as an indicator of aerosol composition. Although AAE=1 is often taken within the community to indicate black carbon, some studies show that this largely depends on aerosol composition and size, as well as the age of the particle and atmospheric processing that it endures (Lack and Langridge, 2013; Saleh et al., 2014; Costabile et al., 2017; Moosmüller et al., 2011). Furthermore, the accuracy of these aerosol classification methods are only as accurate as the AAE value is an indication of the aerosol composition. As the scientific community advances our understanding of AAE and its relationship to aerosol composition and size, these aerosol classification schemes should be refined.

A major missing piece of the currently available aerosol classification methods is identification and validation of optical property thresholds to identify sea salt aerosol. To the authors' knowledge, only one study includes marine aerosol identification; Costabile et al. (2013) provide values of SSA > 0.95, SAE < 0.5, dSSA=0-0.05 and AAE > 2 for coarse marine mode aerosol. Many studies ignore the contribution of sea salt altogether (or do not use data that would have sea salt aerosol contribution), while other studies do not include sea salt aerosol in their typing scheme because sea salt has negligible absorption and thus poorly defined AAE (Russell et al., 2010). The best match with sea salt aerosol in the Cappa et al. (2016) matrix presented here is likely the 'large particles low absorption' classification. Since sea salt aerosols are dominated by large particles, there is a general consensus that marine particles are characterized by low SAE values and high SSA values (Russell et al., 2010; Costabile et al., 2013; Smirnov et al., 2002; Dubovnik et al., 2002). Of the 24 stations analyzed in this study, sea salt aerosol is expected at CPT, CPR, GRW, PYE, THD, and to a lesser extent at ARN, AMY, GSN and PVC. With the exception of ARN, AMY, GSN and PVC, which often measured polluted air masses (see scattering coefficient values for these four stations in Table 3 and back trajectories for PVC), these coastal stations have median values of SAE < 1 and SSA > 0.95. Median values of AAE, however, range from 0.5-2.0. Further back trajectory analysis (not shown here) relating air masses of oceanic origin at these sites to aerosol optical properties does not show specific patterns in AAE values for marine aerosols. Although no new marine aerosol typing information is included here, the authors do encourage consideration of SAE and SSA thresholds for sea salt to be included in future aerosol classification analyses. Furthermore, the authors acknowledge that although no sea salt aerosol types are designated here explicitly at coastal stations, some of the aerosol types are likely sea salt aerosol mixed (however slightly) with some absorbing component. Cappa et al. (2016) in some ways account for sea salt aerosol by changing the categorization in the lower left of the box to "large particle/lower absorption mix" although in the original matrix they also suggest this regime could be represented by "large black particles".

Although this study generally affirms existing aerosol typing schemes, the results here are only applicable given certain conditions and for specific aerosol types. One stipulation of this analysis is that results were compared to aerosol typing schemes from studies that used optical property data from in situ surface measurements, aircraft campaigns, and AERONET measurements. There are few studies (e.g., Cappa et al., 2016) that evaluate the differences that may exist in aerosol typing schemes/thresholds based on the type of data (in situ vs. remote sensing, column vs. point, dry vs. ambient measurements) used. The difference in RH between dry (most in situ surface) and ambient (AERONET) measurements could have some effect on the determined thresholds. A higher RH would decrease SAE (larger aerosol), SSA thresholds might shift up (whiter aerosol), scattering coefficients would get larger, and AAE might change due to coating on absorbing particles. Future analysis

24

comparing dry and ambient aerosol, as well as surface measured vs. remotely sensed, typing schemes would be useful for determining the validity of the comparisons made in this study.

An additional caveat in the parameter clustering analysis and back trajectory cluster analysis is the presence of externally mixed aerosol with size-dependent composition that renders the analysis ambiguous for a given aerosol class. Future work on this would add much needed information to the subject of aerosol typing from optical properties.

Another limitation to the classification analyses presented here is that aerosol aging during transport can influence aerosol type. A study by Devi et al. (2016) shows that prior to atmospheric aging, mobile sources and biomass burning sources can have relatively high (~1.2-2.0) AAE values; however, after aging during transport (~1-2 days), the brown carbon signal can go away, reducing the AAE value. There may be a point when source information from aerosol intensive optical properties can be lost during transport. In that case, aerosol classification schemes may no longer be applicable.

There are still many ways in which this analysis can be expanded. The incorporation of aerosol shape into the typing analysis could be helpful, particularly in determining the differences between particles with similar optical properties. Further stratification of the measurement data by season, time of day, composition or hygroscopicity would elucidate more about the variability of aerosol type with time. And finally, more analyses of stations that have concurrent chemistry measurements and aerosol optical property measurements could help verify existing aerosol classification schemes (e.g., Cappa et al. (2016) and Costabile et al. (2017)).

8. Conclusion

Surface in situ aerosol optical properties obtained at 24 stations in the NOAA Federated Network were used to classify aerosol type at the site, using aerosol classification schemes from the literature, cluster analyses, and general knowledge of station location and characteristics. The monitoring sites utilized for the analysis offered a diverse range of station locations and aerosol types, providing a look at fossil fuel burning, biomass burning, sea salt, dust, as well as regionally mixed aerosols observed at various continental sites. Plotting station optical property medians in an AAE vs. SAE plot space, overlaid by the Cappa et al. (2016) classification matrix, for the most part yielded inferences of aerosol types that were to be expected based on knowledge of the monitoring station location. A handful of stations, however, yielded unexpected results that appeared uncharacteristic of the site, which indicated a need for a different visualization or analysis method. Furthermore, the interquartile values of the optical properties from each station in an AAE vs. SAE parameter space showed that there is often large variability in optical properties at any given location, suggesting that a single 'dominant aerosol type' is not realistic at all stations.

A multivariate cluster analysis was performed as a means of grouping together monitoring sites with not only similar aerosol type, but similar site conditions (frequency of aerosol type, loadings, proximity to source, location, etc.). The multivariate cluster analysis yielded 6 clusters of stations with similar median AAE, SAE, SSA and $\log(\sigma_{sp})$ values. Sites that grouped within the same cluster most often had similar expected aerosol types that aligned with the aerosol type predicted by the aerosol typing scheme. Incorporation of the scattering coefficient into the multivariate cluster analysis improved the inference of aerosol type and conditions (i.e., aerosol loading, source) from optical property measurements.

25

In order to further explore the complexity of aerosol populations and allow for multiple aerosol types at some sites, an additional analysis was presented using air mass back trajectories. Air mass back trajectories were clustered based on similar direction, altitude and speed, and these clusters were paired with optical property data and plotted in the AAE vs. SAE parameter space. More detailed results from 4 of the 24 stations – WLG, NIM, PVC and FKB – were discussed in order to show the range of success (or lack thereof) of this approach. At complex sites like WLG, NIM, and PVC, multiple dominant aerosol types emerged, unique to different clusters of air mass back trajectories. The classification of numerous aerosol types, along with the information from the back trajectory clusters on how often those aerosol types were measured, allowed for a more complete picture of the heterogeneous aerosol populations at those sites. In the case of FKB, only one aerosol type is inferred in each of the different trajectory clusters, suggesting a homogenous aerosol population that is readily predicted by the simpler analysis of just the median optical properties in the AAE vs. SAE parameter space.

Combining back trajectory clusters and classifications from all 24 sites showed that comparing optical characteristics with trajectory characteristics yields results that further inform aerosol typing schemes. While all trajectory clusters that were classified as marine polluted or continental polluted had optical properties that were well defined, other trajectory clusters classified as continental Arctic or remote marine had highly variable optical parameters that were not informative in aerosol typing.

This study has further assessed existing aerosol typing schemes, provided additional methods that can be implemented to reduce ambiguity in typing schemes, elucidate aerosol conditions that accompany aerosol type, and allow for identification of multiple aerosol types at one site. A major conclusion from the analysis, however, is that there is no combination of extensive and/or intensive optical properties that allow for a perfect classification of aerosol types. Prior knowledge of the measurement site can help inform aerosol classification schemes, but obscurity remains in these techniques. Furthermore, this paper highlighted the need for further analyses and suggests specific ideas for future work needed to progress and refine aerosol typing schemes that infer aerosol type from optical properties. Namely, repeating this analysis with concurrent aerosol chemical and optical measurements to verify aerosol classification thresholds will be essential to expand and improve aerosol classification schemes.

Data availability

Data for AMF sites are available from the DOE/ARM website (http://www.arm.gov). Data from all other sites (except WLG) are available from the World Data Center for Aerosols (http://ebas.nilu.no/). WLG data are available from Junying Sun at CAMS.

26

References

Allan, J. D., Baumgardner, D., Raga, G. B., Mayol-Bracero, O. L., Morales-García, F., García-García, F., Montero-Martínez, G., Borrmann, S., Schneider, J. and Mertes, S.: Clouds and aerosols in Puerto Rico–a new evaluation, Atmospheric Chemistry and Physics, 8, 1293-1309, 2008.

Anderson, T. L. and Ogren, J. A.: Determining aerosol radiative properties using the TSI 3563 integrating nephelometer, Aerosol Science and Technology, 29, 57-69, 1998.

Angelov, C., Angelov, I., Arsov, T., Archangelova, N., Boyukliiski, A., Damianova, A., Drenska, M., Georgiev, K., Kalapov, I. and Nishev, A.: BEO Moussala–A New Facility for Complex Environment Studies, in: Sustainable Development in Mountain Regions, Springer, 123-139, 2011.

Andreae, M. O. and Gelencsér, A.: Black carbon or brown carbon? The nature of light-absorbing carbonaceous aerosol, Atmospheric Chemistry and Physics, 6, 3131-3148, 2006.

Bahadur, R., Praveen, P. S., Xu, Y. and Ramanathan, V.: Solar absorption by elemental and brown carbon determined from spectral observations, Proceedings of the National Academy of Sciences, 109, 17366-17371, 2012.

Bergin, M., Schwartz, S., Halthore, R., Ogren, J. and Hlavka, D.: Comparison of aerosol optical depth inferred from surface measurements with that determined by sun photometry for cloud-free conditions at a continental U. S. site, Journal of geophysical research, 105, 6807-6816, 2000.

Bergstrom, R. W., Pilewskie, P., Russell, P., Redemann, J., Bond, T., Quinn, P. and Sierau, B.: Spectral absorption properties of atmospheric aerosols, Atmospheric Chemistry and Physics, 7, 5937-5943, 2007.

Bergstrom, R. W., Russell, P. B. and Hignett, P.: Wavelength dependence of the absorption of black carbon particles: Predictions and results from the TARFOX experiment and implications for the aerosol single scattering albedo, J. Atmos. Sci., 59, 567-577, 2002.

Bergstrom, R., Schmidt, K., Coddington, O., Pilewskie, P., Guan, H., Livingston, J., Redemann, J. and Russell, P.: Aerosol spectral absorption in the Mexico City area: results from airborne measurements during MILAGRO/INTEX B, Atmospheric Chemistry and Physics, 10, 6333-6343, 2010.

27

Berkowitz, C. M., Jobson, B. T. T., Alexander, M. L., Laskin, A. and Laulainen, N. S. , Aerosol Composition and Morphology during the 2005 Marine Stratus Radiation Aerosol and Drizzle Study, 2005.

Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang: Clouds and Aerosols. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Bodhaine, B. A.: Aerosol absorption measurements at Barrow, Mauna Loa and the south pole, Journal of Geophysical Research: Atmospheres (1984–2012), 100, 8967-8975, 1995.

Bond, T. C., Anderson, T. L. and Campbell, D.: Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols, Aerosol. Sci. Technol., 30, 582-600, 1999.

Borys, R. D. and Wetzel, M. A.: Storm Peak Laboratory: A research, teaching, and service facility for the atmospheric sciences, Bull. Am. Meteorol. Soc., 78, 2115-2123, 1997.

Brunke, E., Labuschagne, C., Parker, B., Scheel, H. and Whittlestone, S.: Baseline air mass selection at Cape Point, South Africa: application of 222 Rn and other filter criteria to CO 2, Atmos. Environ., 38, 5693-5702, 2004.

Cappa, C. D., Kolesar, K. R., Zhang, Z., Atkinson, D. B., Pekour, M.S., Zaveri, R. H., Zelenyuk, A. and Zhang, Q.: Understanding the optical properties of ambient sub and supermicron particulate matter: results from the CARES 2010 field study in northern California, Atmos. Chem. Phys., 16, 6511–6535, 2016.

Cazorla, A., Bahadur, R., Suski, K., Cahill, J. F., Chand, D., Schmid, B., Ramanathan, V. and Prather, K.: Relating aerosol absorption due to soot, organic carbon, and dust to emission sources determined from in-situ chemical measurements, Atmospheric Chemistry and Physics, 13, 9337-9350, 2013.

Che, H., Wang, Y. and Sun, J.: Aerosol optical properties at Mt. Waliguan observatory, China, Atmos. Environ., 45, 6004-6009, 2011.

Clarke, A., Shinozuka, Y., Kapustin, V., Howell, S., Huebert, B., Doherty, S., Anderson, T., Covert, D., Anderson, J. and Hua, X.: Size distributions and mixtures of dust and black carbon aerosol in Asian outflow: Physiochemistry and optical properties, Journal of Geophysical Research: Atmospheres (1984–2012), 109, 2004.

Clarke, A., McNaughton, C., Kapustin, V., Shinozuka, Y., Howell, S., Dibb, J., Zhou, J., Anderson, B., Brekhovskikh, V. and Turner, H.: Biomass burning and pollution aerosol over North America: Organic components and their influence on spectral optical properties and humidification response, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

Colbeck, I.: Aerosol Science: Technology and Applications, John Wiley & Sons, 2013.

Costabile, F., Barnaba, F., Angelini, F. and Gobbi, G.: Identification of key aerosol populations through their size and

composition resolved spectral scattering and absorption, Atmospheric Chemistry and Physics, 13, 2455-2470, 2013.

Costabile, F., Gilardoni, S., Barnaba, F., Di Ianni, A., Di Liberto, L., Dionisi, D., Manigrasso, M., Paglione, M., Poluzzi, V., Rinaldi, M., Facchini, M. C., and Gobbi, G. P.: Characteristics of brown carbon in the urban Po Valley atmosphere, Atmos. Chem. Phys., 17, 313-326, doi:10.5194/acp-17-313-2017, 2017.

5   Delene, D. J. and Ogren, J. A.: Variability of aerosol optical properties at four North American surface monitoring sites, J. Atmos. Sci., 59, 1135-1150, 2002.

Denjean, C., Cassola, F., Mazzino, A., Triquet, S., Chevaillier, S., Grand, N., Bourrianne, T., Momboisse, G., Sellegri, K. and Schwarzenbock, A.: Size distribution and optical properties of mineral dust aerosols transported in the western Mediterranean, Atmospheric Chemistry and Physics, 16, 1081-1104, 2016.

10   Devi, J. J., Bergin, M. H., Mckenzie, M., Schauer, J. J. and Weber, R. J.: Contribution of particulate brown carbon to light absorption in the rural and urban Southeast US, Atmos. Environ., 136, 95-104, 2016.

Draxler, R. R. and Rolph, G.: HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model access via NOAA ARL READY website (http://www.arl.noaa.gov/ready/hysplit4. html). NOAA Air Resources Laboratory, Silver Spring, 2003.

15   Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D. and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590-608, 2002.

Giles, D. M., Holben, B. N., Eck, T. F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R., Thompson, A. and Schafer, J.: An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions, Journal of Geophysical Research: Atmospheres (1984–2012), 117, 2012.

20   Gyawali, M., Arnott, W. P., Zaveri, R. A., Song, C., Moosmüller, H., Liu, L., Mischenko, M., Chen, L., Green, M. and Watson, J.: Photoacoustic optical properties at UV, VIS, and near IR wavelengths for laboratory generated and winter time ambient urban aerosols, Atmospheric Chemistry and Physics, 12, 2587-2601, 2012.

Hagler, G. S., Bergin, M. H., Smith, E. A. and Dibb, J. E.: A summer time series of particulate carbon in the air and snow at Summit, Greenland, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

25   Hansen, J., Rossow, W., Carlson, B., Lacis, A., Travis, L., Del Genio, A., Fung, I., Cairns, B., Mischenko, M. and Sato, M.: Low-cost long-term monitoring of global climate forcings and feedbacks, in: Long-Term Climate Monitoring by the Global Climate Observing System, Springer, 117-141, 1996.

Hansen, J., Sato, M. and Ruedy, R.: Radiative forcing and climate response, Journal of Geophysical Research: Atmospheres, 102, 6831-6864, 1997.

30   Huffman, J.A., Docherty, K.S., Mohr, C., Cubison, M.J., Ulbrich, I.M., Ziemann, P.J., Onasch, T.B., and Jimenez, J.L.: Chemically-resolved volatility measurements of organic aerosol from different sources, Environ. Sci. Technol., 43, 5351-

5357, 2009.

Ion, A., Vermeylen, R., Kourtchev, I., Cafmeyer, J., Chi, X., Gelencsér, A., Maenhaut, W. and Claeys, M.: Polar organic compounds in rural PM 2.5 aerosols from K-puszta, Hungary, during a 2003 summer field campaign: Sources and diel variations, Atmospheric Chemistry and Physics, 5, 1805-1814, 2005.

Jacobson, M. Z.: Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, Nature, 409, 695-697, 2001.

Jefferson, A.: Empirical estimates of CCN from aerosol optical properties at four remote sites, Atmospheric Chemistry and Physics, 10, 6855-6861, 2010.

Kalashnikova, O. V. and Kahn, R. A.: Mineral dust plume evolution over the Atlantic from MISR and MODIS aerosol retrievals, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Kelly, G., Taubman, B., Perry, L., Sherman, J., Soulé, P. and Sheridan, P.: Relationships between aerosols and precipitation in the southern Appalachian Mountains, Int. J. Climatol., 33, 3016-3028, 2013.

Kim, S., Yoon, S., Jefferson, A., Ogren, J. A., Dutton, E. G., Won, J., Ghim, Y. S., Lee, B. and Han, J.: Aerosol optical, chemical and physical properties at Gosan, Korea during Asian dust and pollution episodes in 2001, Atmos. Environ., 39, 39-50, 2005.

Kirchstetter, T. W., Novakov, T. and Hobbs, P. V.: Evidence that the spectral dependence of light absorption by aerosols is affected by organic carbon, Journal of Geophysical Research: Atmospheres (1984–2012), 109, 2004.

Kivekäs, N., Sun, J., Zhan, M., Kerminen, V., Hyvärinen, A., Komppula, M., Viisanen, Y., Hong, N., Zhang, Y. and Kulmala, M.: Long term particle size distribution measurements at Mount Waliguan, a high-altitude site in inland China, Atmospheric Chemistry and Physics, 9, 5461-5474, 2009.

Kotamarthi, V.: Ganges Valley Aerosol Experiment (GVAX) Final Campaign Report, 2013.

Lack, D. A. and Langridge, J.M.: On the attribution of black and brown carbon light absorption using the Angstrom exponent, Atmospheric Chemistry and Physics, 13, 10535-10543, 2013.

Lee, S., Yoon, S., Kim, S., Kim, Y. P., Ghim, Y. S., Kim, J., Kang, C., Kim, Y. J., Chang, L. and Lee, S.: Spectral dependency of light scattering/absorption and hygroscopicity of pollution and dust aerosols in Northeast Asia, Atmos. Environ., 50, 246-254, 2012.

Levy, R. C., Remer, L. A. and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

Liu, J. and Li, Z.: Estimation of cloud condensation nuclei concentration from aerosol optical quantities: influential factors and uncertainties, Atmospheric Chemistry and Physics, 14, 471-483, 2014.

30

McFarlane, S. A., Kassianov, E. I., Barnard, J., Flynn, C. and Ackerman, T. P.: Surface shortwave aerosol radiative forcing during the Atmospheric Radiation Measurement Mobile Facility deployment in Niamey, Niger, Journal of Geophysical Research: Atmospheres, 114, 2009.

Mendes, L., Eleftheriadis, K., and Biskos, G.: Performance comparison of two thermal denuders in volatility tandem DMA measurements, J. Aerosol Sci., 92, 38-52, 2016.

Moosmüller, H., Chakrabarty, R. K., Ehlers, K. M., and Arnott, W. P.: Absorption Angstrom coefficient, brown carbon, and aerosols: basic concepts, bulk matter, and spherical particles, Atmospheric Chemistry and Physics, 11, 1217-1225, 2011.

Ogren, J. A.: Comment on "Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols", Aerosol Science and Technology, 44, 589-591, 2010.

Ogren, J. A.: A systematic approach to in situ observations of aerosol properties, Aerosol forcing of climate, 215-226, 1995.

Oltmans, S. J., Lefohn, A. S., Harris, J. M. and Shadwick, D. S.: Background ozone levels of air entering the west coast of the US and assessment of longer-term changes, Atmos. Environ., 42, 6020-6038, 2008.

Omar, A. H., Won, J., Winker, D. M., Yoon, S., Dubovik, O. and McCormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, Journal of Geophysical Research: Atmospheres (1984–2012), 110, 2005.

Omar, A. H., Won, J., Winker, D. M., Yoon, S., Dubovik, O. and McCormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, Journal of Geophysical Research: Atmospheres (1984–2012), 110, 2005.

Osborne, S., Johnson, B., Haywood, J., Baran, A., Harrison, M. and McConnell, C.: Physical and optical properties of mineral dust aerosol during the Dust and Biomass-burning Experiment, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Park, S., Panicker, A., Lee, D., Jung, W., Jang, S., Jang, M., Kim, D., Kim, Y. and Jeong, H.: Characterization of chemical properties of atmospheric aerosols over Anmyeon (South Korea), a super site under Global Atmosphere Watch, J. Atmos. Chem., 67, 71-86, 2010.

Prospero, J. M., Collard, F., Molinié, J. and Jeannot, A.: Characterizing the annual cycle of African dust transport to the Caribbean Basin and South America and its impact on the environment and air quality, Global Biogeochem. Cycles, 28, 757-773, 2014.

Reid, J. S., Kinney, J. E., Westphal, D. L., Holben, B. N., Welton, E. J., Tsay, S., Eleuterio, D. P., Campbell, J. R., Christopher, S. A. and Colarco, P.: Analysis of measurements of Saharan dust by airborne and ground-based remote sensing methods during the Puerto Rico Dust Experiment (PRIDE), Journal of Geophysical Research: Atmospheres (1984–2012), 108, 2003.

Russell, P., Bergstrom, R., Shinozuka, Y., Clarke, A., DeCarlo, P., Jimenez, J., Livingston, J., Redemann, J., Dubovik, O. and

31

Strawa, A.: Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition, Atmospheric Chemistry and Physics, 10, 1155-1169, 2010.

Russell, P. B., Kacenelenbogen, M., Livingston, J. M., Hasekamp, O. P., Burton, S. P., Schuster, G. L., Johnson, M. S., Knobelspiesse, K. D., Redemann, J. and Ramachandran, S.: A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry, Journal of Geophysical Research: Atmospheres, 119, 9838-9863, 2014.

Saleh, R., Robinson, E.S., Tkacik, D.S., Ahern, A.T., Liu, S., Aiken, A.C., Sullivan, R.C., Presto, A.A., Dubey, M., Yokelson, R.J., Donahue, N.M., and Robinson, A.L.: Brownness of organics in aerosols from biomass burning linked to their black carbon content, Nature Geoscience, 7, 647-650, 2014.

Schuster, G. L., Dubovik, O. and Holben, B. N.: Angstrom exponent and bimodal aerosol size distributions, Journal of Geophysical Research: Atmospheres (1984–2012), 111, 2006.

Sharma, S., Brook, J., Cachier, H., Chow, J., Gaudenzi, A. and Lu, G.: Light absorption and thermal measurements of black carbon in different regions of Canada, Journal of Geophysical Research: Atmospheres (1984–2012), 107, AAC 11-1-AAC 11-11, 2002.

Sheridan, P., Andrews, E., Schmeisser, L., Vasel, B. and Ogren, J.: Aerosol Measurements at South Pole: Climatology and Impact of Local Contamination, Aerosol Air Qual.Res, 16, 855-872, 2016.

Sherman, J. P., Sheridan, P. J., Ogren, J. A., Andrews, E., Hageman, D., Schmeisser, L., Jefferson, A. and Sharma, S.: A multi-year study of lower tropospheric aerosol variability and systematic relationships from four North American regions, Atmospheric Chemistry and Physics, 15, 12487-12517, 2015.

Smirnov, A., Holben, B. N., Kaufman, Y. J., Dubovik, O., Eck, T. F., Slutsker, I., Pietras, C. and Halthore, R. N.: Optical properties of atmospheric aerosol in maritime environments, J. Atmos. Sci., 59, 501-523, 2002.

Titos, G., Jefferson, A., Sheridan, P., Andrews, E., Lyamani, H., Alados-Arboledas, L. and Ogren, J.: Aerosol light-scattering enhancement due to water uptake during TCAP campaign, Atmospheric Chemistry and Physics Discussions, 14, 3361-3393, 2014.

Toledano, C., Cachorro, V., Berjon, A., De Frutos, A., Sorribas, M., De la Morena, B. and Goloub, P.: Aerosol optical depth and Ångström exponent climatology at El Arenosillo AERONET site (Huelva, Spain), Q. J. R. Meteorol. Soc., 133, 795-807, 2007.

Wai, K. M., Lin, N., Wang, S. and Dokiya, Y.: Rainwater chemistry at a high-altitude station, Mt. Lulin, Taiwan: Comparison with a background station, Mt. Fuji, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Yang, M., Howell, S., Zhuang, J. and Huebert, B.: Attribution of aerosol light absorption to black carbon, brown carbon, and dust in China–interpretations of atmospheric measurements during EAST-AIRE, Atmospheric Chemistry and Physics, 9, 2035-2050, 2009.
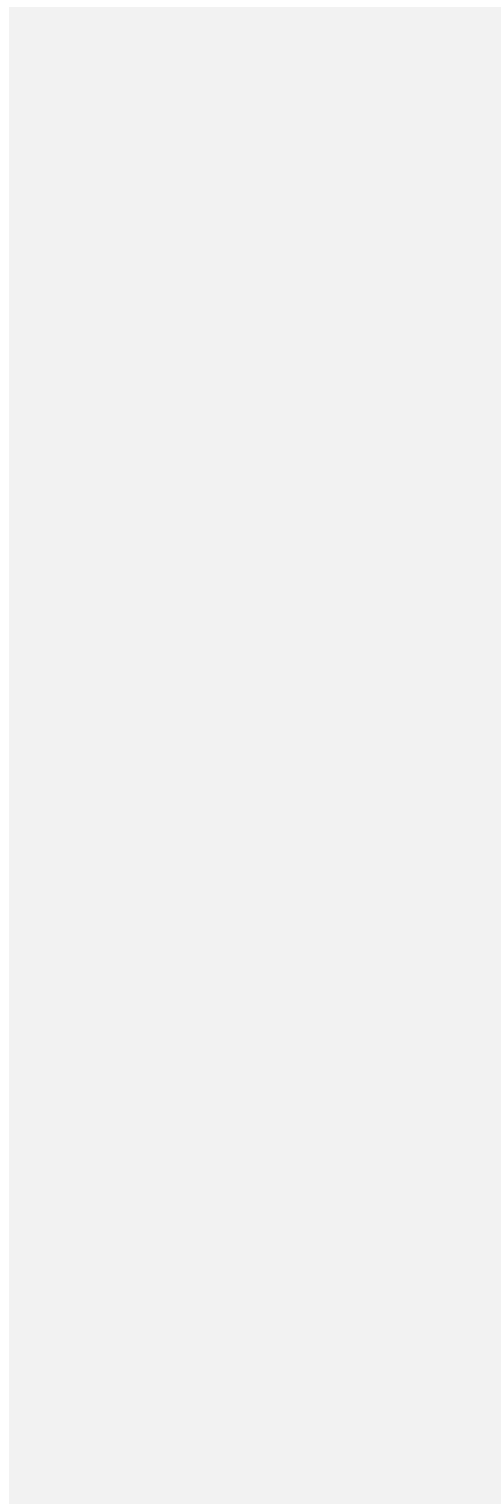
32

33

Table 1 Aerosol optical property thresholds used to determine aerosol type in previous studies. Values in parentheses represent standard deviations, when provided

Study	Measurement Type	Dust	Fossil Fuel Burning	Sea salt	Biomass Burning
Bahadur et al. (2012)	AERONET	$AAE_{440/675nm} \sim 2.2$ (±0.50) $SAE_{440/675nm} < 0.5$	$AAE_{440/675nm} \sim 0.55$ (± 0.24) $SAE_{440/675nm} > 1.2$ *(referred to as BC/EC/soot)*		$AAE_{440/675nm} \sim 4.55$ (± 2.01) $SAE_{440/675nm} > 1.2$ *(referred to as OC)*
Cazorla et al. (2013)	AERONET and aircraft campaign	$AAE_{440/675nm} > 1.5$ $SAE_{440/675nm} < 1$	$AAE_{440/675nm} \leq 1$ $SAE_{440/675nm} > 1.5$ *(referred to as EC dominated)*		$AAE_{440/675nm} \geq 1.5$ $SAE_{440/675nm} > 1.5$ *(referred to as OC dominated)*
Cappa et al. (2016)	Surface in situ	$AAE_{532/660nm} \geq 2$ $SAE_{450/550nm} \leq 0$	$1 < AAE_{532/660nm} \leq 1.5$ $SAE_{450/550nm} \geq 1$ *(referred to as BC dominated)*		$AAE_{532/660nm} \geq 2$ $SAE_{450/550nm} > 1.5$ *(referred to as BrC)*
Russell et al. (2010)	AERONET and aircraft campaign	$AAE = 1.5\text{-}2.5$ $EAE = 0.2\text{-}1$	$AAE = 0.8\text{-}1.5$ $EAE = 1.5\text{-}1.8$		$AAE = 1\text{-}1.7$ $EAE = 1.8\text{-}2$
Clarke et al. (2007)	Aircraft campaign		$AAE_{470/660nm} \sim 1.1$ *(referred to as pollution)*		$AAE_{470/660nm} \sim 2.1$
Costabile et al. (2013)	Surface in situ	$AAE_{467/660nm} \sim 2$ $SAE_{467/660nm} < 0.5$ $SSA_{530nm} > 0.85$ *(referred to as coarse dust mode, CDM)*	$AAE_{467/660nm} < 1.5$ $SAE_{467/660nm} \sim 4$ $SSA_{530nm} < 0.8$ *(referred to as soot mode, STM)*	$AAE_{467/660nm} > 2$ $SAE_{467/660nm} < 0.5$ $SSA_{530nm} > 0.95$ *(referred to as coarse marine mode, CMM)*	$AAE_{467/660nm} < 2$ $SAE_{467/660nmnm} \sim 1\text{-}3$ $SSA_{530nm} < 0.85$ *(referred to as biomass burning smoke mode, BBM)*
Lee et al. (2012)	Surface in situ	$AAE_{450/700nm} \sim 1.2\text{-}1.7$ $SAE_{450/700nm} \sim 0\text{-}1.2$ *(referred to as PD)*	$AAE_{450/700nm} \sim 1\text{-}1.5$ $SAE_{450/700nm} \sim 1.4\text{-}1.8$ *(referred to as P2)*		$AAE_{450/700nm} \sim 0.8\text{-}1.4$ $SAE_{450/700nm} \sim 0.8\text{-}1.5$ *(referred to as P1, higher in OC than P2)*
Yang et al. (2009)	Surface in situ	$AAE_{370/950nm} \sim 1.82$ (±0.90) $SAE_{450/700nm} \sim 0.59$ (± 0.41) $SSA_{550nm} \sim 0.9\ 0.8$ (± 0.04)	$AAE_{370/950nm} \sim 1.46$ (± 0.15) $SAE_{450/700nm} \sim 1.39$ (± 0.20) $SSA_{550nm} \sim 0.8$ (± 0.05) *(referred to as coal pollution)*		$AAE_{370/950nm} \sim 1.49$ (± 0.08) $SAE_{450/700nm} \sim 1.52$ (± 0.18) $SSA_{550nm} \sim 0.89$ (± 0.01)

5

Table 2 Monitoring Site Locations and Descriptions. Rows in bold indicate stations that are part of the ARM Mobile Facility (AMF) program and are temporary measurement sites.

Station Abbreviation	Station Location	Latitude Longitude Altitude (m asl)	Absorption Instrument*	Measurement Dates	Site Classification	Site Description (and references)
ALT	Alert, Canada	+82.45 -62.52 210	PSAP-3W	2012-2013	Arctic	Remote Arctic site, situated away from major anthropogenic and industrial areas, and the most northerly site in the network. (Sharma et al., 2002)
AMY	Anmyeon-do, South Korea	+36.54 +126.33 45	CLAP-3W	2012-2013	Polluted Marine	Polluted marine site that receives both continental and marine air masses, located on Anmyeon Island off the coast of South Korea (Park et al., 2010)
APP	Boone, North Carolina, USA	+36.2 -81.7 1100	PSAP-3W	2012-2013	Continental Polluted	Semi-rural continental site, located in the Appalachian Mountains, a region high in biogenically-derived aerosol (Sherman et al., 2015)
ARN	El Arenosillo, Spain	+37.10 -6.73 41	CLAP-3W	2012-MAY-15 to 2013	Marine Polluted	Located near the Atlantic Ocean and Huelva City. Site is located in protected coastal area of Doñana National Park and experiences episodes of desert dust and pollution (Toledano et al., 2007)
BEO	BEO-Moussala, Bulgaria	+42.18 +23.59 2925	CLAP-3W	2012-JUN-03 to 2013	Continental Polluted, Mountaintop	The Basic Environmental Observatory (BEO) sits atop Moussala Peak, the tallest point on the Balkan Peninsula. Given the site's altitude, it is considered to be in the free troposphere and more or less unperturbed by regional pollution sources (Angelov et al., 2011)
BND	Bondville, Illinois, USA	+40.05 -88.37 230	CLAP-3W	2012-2013	Continental Polluted	Anthropogenically influenced rural site located in Champaign County, Illinois, USA, near soy and corn farms south of Bondville (Delene and Ogren, 2002; Sherman et al., 2015)
BRW	Barrow, Alaska, USA	+71.32 -156.6 11	CLAP-3W	2012-2013	Arctic	Coastal Arctic site 3 km from Arctic Ocean, located north of the Arctic Circle near the small town of Barrow. Though the site is remote, drilling activities nearby may influence aerosol populations (Bodhaine, 1995)
CPR	Cape San Juan, Puerto Rico	+18.48 -66.13 17	CLAP-3W	2012-MAR-30 to 2013	Marine Polluted	Marine site, located on the northeast edge of the Caribbean island of Puerto Rico on Las Cabezas de San Juan nature reserve. Prone to African desert dust episodes (Allan et al., 2008)
CPT	Cape Point, South Africa	-34.35 +18.49 230	PSAP-3W	2010-2011**	Marine Clean	Marine site, located on the southwest tip of South Africa. Site is influenced by remote marine air and polluted and/or dusty continental air (Brunke et al., 2004)
FKB	**Heselbach, Germany**	+48.54 +8.40 511	PSAP-3W	2007-MAR-23 to 2007-DEC-31	Continental Polluted	Continental site in the Black Forest region of Germany surrounded by coniferous trees. The site is in the agricultural Murg valley, and experiences heavy precipitation and influence from anthropogenic industrial

Formatted Table

35

						activities (Jefferson, 2010)
GRW	**Graciosa Island, Azores, Portugal**	+39.09 -28.03 15.24	PSAP-3W	2009-APR-18 to 2010-DEC-31	Marine Clean	Marine site located on the remote Azores Islands surrounded by the Atlantic Ocean. Site may be influenced at times by local pollution and African desert dust episodes (Jefferson, 2010)
GSN	Gosan, Jeju Island, South Korea	+33.28 +126.17 72	CLAP-3W	2012-2013	Marine Polluted	Coastal site located on the western edge of Jeju Island, and prone to influence from marine aerosols, anthropogenic pollution, and long-range Asian desert dust (Kim et al., 2005)
KPS	K-puszta, Hungary	+46.96 +19.58 125	CLAP-3W	2012-2013	Continental Polluted	Continental site located in the Hungarian Great Plain 70 km southeast of Budapest. Measures regional background air, and although it is situated as remotely as possible, is still influenced by biomass burning aerosol from home heating in the winter (Ion et al., 2005)
LLN	Lulin, Taiwan	+23.47 +120.87 2862	CLAP-3W	2012-2013	Continental Polluted, Mountaintop	High altitude site influenced by air masses from polluted biomass and industrial continental Asian sources, as well as clean marine regions (Wai et al., 2008)
MLO	Mauna Loa, Hawaii, USA	+19.54 -155.58 3397	CLAP-3W PSAP-3W	2012-2013	Marine Polluted, Mountaintop	High altitude site on the northern side of the Mauna Loa volcano on the big island of Hawaii. Distinct diurnal patterns in upslope/downslope air flow, with minimal influence from regional aerosol sources (Bodhaine, 1995)
NIM	**Niamey, Niger**	+13.48 +2.18 205	PSAP-3W	2005-DEC-01 to 2006-DEC-31	Continental Dust/Biomass	Continental site susceptible to biomass burning and African desert dust, prone to high heat and heavy rains in the monsoon season (Liu and Li, 2014)
PGH	**Nainital, India**	+29.36 +79.46 1951	CLAP-3W	2011-JUN-09 to 2012-MAR-27	Continental Dust/Biomass	Continental site located in the Ganges Valley in the remote foothills of the Himalayas. Biomass burning, dust, and growth in nearby industrial activities, sporadically influence the site (Liu and Li, 2014)
PVC	**Cape Cod, Massachusetts, USA**	+42.07 -70.20 1	CLAP-3W	2012-JUL-16 to 2013-JUN-24	Marine Polluted	Marine site on a peninsula of Massachusetts reaching into the Atlantic Ocean. Site is also near major urban areas, including Boston, Massachusetts and Providence, Rhode Island, and is thus influenced by both polluted and clean air masses (Titos et al., 2014)
PYE	**Pt. Reyes, California, USA**	+38.09 -122.96 5	PSAP-3W	2005-MAR-21 to 2005-SEP-15	Marine Clean	Marine site on the California coast north of San Francisco. Air masses from the west are strictly maritime, while air masses from the north, south, and east are influenced by continental pollution (Berkowitz et al., 2005)
SGP	Southern Great Plains, Oklahoma, USA	+36.61 -97.49 315	CLAP-3W	2012-2013	Continental Polluted	Rural continental site located near wheat fields and cattle pastures southeast of Lamont, Oklahoma. There are no large urban areas nearby, but point sources, like power plants and oil operations, influence the site occasionally (Delene and Ogren, 2002; Sherman et al., 2015)
SPL	Storm Peak,	+40.45	CLAP-3W	2012-2013	Continental Polluted,	High altitude forested site in the Rocky Mountains of northwestern Colorado.

		Colorado, USA	-106.73 3220			Mountaintop	Located near the town of Steamboat Springs and agricultural Yampa Valley, though the station frequently measures uncontaminated free troposphere. (Borys & Wetzel, 1997)
SUM	Summit, Greenland	+72.58 -38.48 3238	CLAP-3W	2012-2013	Arctic, Mountaintop	Arctic station atop the Greenland Ice Sheet. Remote and clean, with occasional influence from long-range biomass and industrial pollution (Hagler et al., 2007)
THD	Trinidad Head, California, USA	+41.05 -124.15 107	CLAP-3W	2012-2013	Marine Clean	Marine site on the northern California coast, with Pacific Ocean to the west and redwood forests to the east. Though maritime airflow is predominant, some anthropogenic influences from other airflows is observed (Oltmans et al., 2008)
WLG	Mt. Waliguan, China	+36.28 +100.90 3816	PSAP-3W	2012-2013	Continental Dust/Biomass, Mountaintop	High altitude station located on the dry, arid Tibet plateau in China. The site experiences clean or dusty air masses coming in from the west, and anthropogenically influenced and polluted air masses coming from the east (Kivekäs et al., 2009; Che et al., 2011)

*All scattering instruments are TSI nephelometers

**Cape Point (CPT) had data loss issues in the 2012-2013 time period, so the period 2010-2011 was used instead

5

Table 3 Number of hourly data points, plus median values and lower and upper quartiles for Scattering Ångström exponent and Absorption Ångström Exponent, Single Scattering Albedo, scattering coefficient (σ_{sp}), absorption coefficient (σ_{ap}) and inferred aerosol type at each monitoring station. All data are filtered by thresholds $\sigma_{sp} > 1$ Mm^{-1} and $\sigma_{ap} > 0.5$ Mm^{-1}

Station	# data points	SAE (lq, uq)	AAE (lq, uq)	SSA (lq, uq)	σ_{sp} (Mm^{-1}) (lq, uq)	σ_{ap} (Mm^{-1}) (lq, uq)	Aerosol type based on Cappa et al. (2016) scheme	Aerosol type based on clustering of aerosol optical properties
ALT	1648	1.27 (1.05, 1.43)	0.86 (0.79, 0.95)	0.93 (0.92, 0.94)	9.69 (8.16, 12.11)	0.75 (0.63, 0.90)	Large particles low absorption	Small particles low absorption + BC dominated
AMY	8914	1.57 (1.36, 1.75)	1.22 (0.94, 1.42)	0.92 (0.90, 0.95)	107.72 (61.81, 189.54)	8.72 (5.53, 13.44)	BC dominated	BC dominated
APP	15547	2.11 (1.94, 2.26)	1.20 (0.87, 1.48)	0.92 (0.89, 0.94)	24.46 (14.59, 38.17)	2.13 (1.38, 3.19)	BC dominated	BC dominated
ARN	8237	1.37 (0.97, 1.70)	1.32 (1.16, 1.50)	0.89 (0.85, 0.92)	26.10 (16.7, 40.73)	3.15 (1.83, 5.04)	BC dominated	BC dominated
BEO	5775	1.87 (1.44, 2.07)	1.31 (1.05, 1.55)	0.92 (0.90, 0.94)	22.64 (11.52, 40.04)	1.94 (1.07, 3.21)	BC dominated	BC dominated
BND	15257	2.01 (1.84, 2.17)	1.15 (0.93, 1.34)	0.93 (0.89, 0.95)	33.06 (19.90, 55.14)	2.69 (1.58, 4.17)	BC dominated	BC dominated
BRW	2612	1.17 (0.78, 1.52)	0.99 (0.89, 1.10)	0.93 (0.90, 0.96)	10.47 (7.87, 15.97)	0.73 (0.60, 1.00)	Small particles low absorption	Small particles low absorption + BC dominated
CPR	5744	0.28 (0.17, 0.54)	2.00 (1.16, 2.65)	0.97 (0.96, 0.98)	35.32 (24.33, 50.22)	1.01 (0.71, 1.5)	Mixed dust/BC/BrC	Mixed Dust/BC/BrC
CPT	3158	0.67 (0.34, 1.14)	1.12 (0.97, 1.31)	0.96 (0.94, 0.97)	21.31 (13.76, 29.79)	1.14 (0.73, 2.45)	Large particle/BC mix	Large particle/BC Mix
FKB	5543	1.80 (1.59, 1.95)	1.07 (0.98, 1.16)	0.85 (0.79, 0.88)	32.37 (18.12, 57.77)	5.75 (3.17, 9.96)	BC dominated	BC dominated
GRW	7960	-0.12 (-0.34, 0.19)	0.62 (0.31, 0.85)	0.97 (0.95, 0.98)	30.73 (19.37, 47.42)	0.84 (0.64, 1.29)	Large particles low absorption	Large particle/BC Mix
GSN	10731	1.51 (1.29, 1.70)	1.21 (1.03, 1.34)	0.93 (0.92, 0.95)	61.85 (37.92, 106.47)	4.59 (2.70, 7.40)	BC dominated	BC dominated
KPS	8923	2.06 (1.90, 2.19)	1.39 (1.24, 1.60)	0.88 (0.85, 0.90)	45.11 (25.27, 90.90)	6.27 (3.61, 12.02)	BC dominated	BC dominated

Site	N							
LLN	8294	1.94 (1.82, 2.08)	1.11 (0.97, 1.25)	0.91 (0.88, 0.93)	24.02 (11.81, 40.00)	2.39 (1.20, 4.56)	BC dominated	BC dominated
MLO	2351	1.40 (0.85, 1.76)	1.42 (1.08, 1.89)	0.92 (0.85, 0.95)	9.38 (4.88, 18.39)	0.85 (0.64, 1.19)	Large particle/BC Mix	Small particles low absorption + BC dominated
NIM	4527	0.32 (0.14, 0.64)	1.66 (1.46, 1.22)	0.91 (0.86, 0.94)	91.02 (50.67, 185.24)	9.25 (5.68, 16.05)	Mixed Dust/BC/BrC	Large particle/BC mix
PGH	4079	0.75 (0.53, 0.92)	1.03 (0.88, 1.22)	0.94 (0.92, 0.95)	126.31 (66.48, 232.01)	8.14 (4.52, 126.31)	Large particles/BC mix	Large particle/BC mix
PVC	4990	2.15 (1.64, 2.50)	0.99 (0.68, 1.25)	0.93 (0.90, 0.95)	16.08 (10.19, 27.87)	1.10 (0.75, 1.82)	Small particles low absorption	BC dominated
PYE	481	0.98 (0.53, 1.29)	0.50 (0.30, 1.52)	0.98 (0.97, 0.99)	40.00 (26.59, 59.97)	0.69 (0.58, 1.00)	Large particles low absorption	Large particle/BC Mix
SGP	14610	1.77 (1.43, 2.06)	1.30 (1.05, 1.51)	0.92 (0.89, 0.94)	26.75 (16.06, 42.27)	2.31 (1.41, 3.42)	BC dominated	BC dominated
SPL	8509	1.69 (1.24, 2.03)	1.37 (1.22, 1.51)	0.92 (0.90, 0.94)	11.50 (7.79, 17.70)	0.93 (0.69, 1.35)	BC dominated	Small particles low absorption + BC dominated
SUM	462	1.93 (1.62, 2.07)	1.04 (0.93, 1.16)	0.93 (0.91, 0.95)	8.06 (6.27, 11.58)	0.64 (0.55, 0.81)	BC dominated	Small particles low absorption + BC dominated
THD	5283	0.96 (0.62, 1.43)	1.43 (1.14, 1.70)	0.95 (0.93, 0.97)	21.51 (13.09, 34.56)	0.94 (0.68, 1.4)	Large particle/BC mix	Large particle/BC Mix
WLG	6494	1.10 (0.72, 1.35)	1.37 (1.22, 1.54)	0.93 (0.92, 0.95)	42.19 (20.08, 101.06)	3.01 (1.67, 6.16)	Large particle/BC mix	Large particle/BC Mix

Table 4. Median AAE, SAE, SSA, and log(σ_sp) values (along with corresponding interquartile spread) for each cluster resulting from the cluster analysis.

Cluster #	AAE	SAE	SSA	log(σ_sp)	Sites included in cluster	Aerosol type according to Cappa et al. (2016) matrix	Cluster commonality/ Site descriptions
1	1.04 (0.99, 1.37)	1.40 (1.27, 1.69)	0.93 (0.92, 0.93)	2.27 (2.23, 2.34)	ALT, BRW, MLO, SPL, SUM	Small particles low absorption + BC dominated	Remote Arctic or mountaintop with long-range transport aerosol or occasional local influence
2	1.22 (1.21, 1.22)	1.54 (1.53, 1.55)	0.93 (0.92, 0.93)	4.44 (4.29, 4.57)	AMY, GSN	BC dominated	Heavily polluted South Korean coastal sites
3	1.20 (1.11, 1.31)	1.94 (1.80, 2.06)	0.92 (0.89, 0.92)	3.26 (3.18,3.48)	APP, ARN, BEO, BND, FKB, KPS, LLN, PVC, SGP	BC dominated	Primarily continental sites experiencing urban or biomass burning aerosol
4	1.34 (1.19, 1.50)	0.53 (0.43, 0.64)	0.92 (0.92, 0.93)	4.67 (4.59, 4.76)	NIM, PGH	Large particle/BC mix	Continental sites experiencing heavy dust loading and biomass burning aerosol
5	2.00	0.28	0.97	3.56	CPR	Mixed Dust/BC/BrC	Coastal site experiencing occasional dust, biomass burning or pollution
6	1.12 (0.62, 1.37)	0.96 (0.67, 0.98)	0.95 (0.94, 0.97)	3.43 (3.07, 3.69)	CPT, GRW, PYE, THD, WLG	Large particle/BC Mix	Coastal or remote sites experiencing occasional sea salt, dust, biomass burning or pollution aerosol

5

40

Deleted: log

Formatted Table

Deleted: Cazorla

Deleted: 2013

Deleted: Large coated

Deleted: + EC/OC

Deleted: EC/OC

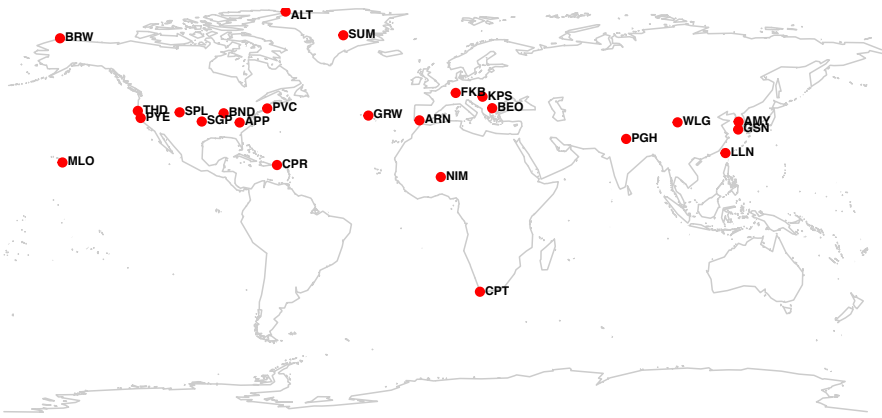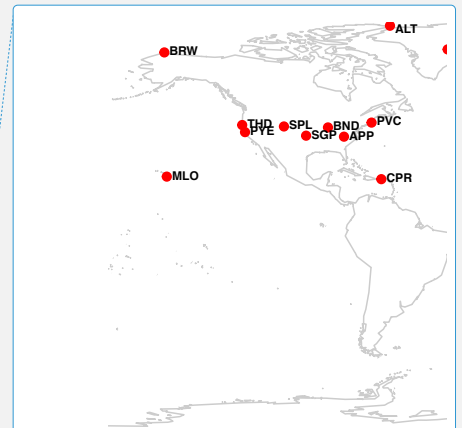**Deleted:** EC/OC

Deleted: Dust

Figure 1 Map of 24 in situ monitoring stations within the NOAA/ESRL Federated Aerosol Network that were utilized in this study. Locations are labeled with each site's 3-letter station abbreviation.
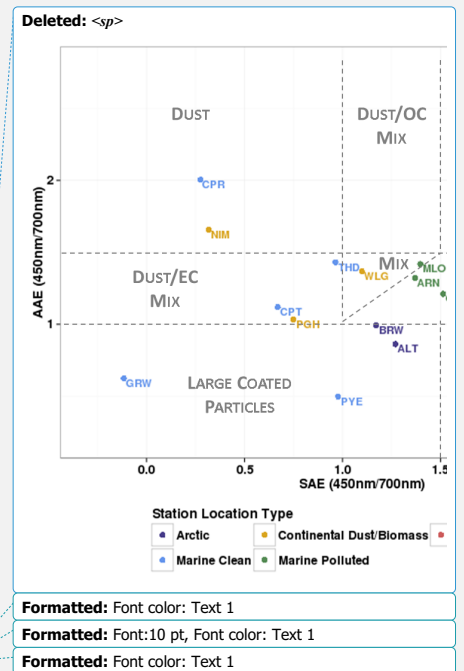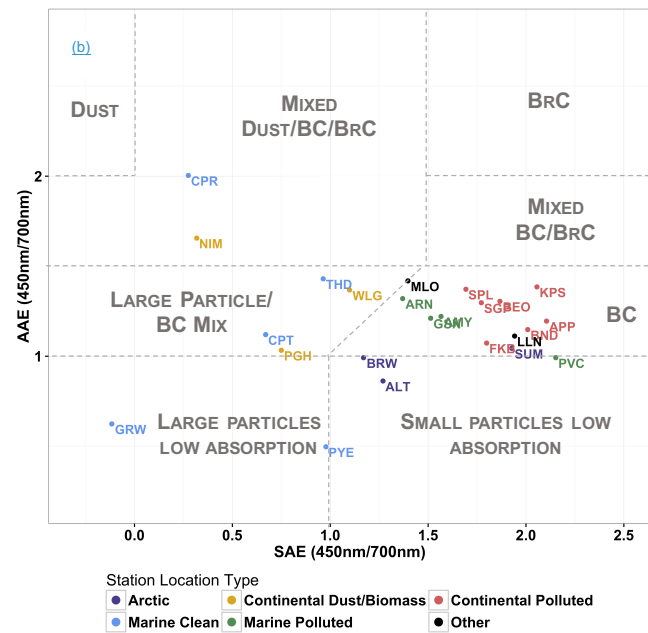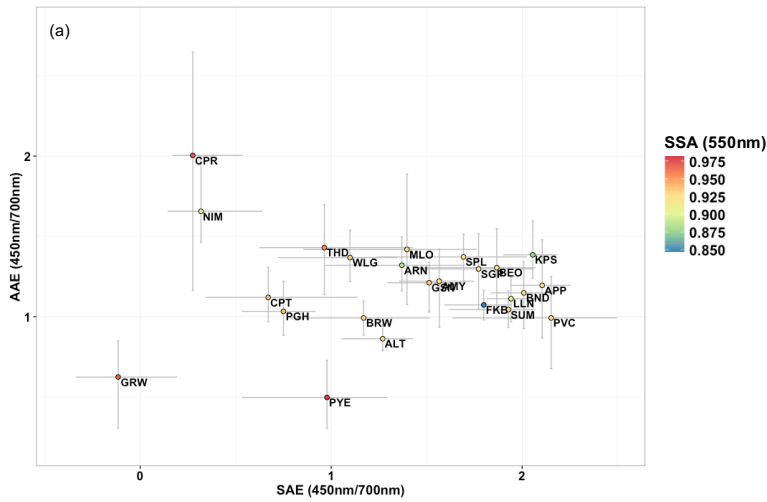
41

Figure 2. AAE vs. SAE medians plotted for 24 in situ monitoring stations in the NOAA/ESRL federated network. (a) Bars represent interquartile values, and points are color coded by median SSA value at the station, (b) Points are color coded by station location type, and plot is overlaid with aerosol classification matrix from Cazorla et al. (2013)

5
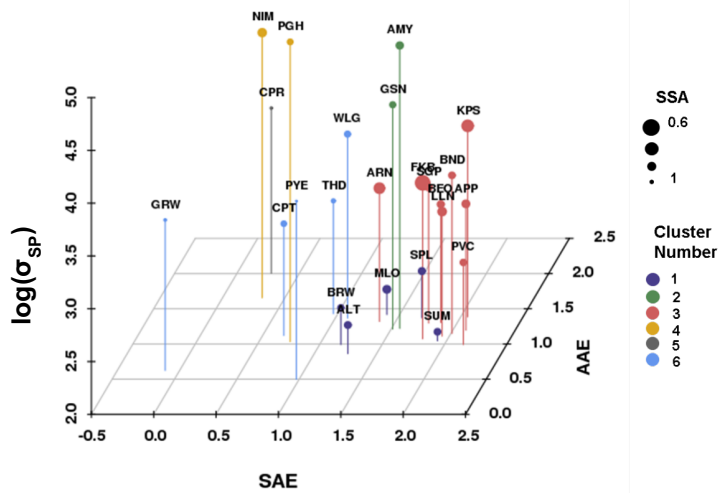
Figure 3. 3D parameter space of SAE vs. AAE vs. log of scattering coefficient, σ_{sp} (Mm^{-1}). Station points are colored by cluster number resulting from the clustering analysis, and sized by median SSA value.
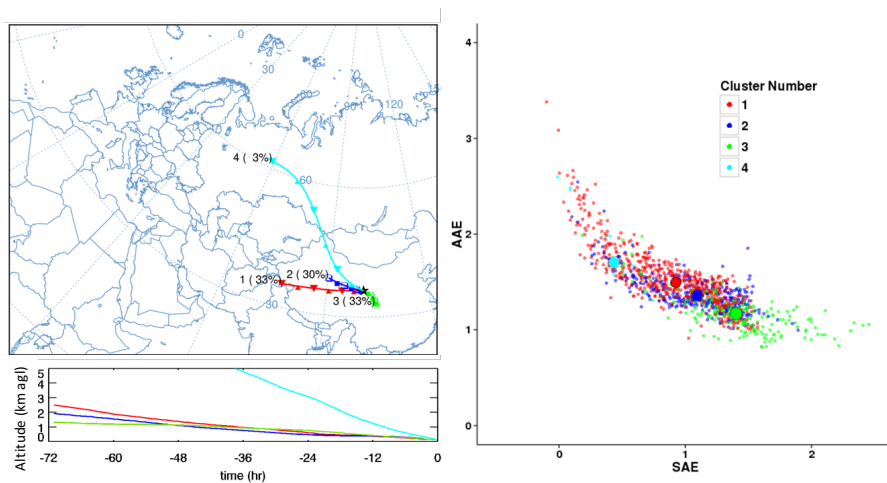

Figure 4. Back trajectory map, back trajectory height (in kilometers above ground level) vs. time, and AAE vs. SAE plot space for Mt. Waliguan, China station WLG, all color coded by back trajectory cluster number. The percentage of air-mass back trajectories corresponding to each cluster are also shown next to the mean trajectories.
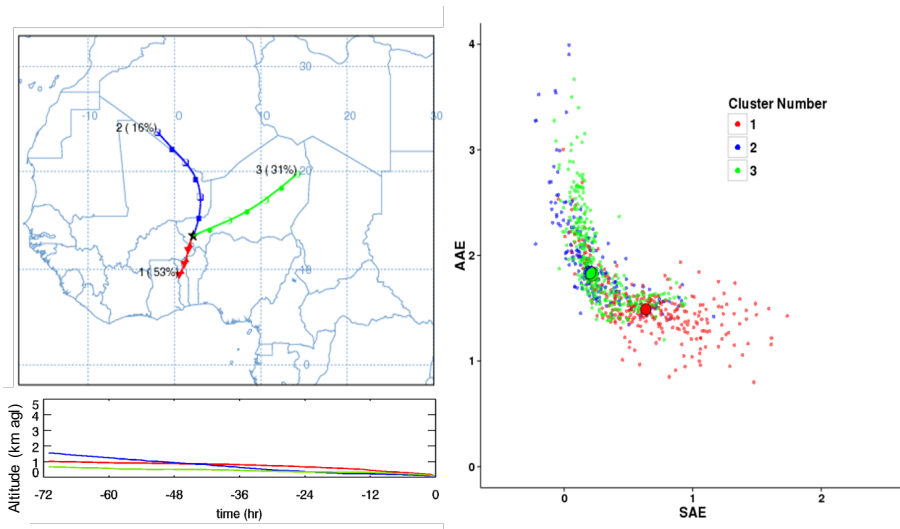
43

Figure 5. Back trajectory map, back trajectory height (in kilometers above ground level) vs. time, and AAE vs. SAE plot space for Niamey, Niger station NIM, all color coded by back trajectory cluster number
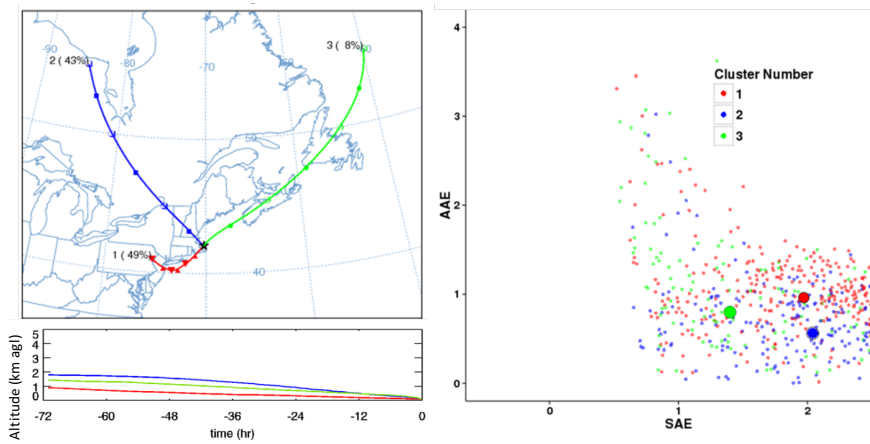

Figure 6. Back trajectory map, back trajectory height (in kilometers above ground level) vs. time, and AAE vs. SAE plot space for Cape Cod, Massachusetts, USA station PVC, all color coded by back trajectory cluster number
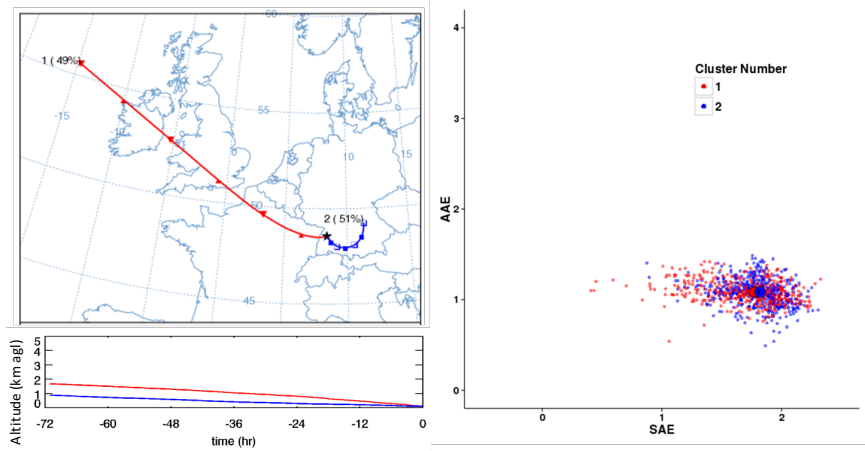
44

Figure 7. Back trajectory map, back trajectory height (in kilometers above ground level) vs. time, and AAE vs. SAE plot space for Black Forest, Germany station FKB, all color coded by back trajectory cluster number
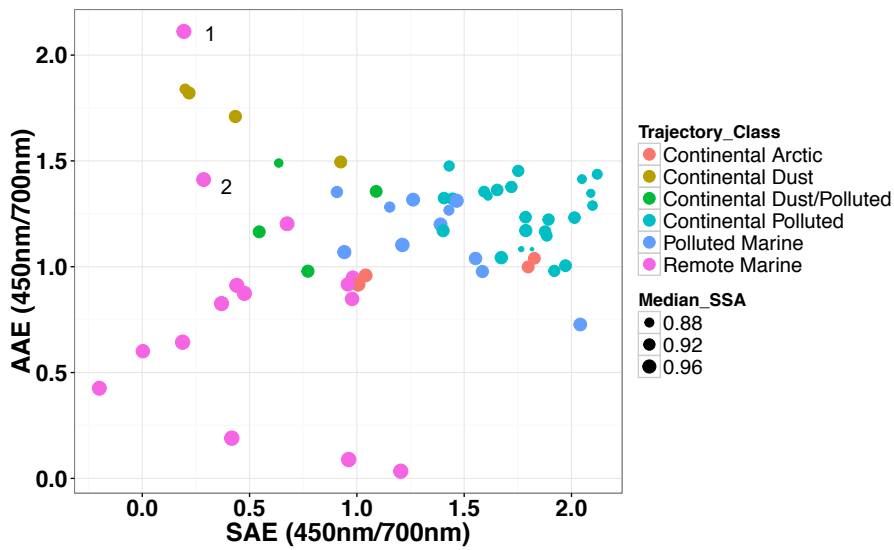

Figure 8. AAE vs. SAE medians plotted for all back trajectory clusters from 24 in situ monitoring stations in the NOAA/ESRL federated network. Points are colored by the trajectory classification, and sized by the median SSA value of measurements from that trajectory cluster, such that smaller points indicate low SSA values and larger point indicate high SSA value. The points labeled 1 and 2 are back trajectories from CPR that are outliers discussed in the text in Sect. 6.
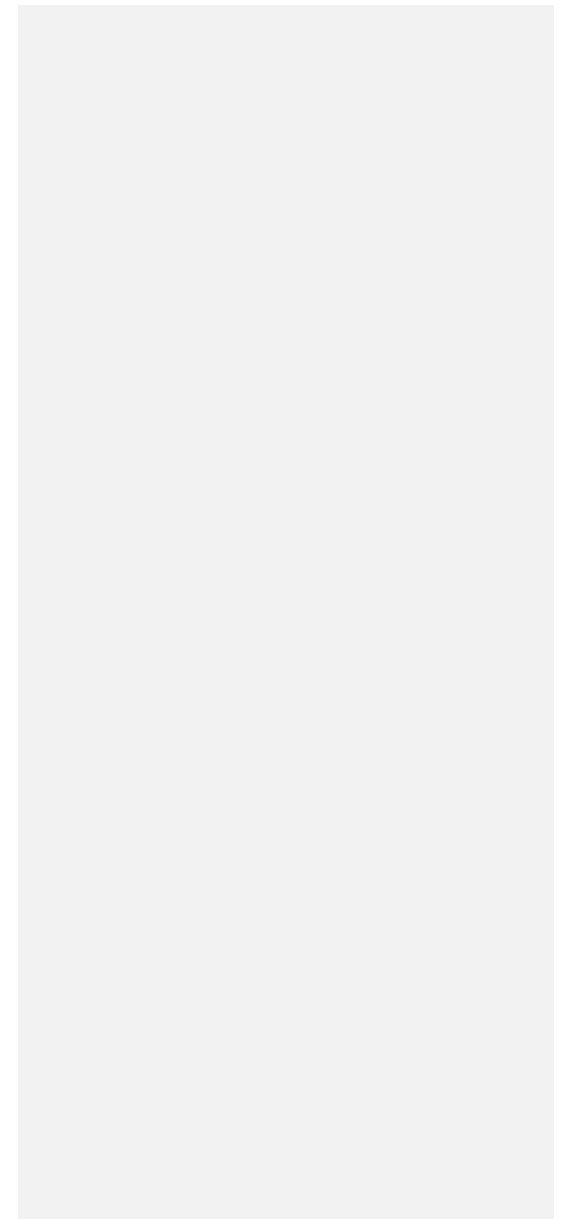
45

46

In order to infer a more accurate representation of aerosol type using intensive optical properties as an indication of aerosol size/composition and extensive optical properties as an indication of loading, a multivariate clustering analysis is performed.

A cluster analysis is the process of statistical grouping that yields 'clusters' with similar characteristics. A few other studies also implement multi-dimensional clustering as a means of solidifying aerosol property thresholds for different aerosol types (Russell et al., 2010; Omar et al., 2005; Levy et al., 2007). In this study, a cluster analysis is used to determine groups of stations with similar aerosol type based on aerosol optical properties. The clusters are then plotted in a 3D parameter space (AAE vs. SAE vs. $\log(\sigma_{sp})$) as a means of visualizing any spatial patterns that emerge.

The k-means clustering algorithm was run using medians of four aerosol optical property parameters – SAE, AAE, SSA, and the log of the scattering coefficient ($\log(\sigma_{sp})$) – from hourly averaged records at each monitoring station. The scattering coefficient, σ_{sp}, is an indication of aerosol loading and is implemented here as an additional parameter to improve the inference of aerosol types. The log of σ_{sp} (in Mm^{-1}) is used rather than the raw σ_{sp} median in order to make the scattering coefficient values more comparable with the magnitude of the optical property values, so the clustering is not dominated by one parameter. While the magnitude of loading (σ_{sp}) alone does not correspond to a specific aerosol type (for example, high loadings can be observed for dust, pollution, or biomass burning events) it may act as a secondary indicator of aerosol conditions (i.e., frequency of aerosol type occurrence, loading) and source contributions, so it is included in the clustering analysis.

To run the clustering algorithm, a number of clusters 'k' is selected. Choosing the 'k' initial seed points is inherently subjective – in this analysis, k needs to be small enough such that the number of stations that fall into each cluster makes for a meaningful grouping, and large enough such that a distinction between station groups is apparent. The algorithm then takes 'k' initial seed points at random and iteratively assigns each point to the nearest cluster centroid taking into account the clustering properties. The next iteration chooses 'k' new seed points and repeats the process until the algorithm converges. In this study, six clusters are selected, creating six unique groups each with similar SAE, AAE, SSA, and $\log(\sigma_{sp})$ characteristics. Each monitoring station was assigned to one of the six clusters produced from the algorithm, and the groupings were used to further analyze aerosol type and conditions.

Given the location of these sites within a highly polluted region and occasional local biomass burning, the inferred aerosol types make sense. The relatively low SAE compared to the sites in Cluster 3 can be attributed to the

contribution of sea salt aerosol.

Cluster 3 is comprised of ARN, SGP, FKB, BEO, LLN, PVC, APP, BND and KPS- all of which are continental polluted stations, with the exception of PVC and ARN, which are marine polluted sites characterized by both oceanic air masses and continental polluted air masses. The stations in Cluster 3 have the highest SAE values of all the sites (SAE~2) and thus the smallest particles, and meet the optical property thresholds for fossil fuel and biomass burning aerosols observed in previous studies (Cappa et al., 2016; Cazorla et al., 2013; Clarke et al., 2007; Lee et al., 2012). What separates Cluster 2 from Cluster 3 is the slightly higher SSA value for sites in Cluster 2, which may indicate a stronger sea salt aerosol signal at the Cluster 2 marine polluted sites that is not present at most sites in Cluster 3 (except ARN). Additionally, aerosol loading at sites in Cluster 2 is substantially higher (factor of 2-3) than at the sites in Cluster 3.

Cluster 4 contains NIM and PGH, both characterized by very high aerosol loadings (high σ_{sp}), low SAE (SAE~0.5), indicating large particle, and AAE~1.3. The median optical signature of the aerosol at these monitoring stations is representative of dust aerosol (Costabile et al., 2013; Cazorla et al., 2013; Lee et al., 2012; Yang et al., 2009; Russell et al., 2010; Bahadur et al., 2012) potentially mixed with a black and/or brown carbon (Cappa et al., 2016). Previous studies report that high loadings of dust aerosol are found at both PGH (Kotamarthi, 2013) and NIM (Osborne et al., 2008), validating the inferred aerosol type of this cluster. NIM also experiences some biomass burning aerosol (Osborne et al., 2008; MacFarlane et al., 2009). PGH experiences fresh biofuel burning daily as well as seasonal pollen and large bioaerosol from the surrounding forest.

Cluster 5 is comprised of just one monitoring site, CPR. Although it is typically undesirable to have a cluster with only one member, the algorithm placed CPR in its own cluster for many choices of 'k' clusters, indicating that indeed it is unique enough to be in its own cluster. Cluster 5 doesn't have interquartile spread on the values since it only has one station. Cluster 5 (CPR) is characterized by the highest AAE value (AAE median = 2.00), the smallest SAE value (SAE median = 0.28), and a high particle loading. Both marine aerosols and African dust event aerosols have been measured at Puerto Rican sites (e.g., Denjean et al., 2016; Propsero et al., 2014; Kalashnikova and Kahn, 2008; Reid et al., 2003), as well as occasional anthropogenic aerosol (Allan et al., 2008) biomass burning aerosol and volcanic aerosol. Based on median optical properties, the CPR aerosol falls into the dust regime suggested by previous classification schemes (Costabile et al., 2013; Cazorla et al., 2013; Lee et al., 2012; Yang et al., 2009; Russell et al., 2010; Bahadur et al., 2012), though knowledge of the CPR site suggests the station would also measure sea salt and occasional biomass burning aerosol.

Monitoring stations CPT, GRW, PYE, THD and WLG make up Cluster 6. The cluster is defined by high SSA values (SSA~0.95), low σ_{sp} and moderate SAE (SAE~0.96) and AAE (AAE~1.12) values. It is worth noting that the AAE values at stations in this cluster have a large spread, and GRW only fits the criteria of this cluster with its high SSA, but is otherwise anomalous in its optical properties. The optical property values fall within the bounds of multiple aerosol type thresholds suggested by various studies, and thus the aerosol type will be considered mixed. The majority of stations within this cluster (CPT, GRW, PYE, THD) are remote marine sites that may receive occasional dust,

biomass burning and/or pollution events amid typical sea salt particle measurements, so a mixed dominant aerosol type might be expected. A slight outlier to this group is WLG, which is a remote mountaintop site. Although sea salt particles are very unlikely to be a constituent of the WLG aerosol, WLG does experience strong dust and pollution events, depending on the season and wind direction (Kivekäs et al., 2009). The main difference between Clusters 5 and 6 is the higher AAE and lower SAE values in Cluster 5 (CPR), indicating a stronger presence of large dust particles and smaller influence of EC/OC pollution at CPR compared to the locations in Cluster 6.

Page 19: [4] Deleted	Microsoft Office User	7/13/17 11:48:00 AM

variation in aerosols over time at a site that are dependent on the geography from which the air masses arrived at the station.

5