Author Response

acp-2018-1297

We thank the two anonymous peer reviewers for their helpful comments and suggestions. Each of the reviewer comments is listed below in bold, followed by our response and the changes made to address each comment in plain text. In some cases, comments which address similar issues were reordered for clarity. In these cases, relevant corrections that were made for multiple comments are listed after each comment.

An updated version of the manuscript with changes highlighted follows the comment responses.

Anonymous Referee #1 Comments

Pp1 line 29: The sentence starting on this line needs to be reworded. It is not readily understandable upon a reading.

This sentence and the remainder of the abstract were simplified and clarified to read:

"Differences between many of the terrestrial and marine population types in total CCN number concentrations active at a specific supersaturation were often not as pronounced as the associated differences in the corresponding activated fraction spectra, particularly for supersaturations below about 0.4%. This finding was due to the generally higher number concentrations in terrestrial airmasses offsetting the lower fraction of particles activating at low supersaturations. At higher supersaturations, CCN concentrations for aged terrestrial types were typically above those of the marine types due to their higher number concentrations."

Pp2 line 4: ACAPEX is not defined upon initial use.

This was updated with the full ACAPEX title as:

"U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Climate Research Facility Cloud-Aerosol-Precipitation Experiment (ACAPEX)"

Pp2 line 19: A source should be given for the sentence that starts on this line.

This sentence was updated as:

"As the aerosol ages, chlorine replacement by uptake of acidic gases can reduce the hygroscopicity of the salts (Finlayson-Pitts and Pitts, 1999; Song and Carmichael, 1999)."

Pp3 line 15, Pp4 line 28: The references in this paper require some attention. In these lines, the punctuation appears incorrect.

Pp7 line 21 and Pp11 lines 13 and 14: The reference manager again seems to be troubled. These references need to be fixed.

The reference manager was evidently creating broken links. Reference fields throughout the manuscript have now been regenerated, updated, and locked to prevent further errors.

Pp4 line 7: Information about the relative smoothness of size distribution of the local contaminant would lend further credibility to this claim. A local contaminant would often have a choppier character in the individual distributions as compared to particles that are more aged and more processed.

This is a particularly interesting suggestion for improvements to this methodology, as many of the individual size distributions may indeed have what could be identified as essentially increased noise in the measured values. However, in some cases of influence from a local contaminant an increased signal to noise ratio or similar metric did not occur. For instance, in the example number size distribution spectra from Fig. 1 there was some noise or lack of smoothness in the measured values at the smallest bin sizes for the non-contaminated spectrum (Fig. 1a), while a relatively smooth spectrum was seen in the small contamination mode a short time later (Fig. 1b). This was in part due to higher levels of noise in the smallest bin sizes measured by the SMPS (e.g. the individual distributions in Fig. 4), particularly when few particles were observed in these bins. The effect of expected increased noise associated with fresh contaminant modes was therefore competing with the effect of increased measurement noise at small bin sizes in some of the SMPS data points. As a result, a reliable criterion associated with this behavior could not be established in this dataset and we therefore opted not to try to incorporate such a metric into our methodology in this work.

The following text was added to the supplemental materials section on removal of contamination modes to address this consideration:

"Further criteria for identification of contamination periods and modes in the number size distribution were considered, including analysis of signal-to-noise or similar measures of relative smoothness of potential contamination modes. However, such metrics were not always consistent in their identification of such small contamination modes due to competing factors from measurement noise and physical variability. Ultimately, the criteria listed in the main text were found to be sufficiently capable of removing small contamination modes such that the cluster model was able utilize these data points without skewing the results, and so were used for the purposes of this analysis."

Pp6, section starting 2.4.1 requires more attention than the rest of the paper. These sections require significant clarification.

-It appears that the authors used a hierarchical clustering scheme to identify the ideal number of clusters. How was this done? Was it bootstrapped? What portion of the data was used? If all the data why then go back to a k-means scheme?

The purpose of the hierarchical clustering scheme was to help identify the appropriate number of clusters to use for the K-means clustering, and was not otherwise related to the K-means clustering methodology. An agglomerative clustering method was used for this hierarchical clustering using the same distance function as the K-means analysis, wherein subsequent steps merged the two data points or clusters with the smallest value of the distance function between them until only one cluster remained. A dendrogram was then used to check for increases in

distance between subsequent clustering steps. Steps with the largest increase were treated as potential candidate numbers of cluster for the subsequent K-means analysis. This process for selecting potential numbers of clusters appropriate for the K-means analysis was clarified at the end of second 2.4 as:

"Selection of the appropriate number of clusters for the K-Means analysis followed a similar methodology to that described in Atwood et al., (2017), and was based on both an initial hierarchical clustering and the use of internal validity measures (Wilks, 2011). First, a hierarchical agglomerative cluster analysis was created using the cluster.AgglomerativeClustering class of scikit-learn to merge the two data points or clusters with the smallest value of the distance function together into a single cluster in subsequent steps until only one cluster remained. A dendrogram was then used to identify potential numbers of clusters at agglomeration steps that had the largest increase in distance between merged clusters. Various internal validity measures of clusters in the ensuing K-Means clustering (Beddows et al., 2009; Baarsch and Celebi, 2012) were then used to further assess appropriate numbers of clusters. The final selection of appropriate cluster numbers was based on each of these measures, along with verification that the results maintained physically distinct and temporally coherent clusters, in order to select the appropriate number of K-means clusters."

-There is inconstancy in the reported number of variables being used in the classification. In line 14 it is reported as 24, in line 24 it is suggested that it is 20. These should be made consistent AND exactly what each of these 20 or 24 variable is should be clearly identified.

These various values are the number of variables in several categories of variables used in the clustering. These included 20 variables for the size distribution, 20 variables for the activated fraction distribution, 24 variables for the HYSPLIT trajectories, 2 variables for the local wind speed, and 1 variable for the total number concentration, for a total 67 clustering variables used for each data point in the distance function.

This was clarified in section 2.4.1 and now reads:

"A total of 67 clustering variables were used to describe each time stamp in this analysis, which included measurements of aerosol microphysical properties and meteorological parameters at BML. Aerosol property variables included the normalized size distribution (normalized to an integrated value of 1 cm^{-3} by dividing by the total particle number concentration), after correction for local contamination. Values of the normalized number size distributions ($dN/dlog_{10}Dp$) were discretized into 20 logarithmically spaced bins, with each mean bin value then serving as a separate variable in the clustering distance function (e.g. Charron et al. (2008)). Similarly, activated fraction spectra (the fraction of particles activated at supersaturation S) for each time stamp were divided into 20 linearly spaced bins distributed between 0% and 1.1% supersaturation to incorporate CCN properties into the analysis. Total particle number concentration was included as a separate cluster variable.

Meteorological parameters at BML included the local 10m observed wind velocity as perpendicular u and v component variables (2 variables). Additionally, HYSPLIT backtrajectories were assigned to data points closest in time to each trajectory. The backtrajectory was converted to separate variables for the distance function by determining the

distance from the receptor, initial bearing, and altitude, every three hours backwards along the trajectory for 24 hours, yielding a total of 24 trajectory clustering variables for each time stamp."

-The choice of a Pearson Euclidian distance is confusing in this case and should be further justified. Is this the same thing as a Karl Pearson distance in which weights are usually standardized by standard deviation? If normalizing the weights why not just use a Euclidian distance? If a true Karl Pearson Euclidian distance function is being used why aren't the weights a reciprocal of variance rather than evenly distributed by the number of variables? This section requires significant justification and explanation.

We agree that this terminology was confusing and in need of correction, and appreciate the efforts of the reviewer in identifying this. In particular, we did in fact first standardize each of the variables using the standard deviation to account for the different scales of the clustering variables. As such and as noted by the reviewer, we should have referred to this distance function as the "Karl Pearson distance" rather than the "Karl Pearson Euclidian distance". This has been corrected.

We further weighted each of the variables based on the variable categories noted in the previous section. As there was a total of 67 variables, but only five categories of variables (size distribution, activation spectra, number concentration, wind speed, and backtrajectory), the purpose of this additional weighting was to reduce the impact of differing numbers of variables within categories, similar to the intent of the Karl Pearson standardization.

This section has been changed to explain the method more clearly:

"As variables of different scale were used in the cluster modelling, the Karl Pearson distance function (Wilks, 2011) was used, wherein each variable is first standardized to ensure they have equal weight on the distance function. However, as the 67 clustering variables were grouped into five categories of measurements (number size distribution, activation spectrum, total particle number concentration, wind speed, and backtrajectory), a further modification to the weighting (described below) was used to reduce the impact of differing numbers of variables in these categories on the distance function. The resulting distance between any two data points i and j was given by the function:

$$d_{i,j} = \left[\sum_{k=1}^{K} \boldsymbol{w_k}\left(\boldsymbol{x_{i,k}} - \boldsymbol{x_{j,k}}\right)^2\right]^{\frac{1}{2}} \quad (1)$$

where d_(i,j) gives the distance between two data point vectors, x_i and x_j, in a K-dimensional space (i.e. K nominally independently measured or observed, orthogonal variables at each time stamp), for each variable, k, with weight, w_k, calculated for each variable by:

$$\boldsymbol{w_k} = \frac{v_k}{s_k} \quad (2)$$

where s_k is the variance of the variable k, and v_k is the further relative weight used for each variable.

Only data points with valid measurements for number size distribution and activated fraction were used in the clustering analysis, leaving approximately 94% of data points (3357 of a total of

3583) utilized. Of the remaining data, two data points had partially invalid activation spectra and were kept in the analysis. The partial spectra had invalid data points imputed to the variable mean to minimize their impact on the distance function.

The size distribution and activated fraction variable categories, each of which had 20 variables, were given relative weights, v_k, of 1/20 such that their total relative weights summed to 1. As these variables were of primary importance to aerosol population microphysical properties, the other variable groups were decreased in relative importance. The backtrajectory and wind vector categories were each assigned a relative weight of 0.5, and the total particle number concentration variable assigned a relative weight of 0.1. As cluster analysis is by its nature an exploratory data analysis technique, these relative weight values were reached by varying their values and assessing the physical interpretability of the results (discussed further in the next section)."

-In line 20, missing data are referenced. The fraction of missing data and which variables are most often missing should be specified.

In this analysis, only data points with valid number size distribution and activation spectra data were utilized in the cluster analysis, leaving approximately 94% of data points (3357 of 3583). Of these data points, two had only partially valid activation spectra. Invalid values for these spectra were imputed to the variable mean value for the purposes of the distance function. All other clustering variables had valid data for each clustering data point. This information was added to the text as:

"Only data points with valid measurements for number size distribution and activated fraction were used in the clustering analysis, leaving approximately 94% of data points (3357 of a total of 3583) utilized. Of the remaining data, two data points had partially invalid activation spectra and were kept in the analysis. The partial spectra had invalid data points imputed to the variable mean to minimize their impact on the distance function."

In line 8 (Pp7), the "physical interpretability" should be clarified. Is this just empirical judgment?

The main criteria for assessing the proper number of clusters in this analysis were temporal coherence of the clusters (i.e. data points assigned to specific clusters occurring sequentially or close in time, rather than randomly distributed throughout the study period), and clusters with physically interpretable or meaningful results (i.e. the cluster results could be related to reasonable physical interpretations such as changes in cluster types associated with land/sea-breeze shifts).

The meaning of "physically meaningful" has been clarified and expanded upon by adding the following to the first description of these criteria in this section as:

"… temporally coherent (i.e. data points assigned to specific clusters tended to occur near each other rather than randomly distributed throughout the study period) and physically meaningful (i.e. could be related to physical phenomena such as land/sea-breeze shifts) …"

In addition, the last sentence of this section with the noted reference to "physical interpretability" has been clarified as follows:

"The eight-cluster option had been initially identified as potentially appropriate based on the internal validity measures and hierarchical clustering, while other numbers of clusters did not improve results based on the criteria of temporally coherent clusters and physically meaningful interpretation of the cluster results. As such, the eight-cluster model was selected as the most appropriate unsupervised classification result."

Pp11 line1: The modal kappas should be reported as 0.3-0.5 or 0.30-0.54.

We have made this correction.

Pp11 line 2: Is there a data associated with Phillips paper?

The comparison to the Phillips et. al., (2018) results was corrected and clarified as follows:

"Mean measured κ values of 0.49 and 0.46 for the M1 and M3 population types, respectively, were within the range of mean κ values (0.30 to 0.54) for marine aerosol dominated periods in a coastal outflow marine region presented by Phillips et al., (2018). Mean values for terrestrial population types T1–T5 varied between 0.15 and 0.25, and were consistent with the 0.20 value for Aitken mode continental outflow aerosol reported by Phillips et al. (2018)."

Pp22 Figure 2: The CN concentration measurements might be better read in a table rather than represented this way. This is especially true in figure c in which they appear to run over the top of the y axis.

We agree that the figure needed to be updated to improve the interpretability of CN concentration. As CN concentration is independent of supersaturation, the point markers have been replaced with a bar in the background highlighting the range of CN observations, and we have included this range as a text value on the plot.

Anonymous Referee #2 Comments

Page 1/13: Include the season when the study occurred after so the readers can place the rest of the abstract in context.

This has been updated to:

"… are presented for approximately six weeks of observations during the boreal winter/spring as part of the CalWater-2015 field campaign."

Page 1/22: Replace "region" with "regions"

This has been corrected.

Page 1/33: The abstract could be a bit stronger if there was a statement or description of the larger implications or purpose of the study.

Page 2/28-30: This sentence is along the lines of what would help the abstract be stronger.

Page 3/6: Again, a larger implication statement here would help motivate the work. Perhaps move the lines 28-30 on page 2 to here.

We appreciate the reviewer's suggestion and have modified the abstract accordingly.

The beginning of the abstract has now been updated to read:

"Aerosol particle and cloud condensation nuclei (CCN) measurements from a littoral location on the northern coast of California at Bodega Bay Marine Laboratory (BML) are presented for approximately six weeks of observations during the boreal winter/spring as part of the CalWater-2015 field campaign. The nature and variability of surface (marine boundary layer, MBL) aerosol populations were evaluated by classifying observations into periods of similar aerosol and meteorological characteristics using an unsupervised cluster model to derive distinct littoral aerosol population types and link them to source regions. Such classifications support efforts to understand the impact of changing aerosol properties on precipitation and cloud development in the region, including during important atmospheric river (AR) tropical moisture advection events."

Page 2/4: Please spell out "ACAPEX"

This was updated with the full ACAPEX title as:

"U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Climate Research Facility Cloud-Aerosol-Precipitation Experiment (ACAPEX)"

Page 2/19: What do the authors mean here by "distribution-averaged" kappa?

Page 4/26: Again, what is meant by "distribution-averaged"?

We agree that this phrasing was confusing and unnecessary, and it has been removed. To clarify: the average value of kappa was obtained by averaging all measurements of hygroscopicity at various supersaturations during the time periods associated with each cluster.

Page 2/22-25: This sentence is a little unclear. Wouldn't organic aerosol necessarily be influenced by organic material?

This sentence has been clarified to read:

"In contrast, average hygroscopicities for continental aerosol populations are often in the range of 0.1 to 0.3 due to heavy influence of organic and insoluble material, although individual inorganic sulfates and nitrate salts also found in continental aerosols have higher values, in the same range as many remote marine aerosol populations (Andreae and Rosenfeld, 2008)."

Page 2/31: Later in this paragraph a couple of dates are given regarding previous work. It would help to provide the overall dates of the study here to put those into context.

Page 3/6: "outside those periods"- see earlier comment on providing the dates of the study earlier, perhaps at the beginning of this paragraph.

We have added the dates used for observations as part of this study and reworded the first sentence in this paragraph to read:

"In this study, observations from 23 January 2015–5 March 2015 at the BML ground site were used to evaluate the nature and variability of surface (marine boundary layer, MBL) aerosol populations."

Page 3/1: What kind of data?

The referenced study included other sources of measurements, including other aircraft and ship observations, as part of both the CalWater-2015 and ACAPEX experiments. This has been clarified in this sentence as:

"In prior work analyzing ground-, ship-, and aircraft-based meteorological and microphysical observations from the CalWater-2015 and ACAPEX field experiments, Leung (2016) discusses an AR event that was first observed and sampled from aircraft platforms off the coast of California on 5 February 2015, during the observational period analyzed here."

Page 3/13: Is this 5 m above the roof? So total height above ground is?

This height was clarified as:

"… to a height of approximately 5 m above the ground, …"

Page 4/4: I assume these are all number median diameters and number geometric standard deviations?

Page 7/19: Add "number" after "best-fit"

Page 7/23: Add "number" before "size distributions"

This correction has been added to each of these instances to clarify that we are utilizing number size distributions for these parameterizations.

Page 4/9: I'm not sure what the authors mean by smallest fitted mode's dN/dlogDp. At what size? Do they mean this corresponds to the value of dN/dlogDp at the smallest mode's median diameter?

This criterion was to check that the distribution fit for the smallest mode constituted less than half to total value at 50 nm. This has been reworded to read:

"The smallest fitted mode constituted less than 50% of the total fitted number concentration at 50 nm (indicating only a small contribution to total particle count at larger sizes)"

Page 4/17: Include "contamination" between "local" and "mode"

This has been added.

Page 5/7: Refer to figures as Fig. 2a and Fig. 2b.

We have updated this sentence to include this as:

"Example sfCCN concentrations (Fig. 2a) and activated fraction spectra (Fig. 2b) are shown for the same time stamp as the size distribution data in Fig. 1."

Page 5/10: It would help to move these references to after the descriptions of the figure, so "Fig. 2c" to line 11 and Figure 2d to line 12.

This change was made to the sentence as:

"During the contamination period approximately one hour later the effect of the small mode noted in Fig. 1b was seen via increased CN concentrations (reaching as high as 5000 cm-3; Fig. 2c) and decreased activated fractions above approximately 0.1% supersaturation (Fig. 2d)."

Page 5/14: Change "Fig. 2(e&f) to "Fig. 2e and Fig 2f., respectively).

This change was made as:

"After removal of this small contamination mode the corrected CN concentrations and activation spectrum Fig. 2e and Fig. 2f, respectively, were similar in characteristics to the earlier non-contaminated period."

Page 6/7: Was this done using the fit lines shown in Figure 2?

In this sentence the distribution was intended to refer to all data points, which were discretized into bins and then averaged, but we agree that this was confusingly worded. We have updated the sentence to read:

"Values of the normalized number size distributions ($dN/d\log_{10}Dp$) were discretized into 20 logarithmically spaced bins, with each mean bin value then serving as separate variables in the clustering distance function (e.g. Charron et al. (2008))."

Page 6/17: Please include what "i" and "j" refer to.

This has been updated to describe the equation as:

"The resulting distance between any two data points i and j was given by the function: …"

Page 7/22: Please move the figure part (a,b, etc.) before the thing in the figures, so "(a) aerosol and (b) meteorological. . ."

This has been corrected as:

"Figure 3 presents the study timelines of measured (a) aerosol, and (b) meteorological variables, and the (c) corrected and (d) normalized aerosol size distributions."

Page 7/31: Is the normalized size distribution the average of lognormals or the all size distributions averaged and then fit?

All sized distributions were averaged and then fit. This has been clarified earlier in the section as:

"Multi-modal lognormal number size distributions were fit to the average of all spectra associated with each cluster. Fit parameters and best-fit CCN-spectrum activated fraction parameters (see Supplemental Material) for each cluster are provided in Table 2."

Page 10/3: This sentence is somewhat unclear- I think the authors mean that ". . .Fig. 3 indicated that the size distributions associated with the T1 cluster grew. . .", not that that T1 cluster grew.

This is correct and we have included this important clarification as:

"During these times, normalized size distributions and modal median diameter fits shown in Fig. 3 indicated that the size distributions associated with the T1 cluster grew over the course of several days into distributions with larger median diameters, which were then classified as other terrestrial cluster types."

Page 11/2: Add "(2018)" after "Phillips et al."

Page 11/4: Same comment as for line 2.

Page 11/30: Add "(2010)" after "Kammermann et al."

The year has been added to correct each of these citations.

Page 12/20: What RH is considered "dry" here?

In this analysis the dry values indicate mass scattering efficiencies at 0% RH, with the change in scattering efficiency due to RH reflected by the curves in Fig. 8 (constructed assuming a deliquesced aerosol). In order to better represent this effect, we have made changes to our optical reconstruction and discussion to better reflect the effect of RH on measurements.

The following sentence was added:

"The average size distribution for each aerosol population type was assumed to be made for a deliquesced aerosol at the study average SMPS measured RH of 35.5% (Martin et al., 2017) and the cluster average κ given in Table 1"

In addition, the discussion of dry RH was updated to read:

"Mass scattering efficiencies at 550 nm for particles modeled at 0% RH ranged between 3.6 and 7.6 $m^2 g^{-1}$, although actual values would be expected to be lower due to expected particle dry densities higher than 1 g cm^{-3}."

The MODIS comparison was also removed as it is not as directly comparable to mass scattering efficiency on a dry aerosol mass basis and could lead to confusion in its interpretation.

Page 12/24: This is true, but super-micron particles also are typically associated with lower mass scattering efficiencies.

We agree that this is an important point to make and have updated this to read:

"Further, these values represent only the contribution to mass scattering efficiencies from fine mode aerosol. Coarse mode aerosol, including particles generated by sea spray in littoral environments, can represent a large fraction of the total light scattering, although their mass scattering efficiencies are typically lower."

Page 14/9: Check the website provided for data repository. I assume this will be up- dated upon final submission.

This dataset has now been submitted to the repository at doi:10.5281/zenodo.2605668.

Page 20/ 22: Table 1 caption: Add "percentage of observations" to the caption list and remove the "Best-fit size distribution. . ." sentence.

This has been corrected and now reads:

"Table 1 Aerosol and meteorological parameters for each of the cluster time periods with the total number and percentage of observations in each cluster. Cluster mean values are given for total particle number concentration, κ hygroscopicity parameter from the srCCN system, HYSPLIT 24-hour accumulated precipitation along the trajectory, and local wind velocity observations."

Page 20/26: Table 2 caption: Include "number" for "best-fit number size distribution". The caption could be expanded to include something about the cluster types (what 'M' and 'T' mean), different modes and describe "a,b,c" (form of equation)

These changes have been made and the caption now reads:

"Table 2 Cluster best-fit number size distribution and activated fraction parameters are shown, with activated fraction parameters a, b, and c pertaining to the fit model equation given in the supplemental information. Clusters with "M" and "T" names refer to marine and terrestrial aerosol population types, respectively."

Page 27: Figure 7: State in the caption what the colors represent.

The following sentence was added to the caption:

"Colors represent the data point density as a fraction of the maximum density in each plot."

Page 28: Figure 8: The title of this figure reads "dry", although these are clearly a function of relative humidity?

The "dry" was intended to indicate the mass scattering efficiency was calculated on a dry aerosol mass basis, but we agree it is confusing. Instead we have changed the figure to read "Scattering efficiency per unit dry mass– 550 nm", and indicated that this value is calculated against total dry particle mass in the caption as:

"Figure 8 Reconstructed mass scattering efficiencies (per unit dry aerosol mass) for each of the cluster population types across a range of environmental relative humidity values. An assumed dry index of refraction of 1.5 + 0.0i was used for all population types to highlight the differences in aerosol optical properties expected due to differences in population average size distribution, hygroscopicity, and relative humidity."

Supplemental References, Line 36-27; 47: Check reference

The reference manager was evidently creating broken links. Reference fields throughout the manuscript have now been regenerated, updated, and locked to prevent further errors.

Classification of aerosol population type and cloud condensation nuclei properties in a coastal California littoral environment using an unsupervised cluster model

Samuel A. Atwood[1], Sonia M. Kreidenweis[1], Paul J. DeMott[1], Markus D. Petters[2], Gavin C. Cornwell[3], Andrew C. Martin[4], Kathryn A. Moore[1,3]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA
[2]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA
[3]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA
[4]Climate Atmospheric Science and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA, USA

Correspondence to: Sonia M. Kreidenweis (soniak@colostate.edu)

Abstract. Aerosol particle and cloud condensation nuclei (CCN) measurements from a littoral location on the northern coast of California at Bodega Bay Marine Laboratory (BML) are presented for approximately six weeks of observations during the ~~CalWater 2015 field campaign. A combination of aerosol microphysical and meteorological parameters was used to classify variability in the properties of the BML surface aerosol using a K-means cluster model.~~boreal winter/spring as part of the CalWater-2015 field campaign. The nature and variability of surface (marine boundary layer, MBL) aerosol populations were evaluated by classifying observations into periods of similar aerosol and meteorological characteristics using an unsupervised cluster model to derive distinct littoral aerosol population types and link them to source regions. Such classifications support efforts to understand the impact of changing aerosol properties on precipitation and cloud development in the region, including during important atmospheric river (AR) tropical moisture advection events. Eight aerosol population types were identified that were associated with a range of impacts from both marine and terrestrial sources. Average measured total particle number concentrations, size distributions, hygroscopicities, and activated fraction spectra between 0.08% and 1.1% supersaturation are given for each of the identified aerosol population types, along with meteorological observations and transport pathways during time periods associated with each type. Five terrestrially influenced aerosol population types represented different degrees of aging of the continental outflow from the coast and interior of California, and their appearance at the BML site was often linked to changes in wind direction and transport pathway. In particular, distinct aerosol populations, associated with diurnal variations in source ~~region~~regions induced by land/sea-breeze shifts, were classified by the clustering technique. A terrestrial type representing fresh emissions, and/or a recent new particle formation event, occurred in approximately 10% of the observations. Over the entire study period, three marine influenced population types were identified that typically occurred when the regular diurnal land/sea-breeze cycle collapsed and BML was continuously ventilated by air masses from marine regions for multiple days. These marine types differed from each other primarily in the degree of cloud processing evident in the size distributions, and in the presence of an additional large-particle mode for the type associated with the highest wind speeds. One of the marine types was associated with a multi-day period during which an atmospheric river made landfall at

BML. ~~The generally higher total particle number concentrations but lower activated fractions of four~~Differences between many of the terrestrial and marine population types ~~yielded similar~~ in total CCN number concentrations ~~to two of the marine types~~ active at a specific supersaturation were often not as pronounced as the associated differences in the corresponding activated fraction spectra, particularly for supersaturations below about 0.4%. ~~Despite quite different activated fraction spectra, the two remaining marine and~~ This finding was due to the generally higher number concentrations in terrestrial ~~types had~~airmasses offsetting the lower fraction of particles activating at low supersaturations. At higher supersaturations, CCN ~~spectral number~~ concentrations ~~very similar to each other, due in part to~~for aged terrestrial types were typically above those of the marine types due to their higher number concentrations ~~associated with the terrestrial type~~.

1 Introduction

Atmospheric rivers (ARs) are tropical moisture advection phenomena that can account for large fractions of the wintertime precipitation in California (Ralph et al., 2004; Dettinger et al., 2011). The winter-spring 2015 CalWater-2015 study (Ralph et al., 2015), and coordinated ~~ACAPEX study~~U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Climate Research Facility Cloud-Aerosol-Precipitation Experiment (ACAPEX) ~~(Leung, 2016)~~(Leung, 2016) that included aircraft- and ship-based observations in the same region, were designed to probe the atmospheric conditions in and around ARs, and to provide new observations of the characteristics of regional aerosols that may interact with these atmospheric moisture features and thereby influence the downwind formation of precipitation. As part of the CalWater-2015 study, ground-based aerosol observations were conducted at the Bodega Bay Marine Laboratory (BML), a coastal California site that is suitable for observation of aerosols in landfalling marine air masses, and in mixtures of marine and continental air.

In marine regions impacted by continental outflow, aerosol chemical and microphysical properties, including particle number concentrations and size distributions, are often moderated by impacts from terrestrial sources (Nair et al., 2013; Wex et al., 2016; Zhao et al., 2016; Phillips et al., 2018). For example, freshly emitted sea spray aerosol particles comprise a mixture of salts with generally high hygroscopicities ($\kappa \sim 0.6–1.2$), and co-emitted organic species with lower hygroscopicities ($\kappa \sim 0–0.3$) (Prather et al., 2013; Quinn et al., 2014)—classified using the κ hygroscopicity parameter (Petters and Kreidenweis, 2007). Bulk hygroscopicity values above 1 are infrequently observed, but have been reported for some background and precipitation impacted marine aerosol populations (Good et al., 2010; Prather et al., 2013). ~~As the aerosol ages, chlorine replacement by uptake of acidic gases can reduce the hygroscopicity of the salts. Aging and organic components result in reported distribution-averaged~~As the aerosol ages, chlorine replacement by uptake of acidic gases can reduce the hygroscopicity of the salts (Finlayson-Pitts and Pitts, 1999; Song and Carmichael, 1999). Aging and organic components result in reported κ values in the range of 0.4 to 0.7 for more pristine or background marine aerosol populations (Keene et al., 2007; Bates et al., 2012; Prather et al., 2013; Quinn et al., 2014; Zhang et al., 2014; Forestieri et al., 2016; Atwood et al., 2017; Royalty et al., 2017; Phillips et al., 2018). In contrast, average hygroscopicities for continental ~~organic~~ aerosol populations are ~~typically~~often in the range of 0.1 to 0.3 due to heavy influence of organic and insoluble material, although individual inorganic sulfates and nitrate

salts have higher values in the same range as many remote marine aerosol populations (Andreae and Rosenfeld, 2008). The cloud-droplet-nucleating activity of particles in coastal regions therefore depends on the composition, size distribution, and degree of aging of the marine particles, along with the characteristics of particles from non-marine sources that become mixed with the marine aerosol. Thus, observation and analysis of the coastal California aerosol environment will support efforts to better understand implications of aerosol properties on precipitation and cloud development in the region, including aerosol influences on how precipitation develops in landfalling ARs in this region.

In this study, observations from 23 January 2015–5 March 2015 at the BML ground site ~~during the CalWater-2015 field campaign~~ were used to evaluate the nature and variability of surface (marine boundary layer, MBL) aerosol populations ~~at BML~~. Observations were classified into periods of similar aerosol and meteorological characteristics using an unsupervised cluster model (Atwood et al., 2017) to derive distinct littoral aerosol population types and link them to source regions. The clusters were assessed with respect to their aerosol characteristics, particularly the contributions of each population type to cloud condensation nuclei (CCN) concentrations that influence the microphysical properties of liquid-phase clouds and the evolution of precipitation in mixed-phase clouds (e.g., Rosenfeld et al., 2008). In prior work analyzing ~~data~~ground-, ship-, and aircraft-based meteorological and microphysical observations from the CalWater-2015 ~~/~~and ACAPEX field experiments, Leung ~~(2016)~~(2016) discusses an AR event that was first observed and sampled from aircraft platforms off the coast of California on 5 February 2015, during the observational period analyzed here. The AR made landfall near BML on the morning of 6 February, and the AR-associated front, producing heavy precipitation along its trajectory, reached the Central Valley and Sierra Nevada in the afternoon. We thus also contrast the aerosol cluster types identified at BML during AR and marine aerosol dominated events to observations outside those periods.

2 Methods and site description

Measurements occurred at Bodega Bay Marine Laboratory (BML; 38.32° N, 123.07° W) and took place between January and March 2015. Data utilized in this study were gathered between 23 January and 5 March, when relevant instruments were operational. All aerosol measurements used here were made from within the Colorado State University/National Park Service mobile lab, located approximately 200 m from the coast, and approximately 150 m north of the main BML lab building. Air was sampled through several ½" stainless steel tube inlets routed through the roof of the mobile lab to a height of approximately 5 m above the ground, before the flow was split and sent to various instruments described further below. A map and additional description of the site and instrumentation are provided in Martin et al., (2017).

2.1 Meteorological data

Local meteorological measurements were obtained from a 10 m surface met tower located approximately 100 m north of the mobile lab, and operated as part of a NOAA/ESRL observation network (White et al., 2013). Air mass source regions were defined using 24-hour backtrajectories that were initiated every three hours during the study period and computed using the

HYSPLIT version 4.9 Lagrangian parcel model (Draxler and Hess, 1997, 1998; Draxler et al., 1999; Stein et al., 2015) with the 40 km x 40 km Edas40 meteorological dataset. All trajectories were generated with 100 m arrival heights at BML to characterize transport in the MBL.

2.2 Aerosol size distribution measurements (SMPS)

Aerosol size distribution measurements made during CalWater-2015 at BML are described in detail in Martin et al., (2017). This study used size distribution measurements for particle diameters between 14 and 730 nm from a TSI 3080 SMPS using a 0.3 L min^{-1} sample flow rate and 3.0 L min^{-1} sheath flow rate. SMPS scans were conducted approximately every 5 minutes and were subsequently averaged to approximately 15-minute time periods to match additional coincident measurements.

The presence of large numbers of particles smaller than approximately 30 nm in diameter was intermittently observed in the size distribution measurements. These particles were likely from a combination of local sources that included vehicle and other activity at BML, local camp and brush fires, and emissions from the nearby town of Bodega Bay, located to the east of the site. This nucleation mode was generally superimposed on more stable aerosol populations, and thus rather than removing the entire observation during these events, we removed only the contamination mode, as described below and in further depth in the supplemental material.

The best-fit modal parameterization for each size distribution spectrum was first assessed using a lognormal mixture distribution fitting algorithm based on Hussein et al., (2005), and as described in Atwood et al., (2017). The algorithm selects between one and three modes to best represent the number size distribution based on empirical rules and maximum-likelihood fitting criteria, and defines each mode by three parameters (median diameter, geometric standard deviation, and fractional number concentration). Within each of the size spectra, a fitted mode was identified as local contamination if the following criteria were met:

- The combined fit's distribution was identified to have more than one mode
- The median diameter of the smallest fitted mode was at or below 20 nm
- The smallest fitted ~~mode's $dN/dlog_{10}D_p$ value was~~mode constituted less than 50% of the ~~combined fit's $dN/dlog_{10}D_p$ value~~total fitted number concentration at 50 nm (indicating only a small contribution to total particle count ~~in~~at larger ~~fit modes~~sizes)
- The smallest fitted mode did not persist continuously for more than three hours

In the case that an observed size distribution spectrum passed all these criteria, the smallest mode was classified as a contamination mode likely associated with local sources as described above and therefore not representative of the regional aerosol. These modes were removed from the size distributions and total number concentrations while retaining the remaining fitted modes. This was accomplished by multiplying the observed $dN/dlog_{10}D_p$ value by one minus the contamination mode's number fraction of the total distribution for each size bin. An example of the removal is shown in Figure 1, with an observed distribution without a local contamination mode shown in (a). Approximately one hour later, a distinct small mode was observed with a median diameter less than 14 nm and nearly no contribution to particle concentration at diameters above 20

nm (b), while total number concentration showed a rapid increase (not shown). This mode was no longer present approximately 1.5 hours later. Therefore, the observation passed all four criteria to be considered local contamination and was removed from the reported size distribution and total number concentration to give the final corrected distribution (c).

2.3 CCN measurements

As described in Martin et al. (2017), size-resolved CCN concentrations (srCCN) were measured using a DMT cloud condensation nuclei counter (CCN-100) coupled with a TSI 3080 SMPS to scan across a range of particle diameters (12-540 nm at water supersaturation of s = 0.1%, 0.19%, 0.28%, 0.44%, 0.58%, 0.67%). ~~A distribution averaged~~An apparent hygroscopicity parameter κ (Petters and Kreidenweis, 2007) was calculated from these scans as described in Petters and Petters, (2016).

A separate scanning flow CCN system (sfCCN) was used to measure total CCN number concentration as supersaturation was ramped between approximately 0.08% and 1.1% supersaturation. The system used a DMT CCN-100 instrument that had been modified by connecting a voltage modulated proportional flow valve to the bottom of the column to control flow (Suda et al., 2014). The flow rate through the CCN column was increased from 0.2 to 1.2 lpm over 5 minutes while holding the temperature gradient constant, thereby scanning peak column supersaturation as a function of flow rate and column temperature gradient (Moore and Nenes, 2009). A TSI 3010 Condensation Particle Counter (CPC) was placed in parallel to the CCN to measure total particle number concentration (CN). CCN and CN concentrations were recorded at a frequency of one hertz. Each approximately 5-minute scan was repeated three times, after which the temperature gradient was changed to scan a different range of supersaturations. As the column temperatures took approximately 3 minutes to stabilize, the first scan of each three-repetition set was not analyzed. As the residence time in the sfCCN varied with total flow rate when compared to the parallel CPC, the CPC timestamps were empirically adjusted to ensure the 1 Hz data points from both instruments were aligned, as described in more detail in the Supplemental Material. Calibrations of both CCN systems were conducted following the methodology described in Suda et al., (2012).

Example sfCCN concentrations (Figure 2a) and activated fraction spectra (Figure 2b) are shown ~~in Fig. 2(a&b), respectively,~~ for the same time stamp as the size distribution data in Figure 1. The range of CN concentrations as measured by the parallel CPC during this period are shown as red data points (placed for comparison to CCN concentrations at 1.3% supersaturation). During the contamination period approximately one hour later ~~(Fig. 2(c&d))~~ the effect of the small mode noted in ~~Fig. 1~~(Figure 1b~~)~~ was seen via increased CN concentrations (reaching as high as 5000 cm^{-3}; Figure 2c) and decreased activated fractions above approximately 0.1% supersaturation~~.~~ (Figure 2d). Observed CCN concentrations remained relatively constant indicating the additional particles were too small to activate below 1.1% supersaturation. After removal of this small contamination mode the corrected CN concentrations and activation spectrum ~~Fig. 2(e&f~~(Figure 2e and Figure 2f, respectively) were similar in characteristics to the earlier non-contaminated period.

2.4 Classification of aerosol population type

Classification of aerosol population types impacting BML was conducted using an unsupervised K-means cluster analysis. Such clustering methods utilize properties of the aerosol and environment to identify periods of potentially similar impacts and aerosol population types (Wilks, 2011). Cluster analyses have been used to classify aerosol particle size distributions (Tunved et al., 2004), associate them with various environmental and atmospheric processes (Charron et al., 2008; Beddows et al., 2009; Wegner et al., 2012), and conduct aerosol source apportionment studies (Salimi et al., 2014). Information on aerosol chemistry or composition has also been used for clustering purposes (Frossard et al., 2014), and has been integrated with size distribution measurements and atmospheric transport data to produce cluster results based on multiple types of observations (Charron et al., 2008; Atwood et al., 2017).

The K-means clustering methodology involved selection of specific variables that partially defined the state of the aerosol and meteorological environment at the sampling site. The degree of similarity of the state of the environment between any two data points (i.e. specific times) was estimated by the use of a distance function that grouped data points into clusters that had broadly similar values among the input variables. Here, we utilized the *cluster.KMeans* class of the Python scikit-learn package (Pedregosa et al., 2011) to perform the analysis. Selection of the appropriate number of clusters for the K-Means analysis followed a similar methodology to that described in Atwood et al., (2017)~~. A hierarchical cluster analysis was first created using the *cluster.AgglomerativeClustering* class of scikit-learn to identify potential numbers of clusters using a dendrogram. Various internal validity measures of the K-means clusters (Beddows et al., 2009; Baarsch and Celebi, 2012) were then used~~, and was based on both an initial hierarchical clustering and the use of internal validity measures (Wilks, 2011). First, a hierarchical agglomerative cluster analysis was created using the *cluster.AgglomerativeClustering* class of scikit-learn to merge the two data points or clusters with the smallest value of the distance function together into a single cluster in subsequent steps until only one cluster remained. A dendrogram was then used to identify potential numbers of clusters at agglomeration steps that had the largest increase in distance between merged clusters. Various internal validity measures of clusters in the ensuing K-Means clustering (Beddows et al., 2009; Baarsch and Celebi, 2012) were then used to further assess appropriate numbers of clusters. The final selection of appropriate cluster numbers was based on each of these measures, along with verification that the results maintained physically distinct and temporally coherent clusters, in order to select the appropriate number of K-means clusters.

2.4.1 Cluster variables

~~Variables~~A total of 67 clustering variables were used to describe each time stamp in this analysis, which included measurements of aerosol microphysical properties and ~~of~~ meteorological parameters at BML. Aerosol property variables included the normalized size distribution ~~at each time stamp~~ (normalized to an integrated value of 1 cm^{-3} by dividing by the total particle number concentration), after correction for local contamination. ~~The~~Values of the normalized number size distributions ($dN/dlog_{10}Dp$) were discretized into 20 logarithmically spaced bins ~~that served~~, with each mean bin value then serving as a separate ~~variables~~variable in the clustering distance function (e.g. Charron et al. ~~(~~., 2008~~)~~). ~~Activated~~Similarly,

activated fraction spectra (the fraction of particles activated at supersaturation S) for each time stamp were divided into 20 linearly spaced bins distributed between 0% and 1.1% supersaturation to incorporate CCN properties into the analysis. Total particle number concentration was included as a separate cluster variable.

Meteorological parameters at BML included the local 10m observed wind velocity as perpendicular u and v component variables. (2 variables). Additionally, HYSPLIT backtrajectories were assigned to data points closest in time to each trajectory. The backtrajectory was converted to separate variables for the distance function by determining the distance from the receptor, initial bearing, and altitude, every three hours backwards along the trajectory for 24 hours, yielding a total of 24 trajectory clustering variables for each time stamp.

2.4.2 Distance function

TheAs variables of different scale were used in the cluster modelling, the Karl Pearson Euclidean distance function (Wilks, 2011) used in the cluster model was modified to include a relative weight parameter forwas used, wherein each input variable is first standardized to ensure they have equal weight on the distance function. However, as the 67 clustering variables were grouped into five categories of measurements (number size distribution, activation spectrum, total particle number concentration, wind speed, and backtrajectory), a further modification to the weighting (described below) was used to reduce the impact of differing numbers of variables in these categories on the distance function. The resulting distance between any two data points i and j was given by the function:

$$d_{i,j} = \left[\sum_{k=1}^{K} \boldsymbol{w_k}\left(\boldsymbol{x_{i,k}} - \boldsymbol{x_{j,k}}\right)^2\right]^{\frac{1}{2}} \quad (1)$$

where $d_{i,j}$ gives the Euclidean distance between two data point vectors, x_i and x_j, in a K-dimensional space (i.e. K nominally independently measured or observed, orthogonal variables at each time stamp), for each variable, k, with relative weight, w_k. Each, calculated for each variable was first standardized, with missing data imputed to values of zero to minimize their impact on the distance function.by:

The

$$\boldsymbol{w_k} = \frac{v_k}{s_k} \quad (2)$$

where s_k is the variance of the variable k, and v_k is the further relative weightsweight used for each variable.

Only data points with valid measurements for number size distribution and activated fraction were included to prevent properties of the aerosol from becoming over weightedused in the cluster model dueclustering analysis, leaving approximately 94% of data points (3357 of a total of 3583) utilized. Of the remaining data, two data points had partially invalid activation spectra and were kept in the analysis. The partial spectra had invalid data points imputed to the variable mean to having more variables deseribing them. minimize their impact on the distance function.

The size distribution and activated fraction ~~variables~~variable categories, each of which had 20 variables, were given relative weights, $v_{k,}$ of 1/20 such that their total relative weights summed to 1. As these variables were of primary importance to aerosol population microphysical properties, the other variable groups were decreased in relative importance. The backtrajectory and wind vector ~~groups~~categories were each assigned a relative weight of 0.5, and the total particle number concentration variable assigned a relative weight of 0.1. As cluster analysis is by its nature an exploratory data analysis technique, these relative weight values were reached by varying their values and assessing the physical interpretability of the results~~.~~ (discussed further in the next section).

2.4.3 Number of clusters

Hierarchical clustering and internal validity measures indicated two, six, eight, and twelve clusters were potentially appropriate for the K-means analysis. Clusters associated with periods of marine aerosol impacts (discussed further in the next section) became temporally coherent ~~and physically meaningful~~(i.e. data points assigned to specific clusters tended to occur near each other rather than randomly distributed throughout the study period) and physically meaningful (i.e. could be related to physical phenomena such as land/sea-breeze shifts) after the number of clusters was increased to eight.

In the case of the twelve-cluster analysis, several of the clusters were composed of few or even a single data point, indicating the model had begun to separate outliers into distinct clusters. In addition, several of the temporally consistent clusters were split, indicating that too many clusters had been selected. All potential cluster numbers between seven and eleven were then investigated to determine if physically or temporally coherent population types emerged to a greater degree than the eight-cluster model. ~~As the~~The eight-cluster option ~~was~~had been initially identified as potentially appropriate~~,~~ based on the internal validity measures and ~~the~~ hierarchical clustering, while other ~~potential models~~numbers of clusters did not improve ~~physical interpretability~~results based on the criteria of temporally coherent clusters and physically meaningful interpretation of the cluster results. As such, the eight-cluster model was selected as the most appropriate unsupervised classification result.

3 Aerosol population type classification results and discussion

Three of the eight identified clusters were defined as “marine” population types, as backtrajectory data showed evidence of transport pathways primarily over ocean areas. These marine types, denoted as clusters M1–M3, tended to have lower average number concentrations (below approximately 1500 cm^{-3}), while the terrestrial clusters (T1–T5) had typical averages between approximately 2000 and 4000 cm^{-3} and were associated with transport from more terrestrial source regions. The exception to this was cluster T5, which had number concentrations of roughly 1500 cm^{-3}, more oceanic transport pathways, and size distributions with the largest median diameters among the “terrestrial” clusters. Table 1 provides the cluster-averaged number concentrations, wind velocities, HYSPLIT accumulated precipitation along the 24-hour trajectory, hygroscopicity parameters from the srCCN system, and the percentage of all measurements associated with each of the 8 identified clusters. ~~Best-fit size distribution~~Multi-modal lognormal number size distributions were fit to the average of all spectra associated with each cluster.

Fit parameters and best-fit CCN-spectrum activated fraction parameters (see Supplemental Material) for each cluster are provided in Table 2.

Figure 3 presents the study timelines of measured (a) aerosol ~~(a),~~ and (b) meteorological ~~(b)~~ variables, and the (c) corrected ~~(c)~~ and (d) normalized ~~(d)~~ aerosol size distributions. The fitted modal diameters are superimposed on the normalized number size distributions, revealing periods of stability in the size distributions as well as periods that are highly variable. Time periods associated with each cluster are shown as colors in the background of the panels.

Changes between identified cluster types often tracked diurnal changes in wind direction associated with the land/sea-breeze cycle (Fig. 3b; e.g., 28–30 Jan). Several periods occurred during the study during which this diurnal cycle collapsed and BML was ventilated with air masses from marine regions for several days at a time (starting on 4 Feb, 17 Feb, and 26 Feb). Clusters M1, M2, and M3 were selected by the model more consistently during these extended periods and confirmed their characterization as marine aerosol population types.

Normalized size distributions for the eight clusters are shown in grey in Figure 4, along with the average total corrected particle number concentrations and the total number of data points included in each cluster. Clusters with terrestrial and/or anthropogenic influence were ordered by increasing median diameter of the mode with the largest number fraction. HYSPLIT backtrajectories and wind rose plots for each data point included in each cluster are shown in Figure 5.

3.1 Marine population types

Further analysis of the marine clusters showed generally distinct meteorological conditions associated with each. Cluster M1 primarily occurred toward the end of the study period, during a period of high velocity winds from the northwest. Backtrajectories agreed with local winds and showed generally faster transport velocities. This cluster dominated during 26 and 27 February, a period during which the cleanest air masses and lowest number concentrations of the study were observed, reaching as low as 50 cm^{-3} during the height of the event. The normalized size distributions indicated that the Aitken mode dominated the number distribution (Figure 4), and also suggested the presence of a somewhat larger mode (particles larger than ~400 nm) that may have been associated with generation of sea spray by the higher wind velocities. Backtrajectories indicated that airmasses had passed over the ocean to the northwest of BML, while 24-hour accumulated precipitation along the trajectories (Table 1) indicated rainfall had occurred in the days prior to the airmass' arrival at BML. Each of these findings was consistent with classification of the M1 cluster as a precipitation scrubbed, clean marine aerosol population.

The marine cluster with the next highest number concentration was M2, with 78% of its total occurrences between 17 and 21 February. The wind rose for this cluster indicated a primarily northwesterly wind, similar to cluster M1 but with much slower velocities, and backtrajectories with oceanic transport pathways. While HYSPLIT does not always simulate sub-synoptic scale transport with high fidelity, this pattern may be indicative of slower transport of air from a marine region just off the coast, as opposed to the direct fast-transport path from more distant ocean regions seen in the M1 type. As observed for M1, the best-fit average normalized particle size distribution (Figure 4) was primarily bimodal, consistent with many reports of cloud-processed background marine aerosol populations (Hoppel et al., 1986; Bates et al., 2000; Wex et al., 2016; Atwood et al.,

2017; Royalty et al., 2017; Phillips et al., 2018), though with a larger accumulation mode number fraction than the M1 type. Minimal rainfall along the transport pathway was evident in the HYSPLIT accumulated precipitation estimates (Table 1), indicating no recent precipitation scrubbing and suggesting that more cloud processing without rainout led to larger numbers of particles in the accumulation mode compared with M1. In contrast to M1, for which a third fitted mode was found above 400 nm, the third mode in M2 occurred at very small particle sizes (diameters less than 30 nm).

The final marine cluster, M3, shared similarities with the other marine types, including a bimodal normalized size distribution with a minimum near 110 nm and indications of oceanic source regions in local winds and backtrajectories. However, the average total number concentration was nearly double that of the other marine types. The primary period during which this cluster occurred was during 4-9 February, bracketing the time before, during, and after landfall of the AR that impacted the BML region during CalWater-2015. This cluster is therefore interesting as it may be indicative of a unique population type associated with AR meteorological conditions. Some caution is warranted, however, as instrument downtime lead to a gap in the aerosol size distribution dataset during late 6 February through 7 February, during a high-wind and precipitation period that marked the landfall of the AR, and thus some key data that could be used to guide the clustering during this event were missing. When confined to the 5-8 February period noted by Leung ~~(2016)~~(2016) when the AR made landfall at BML, average number concentration was 749 cm^{-3} for time periods associated with the M3 cluster (excluding the data gap on 6-7 February), and 1052 cm^{-3} for the entire 5-8 February AR period (including the intermittent periods classified as terrestrial or anthropogenic aerosol).

Backtrajectories for the M3 cluster indicated source regions from just off the coast of the San Francisco Bay area prior to reaching BML, while HYSPLIT accumulated precipitation along the trajectory was the highest of any cluster (Table 1). Flows associated with AR landfall at the coast can be complex (Neiman et al., 2013), however, emissions from this urban area could potentially have mixed with the relatively low particle number counts that would be expected in a precipitation scrubbed AR air mass as it made landfall, accounting for the elevated (> 1000 cm^{-3}) total number (CN) concentrations that persisted for much of the AR, including during the high-wind and heavy precipitation period (Figure 3b) when SMPS and CCN data were not available. However, local generation of fine-mode sea spray aerosol could also be a factor during high winds.

3.2 Terrestrial population types

During periods dominated by diurnal shifts in aerosol and meteorological observations, and for short-duration periods during times associated primarily with marine clusters, the cluster model identified clusters that corresponded to terrestrially influenced populations. In the case of the multi-day events dominated by the M1 and M2 types, these short duration periods were often identified as either T4 or T5, clusters that were notable for having largely monomodal normalized size distributions with median diameters around 100 nm. Occurrences of these cluster types were often associated with a spike in number concentration and changes to either wind direction or wind velocity. Thus, the cluster model was able to identify and separate short duration periods of impacts from terrestrial sources during multi-day marine aerosol conditions at BML.

Several longer periods (28-31 January and 13-16 February) were observed during which populations T4 and T5 regularly alternated in tandem with the diurnal land/sea-breeze shift. Similar diurnal-shift behavior, but between clusters T2 and T3, occurred during 25–28 January and 1–5 March. During these diurnal shifts between various terrestrial clusters, the cluster with the larger median diameter was typically associated with the sea-breeze and transport from oceanic regions, while the smaller diameter cluster was associated with the land-breeze and transport from terrestrial regions. As aging of terrestrial aerosol typically leads to an increase in the median diameter of the aerosol modes, these four clusters may therefore be indicative of various degrees of aging of regional terrestrial aerosol during "sea-breeze resampling" at BML (Martin et al., 2017). Resampling occurs when terrestrial and marine aerosol populations mix and flow across the coastal boundary at low levels, leading to a region of mixed aerosol populations wherein the larger number concentrations associated with terrestrial types dominate the observed number concentrations in the resulting littoral zone airmass. When this diurnal cycle collapsed and BML was subjected to extended periods of sea-breezes and ventilation by air masses almost exclusively from ocean regions, the cluster model selected marine cluster types, indicative of marine airmasses that had not experienced much mixing with terrestrial air masses.

The terrestrial type T1 featured a dominant mode of particles with median diameters around 30 nm, indicating relatively little aging of the particles had occurred. Both low-level winds and backtrajectories during this cluster type indicated transport pathways from many directions (Fig. 5), though with the highest wind speeds of the terrestrial clusters, consistent with less time between the aerosol source and observation at BML. This cluster occurred primarily during two periods, 23–24 February and 28 February–1 March. During these times, normalized size distributions and modal median diameter fits shown in Figure 3 indicated that the size distributions associated with the T1 cluster grew ~~into other clusters with more dominant accumulation mode sizes~~ over the course of several days into distributions with larger median diameters, which were then classified as other terrestrial cluster types. The T1 cluster may therefore identify a freshly emitted population type or a recent new particle formation event.

3.3 CCN and activated fraction spectra characteristics

The best-fit activated fraction spectra, as functions of water supersaturation, are shown for all valid data points in each of the clusters in Figure 6(a). As a general comparison against other reported values for aerosol activation spectra, Figure 6 also shows the parameterized spectra reported by Paramonov et al. (2015) for a range of measurement locations from the European EUCAARI Network (grey background) and an overall typical average value (black line). The EUCAARI activation spectra were drawn from a range of marine, littoral, and continental sites, and were impacted by both marine and terrestrial airmasses, subject to a variety of emissions. The BML spectra spanned much of the range reported for the EUCAARI network, with the M2 cluster slightly above this range at intermediate superaturations. Activated fraction spectra are independent of total number concentration, thus the effect of particle size on activation is evident for the BML cases, with clusters with larger size modes having higher activated fractions across the range of measured supersaturations. These results show the wide range of activated

fraction spectra at BML associated with differences in aerosol population type, and the corresponding complexity of the population characteristics at this site.

In the CalWater-2015 dataset, marine population types all reached activated fractions of about 0.2 at supersaturations around 0.1% to 0.15%, while the terrestrial types did not reach equivalent fractions until supersaturations between approximately 0.18% and 0.6% were reached. Terrestrial population types with smaller median diameters tended to have less fractional activation across the full range of measured supersaturations, leading to activated fractions at 1.0% supersaturation that varied from approximately 0.3 to 0.85. However, due to the generally higher total particle number concentrations, and despite lower activated fractions of the terrestrial populations, differences in CCN concentrations between marine and terrestrial types (Figure 6b) were smaller than the differences in the activated fraction spectra. Only at supersaturations above approximately 0.5% did the CCN concentration for the terrestrial types (except T1, associated with many fresh, small particles) consistently exceed those of the marine types (Figure 6b). Between approximately 0.1% and 0.4%, CCN concentrations were often similar between the marine and terrestrial types.

3.4 Comparison of reconstructed and directly-measured CCN spectra

Average values for observations of the hygroscopicity parameter κ from the srCCN system are given for each cluster in Table 1. Mean κ values for the three marine population types were higher than for any of the terrestrial clusters, with the κ for the marine populations found to be significantly different ($p < 0.05$) from those for any of the terrestrial clusters.

Mean measured κ values of 0.49 and 0.46 for the M1 and M3 population types, respectively, ~~appeared~~were within the range of mean κ values (0.30 to ~~be an average of the~~ 0.54) for marine aerosol dominated periods in a coastal outflow marine ~~cases~~region presented by Phillips et al., (2018)~~, which had modal values of κ between 0.3 and 0.54. Mean values for terrestrial population types varied between 0.15 and 0.25, and were consistent with the 0.2 value for Aitken mode of continental aerosols reported by Phillips et al.~~. Mean values for terrestrial population types T1–T5 varied between 0.15 and 0.25, and were consistent with the 0.20 value for Aitken mode continental outflow aerosol reported by Phillips et al. (2018). However, we note that κ values measured in this study and used in the closure calculations were derived from supersaturated CCN measurements, whereas Phillips et al. ~~reported κ from humidified growth factors at 80% RH.~~(2018) reported κ from humidified growth factors at 80% RH. Prior work has shown that κ is not always consistent across this large of a span of RH (Irwin et al., 2010; Whitehead et al., 2014), thus contributing to uncertainty in such comparisons. The hygroscopicity for the final marine M2 type was 0.30, near the lower end of typical values for marine aerosol in regions of continental outflow, but still above those of the terrestrial population types. As the M2 cluster was also the only marine population type with no indication of recent precipitation scrubbing of the airmass prior to arrival at BML (Figure 3b), some combination of influences from cloud processing, marine, and terrestrial or anthropogenic sources may result in the observed hygroscopicity values between those of the other population types.

The cluster-average hygroscopicities from Table 1 were combined with the average cluster size distributions from Table 2 to create a reconstructed activated fraction spectrum for each cluster. These reconstructions are compared with the direct-

measured CCN spectra in Fig. S3 and Table 2. Generally, the reconstructed spectra are within one standard deviation of the directly measured spectra from the sfCCN system. However, the reconstruction overpredicts activated fraction for the marine clusters, with the largest discrepancies at low supersaturations and low CCN number concentrations. Similar behaviors in overprediction of CCN concentrations based on reconstructions using hygroscopicity and size distributions have been noted before (McFiggans et al., 2006). Kammermann et al., (2010) reviewed a number of studies that compared such predicted CCN concentrations against observed values and found biases were often largest at supersaturations below 0.3%, where predictions deviated from observations by factors ranging from 0.6 to 3.3, though with most studies finding overprediction occurred. They attributed the increasing bias at decreasing supersaturations to increased uncertainty in the critical activation diameter and associated CCN number prediction, though they also noted that this discrepancy between predicted and observed CCN concentrations has not been fully resolved. At larger particle diameters measurement uncertainty increases due to imprecise particle size cuts, losses in inlets and tubing, and inversion uncertainties, along with generally lower number concentrations than at smaller particle diameters, leading to higher expected uncertainty in CCN predictions and reconstructions when the critical activation diameter is in this range of particle sizes. At low supersaturations where BML data showed an overprediction bias the critical activation diameter would be above 150 nm. The marine types that had the largest biases at low supersaturations also tended to have larger fractions of particles at these large sizes. This would be expected to add to uncertainty in ways similar to those noted by Kammermann et al., ~~and potentially explain the larger discrepancies in CCN prediction at lower supersaturations in the marine aerosol types.~~(2010) and potentially explain the larger discrepancies in CCN prediction at lower supersaturations in the marine aerosol types.

Further investigation of the closure between predicted and observed CCN concentration was conducted using two prediction models. Hygroscopicity derived predictions using cluster average κ and normalized size distribution values were generated and compared against all observed CCN concentrations by the sfCCN system in Figure 7(a). Similarly, the predicted CCN concentration using cluster average activated fraction spectra were compared against observations in Figure 7(b). Both models predicted the activated fraction using the cluster type identified during the observation, which was then multiplied by the observed total number concentration at the observation time. The hygroscopicity and size model showed overprediction compared to both observations and the activated fraction model (Figure 7c), with best-fit slopes of 1.08 and 1.09 respectively. The activated fraction model predictions did improve on the hygroscopicity model, with a slope of 1.00 and R^2 values increasing from 0.87 to 0.89, though this is in part due to the model being based on a direct fit of the observed data. Nevertheless, the degree of closure between model predictions and observed CCN concentrations is similar to closures previously reported in field studies (e.g. Bougiatioti et al., 2009; Kammermann et al., 2010).

3.5 Aerosol optical properties

While optical properties of the various aerosol population types were not directly measured, a simple optical reconstruction was conducted to evaluate potential differences between the population types due to differences in size distribution and hygroscopicity. Average particle size distributions and measured average κ values were used to grow particles to equilibrium

with relative humidity across a range of values between 0% and 99%, assuming deliquesced particles, followed by estimation of mass scattering efficiency on a dry aerosol mass basis using Mie theory (Bohren and Huffman, 1983). For the purposes of this simple optical comparison, supersaturated κ values used for this analysis were assumed to be sufficient to provide an estimate of subsaturated hygroscopic growth. The average size distribution for each aerosol population type was assumed to be made for a deliquesced aerosol at the study average SMPS measured RH of 35.5% (Martin et al., 2017) and the cluster average κ given in Table 1. An assumed dry index of refraction of 1.5 + 0.0i and density of 1.0 g cm^{-3} (e.g. Remer et al., 2006) were used for all population types in order to estimate the relative effect of differences in size distribution, hygroscopicity, and relative humidity on scattering properties at a wavelength of 550 nm. The index of refraction of each humidified particle was adjusted based on volume mixing with water (m = 1.33 + 0.0i). Reconstructed mass scattering efficiencies for each of the population types are shown in Figure 8.

~~Computed dry mass~~Mass scattering efficiencies at 550 nm for particles modeled at 0% RH ranged between ~~4.2~~3.6 and 7.~~7~~6 m^2 g^{-1}, although actual values would be expected to be lower due to expected particle dry densities higher than 1 g cm^{-3}. Further, these values represent only the contribution to mass scattering efficiencies from ~~the sub-micron~~fine mode aerosol~~, whereas super-micron~~. Coarse mode aerosol, including particles generated by sea spray in littoral environments, can represent a large fraction of the total light scattering~~.~~, although their mass scattering efficiencies are typically lower. The M1 and M3 marine types, which included the largest fraction of larger accumulation mode particles, had the highest associated dry mass scattering efficiency. The M2 marine type, which occurred at generally lower wind speeds than the other marine types and thus had fewer larger particles associated with wind generated sea spray aerosol (O'Dowd and Leeuw, 2007), had a lower dry mass scattering efficiency than several of the terrestrial types with the largest median particle diameters. However, at relative humidity values above roughly 60%, as would typically be expected in littoral environments such as BML, the marine population types all yielded expected mass scattering efficiencies above those of the terrestrial types. For the marine types at a relative humidity of 95%, the mass scattering efficiencies on a unit-density dry mass basis were between ~~34 and 49 m^2 g^{-1}, roughly twice the range of the terrestrial types, 16 to 25 m^2 g^{-1}. As an additional point of comparison, similar reconstructions with MODIS fine mode ocean aerosol populations (Remer et al., 2006) yielded dry mass scattering efficiencies between 3.4 and 5.4 m^2 g^{-1}, and 33 and 54 m^2 g^{-1} at 95% RH, indicating the marine types were within the same range as assumptions used for MODIS marine aerosol populations~~31 and 47 m^2 g^{-1}, roughly twice the range of the terrestrial types, 14 to 21 m^2 g^{-1}.

4 Summary

The unsupervised cluster model analysis successfully identified distinct aerosol population types in the littoral zone aerosol at BML during CalWater-2015. The time periods selected by each cluster tended to be both temporally and physically coherent. Clusters also tended to be grouped into periods with physically meaningful microphysical properties that could be associated with meteorological processes and expected sources and transport pathways. For example, the clustering methodology identified regular diurnal swings in aerosol properties associated with land/sea-breeze changes and assigned two distinct,

terrestrially-influenced aerosol types during these periods. Overall, the clustering results for the CalWater-2015 dataset produced a reliable set of aerosol population types, and appropriately screened intermittent periods of impacts from various other sources as an outcome of the classification. Both marine and terrestrially influenced aerosol population types were identified by the unsupervised cluster model. Several marine events that persisted for days were identified as distinct in character from each other—differing in the degree of cloud processing and precipitation removal prior to arrival at the measurement site, and in the extent to which high winds contributed larger sea spray particles. About 10% of the observations were associated with a terrestrial population with a large fraction of small particles, indicating it was affected by relatively fresh emissions and/or new particle formation.

A primary motivation for CalWater-2015 was improving characterization of the regional aerosol and how it might affect the formation of precipitation. The CCN activation spectra observed at BML spanned a full range reported in the literature, from clean marine to strongly terrestrially influenced in character. However, differences in total aerosol number concentrations associated with the marine and terrestrial types partially offset the differences in activated fraction over a range of measured supersaturations, such that the variability in CCN concentrations between some marine and terrestrial aerosol types at some supersaturations was less than expected from the differences in the averaged total aerosol particle concentrations.

In this littoral region sea-breeze resampling and complex mixing between marine, terrestrial, and free-tropospheric air masses lead to complex aerosol populations. Determination of which aerosol types can become incorporated into AR events and thereby affect cloud and precipitation over California depends strongly on the individual nature of the flow regimes and potential aerosol source regions that they tap into. Nevertheless, the observational data set had cases for which marine and terrestrial CCN concentrations were comparable at supersaturations below approximately 0.4%, and thus changes in aerosol types would result in little change to initial drop populations in the resulting stratus cloud, except at very low supersaturations. Thus, mixed-phase microphysical processes occurring in those clouds might also be expected to be similar whether marine or terrestrial aerosols served as the nuclei for the supercooled droplets. In contrast, for clouds forming at higher maximum supersaturations, terrestrial aerosol populations are expected to yield higher drop concentrations than marine types, with the exception of a terrestrial population characterized by primarily small, recently emitted or newly formed particles. Thus the droplet size distributions formed on terrestrial vs. marine types, for liquid clouds formed in stronger updrafts or at higher cooling rates that reach such higher supersaturations, should be distinctly different.

At the BML observational site, apparent aerosol types changed over much shorter time scales than a multi-day classification based on prevailing meteorological conditions would suggest, consistent with complex flows in coastal zones. Inclusion of both meteorological and aerosol properties into the schemes used in unsupervised cluster models therefore yielded improvements in classification of aerosol and air mass types observed at the surface, and successfully identified changes at time resolutions on the order of hours. Further work into classification of aerosol populations in complex or highly variable regions may therefore benefit from inclusion of a wider range of measurements or observed properties into clustering methods.

Data availability

Measurement data for the scanning flow CCN used in this study are available at doi:10.25675/10217/180097.
Additional experimental data used in the analysis and presented in the figures are provided in an online data repository at doi:10.5281/zenodo.~~XXXXXXX (available upon publication).~~2605668.

Author Contributions

Samuel Atwood performed the sfCCN system investigation and data curation, along with conceptualization, methodology, formal analysis, and writing for this work. Sonia Kreidenweis was involved with writing the manuscript and, along with Paul DeMott, conducted conceptualization, funding acquisition, administration, and supervision of the project. Markus Petters performed the srCCN system investigation at BML, along with methodology development for the CCN systems, and formal analysis for the srCCN results. Andrew Martin assisted with review and editing of the manuscript, and along with Gavin Cornwell, assisted with investigation and data curation at the BML sampling location. Kathryn Moore performed the SMPS system investigation and assisted with sfCCN system investigation at BML.

Acknowledgements

This material is based upon research by the Office of Naval Research under Award Number N00014-16-1-2040. This work was supported by NSF award number 1450690 (MDP), NSF award number 1450760 (SAA, SMK, PJD), and NSF award number 1451347 (GCC, ACM, KAM). Assistance in operation of the research site was provided by Nicholas E. Rothfuss, Hans Taylor, Ezra Levin, Christina McCluskey, Yvonne Boose, Gregg Schill, Camille Sultana, and Kim Prather. The assistance by BML staff and Bodega Marine Reserve for lending laboratory space, assisting with measurement site improvements, and permission to measure on Reserve land is gratefully acknowledged. The loan of a CCN instrument by Jeff Reid and the U.S. Naval Research Laboratory, Monterey is also gratefully acknowledged.
We would also like to thank the two anonymous peer reviewers for their helpful suggestions and insights.

References

Andreae, M. O. and Rosenfeld, D.: Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Sci. Rev., 89(1–2), 13–41, doi:10.1016/j.earscirev.2008.03.001, 2008.

Atwood, S. A., Reid, J. S., Kreidenweis, S. M., Blake, D. R., Jonsson, H. H., Lagrosas, N. D., Xian, P., Reid, E. A., Sessions, W. R. and Simpas, J. B.: Size-resolved aerosol and cloud condensation nuclei (CCN) properties in the remote marine South China Sea – Part 1: Observations and source classification, Atmos Chem Phys, 17(2), 1105–1123, doi:10.5194/acp-17-1105-2017, 2017.

Baarsch, J. and Celebi, M. E.: Investigation of internal validity measures for K-means clustering, in Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, pp. 14–16. [online] Available from: http://www.iaeng.org/publication/IMECS2012/IMECS2012_pp471-476.pdf (Accessed ~~20~~19 November 2015), 2012.

Bates, T. S., Quinn, P. K., Covert, D. S., Coffman, D. J., Johnson, J. E. and Wiedensohler, A.: Aerosol physical properties and processes in the lower marine boundary layer: a comparison of shipboard sub-micron data from ACE-1 and ACE-2, Tellus B, 52(2), 258–272, doi:10.1034/j.1600-0889.2000.00021.x, 2000.

Bates, T. S., Quinn, P. K., Frossard, A. A., Russell, L. M., Hakala, J., Petäjä, T., Kulmala, M., Covert, D. S., Cappa, C. D., Li, S.-M., Hayden, K. L., Nuaaman, I., McLaren, R., Massoli, P., Canagaratna, M. R., Onasch, T. B., Sueper, D., Worsnop, D. R. and Keene, W. C.: Measurements of ocean derived aerosol off the coast of California, J. Geophys. Res. Atmospheres, 117(D21), D00V15, doi:10.1029/2012JD017588, 2012.

Beddows, D. C. S., Dall'Osto, M. and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle Size Data, Environ. Sci. Technol., 43(13), 4694–4700, doi:10.1021/es803121t, 2009.

Bohren, C. F. and Huffman, D. R.: Absorption and scattering by a sphere, Absorpt. Scatt. Light Small Part., 82–129, 1983.

Bougiatioti, A., Fountoukis, C., Kalivitis, N., Pandis, S. N., Nenes, A. and Mihalopoulos, N.: Cloud condensation nuclei measurements in the marine boundary layer of the Eastern Mediterranean: CCN closure and droplet growth kinetics, Atmos Chem Phys, 9(18), 7053–7066, doi:10.5194/acp-9-7053-2009, 2009.

Charron, A., Birmili, W. and Harrison, R. M.: Fingerprinting particle origins according to their size distribution at a UK rural site, J. Geophys. Res. Atmospheres, 113(D7), D07202, doi:10.1029/2007JD008562, 2008.

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., Cayan, D. R., Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J. and Cayan, D. R.: Atmospheric Rivers, Floods and the Water Resources of California, Water, 3(2), 445–478, doi:10.3390/w3020445, 2011.

Draxler, R. R. and Hess, G. D.: Description of the HYSPLIT4 modeling system, [online] Available from: http://warn.arl.noaa.gov/documents/reports/arl-224.pdf (Accessed 14 April 2015), 1997.

Draxler, R. R. and Hess, G. D.: An overview of the HYSPLIT_4 modelling system for trajectories, Aust. Meteorol. Mag., 47(4), 295–308, 1998.

Draxler, R. R., Stunder, B., Rolph, G. and Taylor, A.: HYSPLIT4 user's guide, NOAA Tech. Memo. ERL ARL, 230, 35, 1999.

Finlayson-Pitts, B. J. and Pitts, J. N.: Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications, Elsevier., 1999.

Forestieri, S. D., Cornwell, G. C., Helgestad, T. M., Moore, K. A., Lee, C., Novak, G. A., Sultana, C. M., Wang, X., Bertram, T. H., Prather, K. A. and Cappa, C. D.: Linking variations in sea spray aerosol particle hygroscopicity to composition during two microcosm experiments, Atmos Chem Phys, 16(14), 9003–9018, doi:10.5194/acp-16-9003-2016, 2016.

Frossard, A. A., Russell, L. M., Burrows, S. M., Elliott, S. M., Bates, T. S. and Quinn, P. K.: Sources and composition of submicron organic mass in marine aerosol particles, J. Geophys. Res. Atmospheres, 119(22), 2014JD021913, doi:10.1002/2014JD021913, 2014.

Good, N., Topping, D. O., Allan, J. D., Flynn, M., Fuentes, E., Irwin, M., Williams, P. I., Coe, H. and McFiggans, G.: Consistency between parameterisations of aerosol hygroscopicity and CCN activity during the RHaMBLe discovery cruise, Atmos Chem Phys, 10(7), 3189–3203, doi:10.5194/acp-10-3189-2010, 2010.

Hoppel, W. A., Frick, G. M. and Larson, R. E.: Effect of nonprecipitating clouds on the aerosol size distribution in the marine boundary layer, Geophys. Res. Lett., 13(2), 125–128, doi:10.1029/GL013i002p00125, 1986.

Hussein, T., Dal Maso, M., Petäjä, T., Koponen, I. K., Paatero, P., Aalto, P. P., Hämeri, K. and Kulmala, M.: Evaluation of an automatic algorithm for fitting the particle number size distributions, Boreal Environ. Res., 10(5), 337–355, 2005.

Irwin, M., Good, N., Crosier, J., Choularton, T. W. and McFiggans, G.: Reconciliation of measurements of hygroscopic growth and critical supersaturation of aerosol particles in central Germany, Atmospheric Chem. Phys., 10(23), 11737–11752, doi:https://doi.org/10.5194/acp-10-11737-2010, 2010.

Kammermann, L., Gysel, M., Weingartner, E., Herich, H., Cziczo, D. J., Holst, T., Svenningsson, B., Arneth, A. and Baltensperger, U.: Subarctic atmospheric aerosol composition: 3. Measured and modeled properties of cloud condensation nuclei, J. Geophys. Res. Atmospheres, 115(D4), doi:10.1029/2009JD012447, 2010.

Keene, W. C., Maring, H., Maben, J. R., Kieber, D. J., Pszenny, A. A. P., Dahl, E. E., Izaguirre, M. A., Davis, A. J., Long, M. S., Zhou, X., Smoydzin, L. and Sander, R.: Chemical and physical characteristics of nascent aerosols produced by bursting bubbles at a model air-sea interface, J. Geophys. Res. Atmospheres, 112(D21), D21202, doi:10.1029/2007JD008464, 2007.

Leung, L. R.: ARM Cloud-Aerosol-Precipitation Experiment (ACAPEX) Field Campaign Report, DOE ARM Climate Research Facility, Pacific Northwest National Laboratory, Richland, WA. [online] Available from: https://www.osti.gov/biblio/1251152 (Accessed ~~16~~15 October 2018), 2016.

Martin, A. C., Cornwell, G. C., Atwood, S. A., Moore, K. A., Rothfuss, N. E., Taylor, H., DeMott, P. J., Kreidenweis, S. M., Petters, M. D. and Prather, K. A.: Transport of pollution to a remote coastal site during gap flow from California's interior: impacts on aerosol composition, clouds, and radiative balance, Atmos Chem Phys, 17(2), 1491–1509, doi:10.5194/acp-17-1491-2017, 2017.

McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M. C., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A., Lohmann, U., Mentel, T. F., Murphy, D. M., O'Dowd, C. D., Snider, J. R. and Weingartner, E.: The effect of physical and chemical aerosol properties on warm cloud droplet activation, Atmos Chem Phys, 6(9), 2593–2649, doi:10.5194/acp-6-2593-2006, 2006.

Moore, R. H. and Nenes, A.: Scanning Flow CCN Analysis—A Method for Fast Measurements of CCN Spectra, Aerosol Sci. Technol., 43(12), 1192–1207, doi:10.1080/02786820903289780, 2009.

Nair, V. S., Moorthy, K. K. and Babu, S. S.: Influence of continental outflow and ocean biogeochemistry on the distribution of fine and ultrafine particles in the marine atmospheric boundary layer over Arabian Sea and Bay of Bengal, J. Geophys. Res. Atmospheres, 118(13), 7321–7331, doi:10.1002/jgrd.50541, 2013.

Neiman, P. J., Hughes, M., Moore, B. J., Ralph, F. M. and Sukovich, E. M.: Sierra Barrier Jets, Atmospheric Rivers, and Precipitation Characteristics in Northern California: A Composite Perspective Based on a Network of Wind Profilers, Mon. Weather Rev., 141(12), 4211–4233, doi:10.1175/MWR-D-13-00112.1, 2013.

O'Dowd, C. D. and Leeuw, G. de: Marine aerosol production: a review of the current knowledge, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 365(1856), 1753–1774, doi:10.1098/rsta.2007.2043, 2007.

Paramonov, M., Kerminen, V.-M., Gysel, M., Aalto, P. P., Andreae, M. O., Asmi, E., Baltensperger, U., Bougiatioti, A., Brus, D., Frank, G. P., Good, N., Gunthe, S. S., Hao, L., Irwin, M., Jaatinen, A., Jurányi, Z., King, S. M., Kortelainen, A., Kristensson, A., Lihavainen, H., Kulmala, M., Lohmann, U.,

Martin, S. T., McFiggans, G., Mihalopoulos, N., Nenes, A., O'Dowd, C. D., Ovadnevaite, J., Petäjä, T., Pöschl, U., Roberts, G. C., Rose, D., Svenningsson, B., Swietlicki, E., Weingartner, E., Whitehead, J., Wiedensohler, A., Wittbom, C. and Sierau, B.: A synthesis of cloud condensation nuclei counter (CCNC) measurements within the EUCAARI network, Atmos Chem Phys, 15(21), 12211–12229, doi:10.5194/acp-15-12211-2015, 2015.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É.: Scikit-learn: Machine Learning in Python, J Mach Learn Res, 12, 2825–2830, 2011.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos Chem Phys, 7(8), 1961–1971, doi:10.5194/acp-7-1961-2007, 2007.

Petters, S. S. and Petters, M. D.: Surfactant effect on cloud condensation nuclei for two-component internally mixed aerosols, J. Geophys. Res. Atmospheres, 121(4), 1878–1895, doi:10.1002/2015JD024090, 2016.

Phillips, B. N., Royalty, T. M., Dawson, K. W., Reed, R., Petters, M. D. and Meskhidze, N.: Hygroscopicity- and Size-Resolved Measurements of Submicron Aerosol on the East Coast of the United States, J. Geophys. Res. Atmospheres, 123(3), 1826–1839, doi:10.1002/2017JD027702, 2018.

Prather, K. A., Bertram, T. H., Grassian, V. H., Deane, G. B., Stokes, M. D., DeMott, P. J., Aluwihare, L. I., Palenik, B. P., Azam, F., Seinfeld, J. H., Moffet, R. C., Molina, M. J., Cappa, C. D., Geiger, F. M., Roberts, G. C., Russell, L. M., Ault, A. P., Baltrusaitis, J., Collins, D. B., Corrigan, C. E., Cuadra-Rodriguez, L. A., Ebben, C. J., Forestieri, S. D., Guasco, T. L., Hersey, S. P., Kim, M. J., Lambert, W. F., Modini, R. L., Mui, W., Pedler, B. E., Ruppel, M. J., Ryder, O. S., Schoepp, N. G., Sullivan, R. C. and Zhao, D.: Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol, Proc. Natl. Acad. Sci., 110(19), 7550–7555, doi:10.1073/pnas.1300262110, 2013.

Quinn, P. K., Bates, T. S., Schulz, K. S., Coffman, D. J., Frossard, A. A., Russell, L. M., Keene, W. C. and Kieber, D. J.: Contribution of sea surface carbon pool to organic matter enrichment in sea spray aerosol, Nat. Geosci., 7(3), 228–232, doi:10.1038/ngeo2092, 2014.

Ralph, F. M., Neiman, P. J. and Wick, G. A.: Satellite and CALJET Aircraft Observations of Atmospheric Rivers over the Eastern North Pacific Ocean during the Winter of 1997/98, Mon. Weather Rev., 132(7), 1721–1745, doi:10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2, 2004.

Ralph, F. M., Prather, K. A., Cayan, D., Spackman, J. R., DeMott, P., Dettinger, M., Fairall, C., Leung, R., Rosenfeld, D., Rutledge, S., Waliser, D., White, A. B., Cordeira, J., Martin, A., Helly, J. and Intrieri, J.: CalWater Field Studies Designed to Quantify the Roles of Atmospheric Rivers and Aerosols in

Modulating U.S. West Coast Precipitation in a Changing Climate, Bull. Am. Meteorol. Soc., 97(7), 1209–1228, doi:10.1175/BAMS-D-14-00043.1, 2015.

Remer, L. A., Tanre, D., Kaufman, Y. J., Levy, R. and Mattoo, S.: Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 005, Natl. Aeronaut. Space Adm., 1490 [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.385.6530&rep=rep1&type=pdf (Accessed 8 January 2017), 2006.

Rosenfeld, D., Lohmann, U., Raga, G. B., O'Dowd, C. D., Kulmala, M., Fuzzi, S., Reissell, A. and Andreae, M. O.: Flood or Drought: How Do Aerosols Affect Precipitation?, Science, 321(5894), 1309–1313, doi:10.1126/science.1160606, 2008.

Royalty, T. M., Phillips, B. N., Dawson, K. W., Reed, R., Meskhidze, N. and Petters, M. D.: Aerosol Properties Observed in the Subtropical North Pacific Boundary Layer, J. Geophys. Res. Atmospheres, 122(18), 9990–10,012, doi:10.1002/2017JD026897, 2017.

Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., Clifford, S. and Morawska, L.: Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment, Atmos Chem Phys, 14(21), 11883–11892, doi:10.5194/acp-14-11883-2014, 2014.

Song, C. H. and Carmichael, G. R.: The aging process of naturally emitted aerosol (sea-salt and mineral aerosol) during long range transport, Atmos. Environ., 33(14), 2203–2218, doi:10.1016/S1352-2310(98)00301-X, 1999.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D. and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bull. Am. Meteorol. Soc., 96(12), 2059–2077, doi:10.1175/BAMS-D-14-00110.1, 2015.

Suda, S. R., Petters, M. D., Matsunaga, A., Sullivan, R. C., Ziemann, P. J. and Kreidenweis, S. M.: Hygroscopicity frequency distributions of secondary organic aerosols, J. Geophys. Res. Atmospheres, 117(D4), D04207, doi:10.1029/2011JD016823, 2012.

Suda, S. R., Petters, M. D., Yeh, G. K., Strollo, C., Matsunaga, A., Faulhaber, A., Ziemann, P. J., Prenni, A. J., Carrico, C. M., Sullivan, R. C. and Kreidenweis, S. M.: Influence of Functional Groups on Organic Aerosol Cloud Condensation Nucleus Activity, Environ. Sci. Technol., 48(17), 10182–10190, doi:10.1021/es502147y, 2014.

Tunved, P., Ström, J. and Hansson, H.-C.: An investigation of processes controlling the evolution of the boundary layer aerosol size distribution properties at the Swedish background station Aspvreten, Atmos Chem Phys, 4(11/12), 2581–2592, doi:10.5194/acp-4-2581-2004, 2004.

Wegner, T., Hussein, T., Hämeri, K., Vesala, T., Kulmala, M. and Weber, S.: Properties of aerosol signature size distributions in the urban environment as derived by cluster analysis, Atmos. Environ., 61, 350–360, doi:10.1016/j.atmosenv.2012.07.048, 2012.

Wex, H., Dieckmann, K., Roberts, G. C., Conrath, T., Izaguirre, M. A., Hartmann, S., Herenz, P., Schäfer, M., Ditas, F., Schmeissner, T., Henning, S., Wehner, B., Siebert, H. and Stratmann, F.: Aerosol arriving on the Caribbean island of Barbados: physical properties and origin, Atmos Chem Phys, 16(22), 14107–14130, doi:10.5194/acp-16-14107-2016, 2016.

White, A. B., Anderson, M. L., Dettinger, M. D., Ralph, F. M., Hinojosa, A., Cayan, D. R., Hartman, R. K., Reynolds, D. W., Johnson, L. E., Schneider, T. L., Cifelli, R., Toth, Z., Gutman, S. I., King, C. W., Gehrke, F., Johnston, P. E., Walls, C., Mann, D., Gottas, D. J. and Coleman, T.: A Twenty-First-Century California Observing Network for Monitoring Extreme Weather Events, J. Atmospheric Ocean. Technol., 30(8), 1585–1603, doi:10.1175/JTECH-D-12-00217.1, 2013.

Whitehead, J. D., Irwin, M., Allan, J. D., Good, N. and McFiggans, G.: A meta-analysis of particle water uptake reconciliation studies, Atmospheric Chem. Phys., 14(21), 11833–11841, doi:https://doi.org/10.5194/acp-14-11833-2014, 2014.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press., 2011.

Zhang, X., Massoli, P., Quinn, P. K., Bates, T. S. and Cappa, C. D.: Hygroscopic growth of submicron and supermicron aerosols in the marine boundary layer, J. Geophys. Res. Atmospheres, 119(13), 2013JD021213, doi:10.1002/2013JD021213, 2014.

Zhao, Y., Zhang, Y., Fu, P., Ho, S. S. H., Ho, K. F., Liu, F., Zou, S., Wang, S. and Lai, S.: Non-polar organic compounds in marine aerosols over the northern South China Sea: Influence of continental outflow, Chemosphere, 153, 332–339, doi:10.1016/j.chemosphere.2016.03.069, 2016.

Tables

Table 1 Aerosol and meteorological parameters for each of the cluster time periods~~.~~ with the total number and percentage of observations in each cluster. Cluster mean values are given for total particle number concentration, κ hygroscopicity parameter from the srCCN system, HYSPLIT 24-hour accumulated precipitation along the trajectory, and local wind velocity observations. ~~Best-fit size distribution and activated fraction parameters are shown, with activated fraction parameters pertaining to the fit model given in the supplemental information.~~

BML Cluster	Percentage of Observations	Total Number Concentration	κ	Wind Velocity		Trajectory 24-hr
			Mean	u	v	Accum Precip
	% (total N)	(# cm^{-3}) (std dev)	(-) (std dev)	(m/s)	(m/s)	(mm) (std dev)
M1	5% (169)	592 (435)	0.49 (0.23)	6.76	-4.22	0.31 (0.73)
M2	9% (308)	774 (462)	0.30 (0.10)	2.77	1.16	0.07 (0.30)
M3	10% (345)	1547 (1177)	0.46 (0.22)	-2.63	5.70	1.02 (1.76)
T1	9% (304)	1975 (2166)	0.25 (0.17)	0.25	4.12	0.51 (1.21)
T2	17% (566)	3189 (2055)	0.18 (0.07)	-1.90	2.70	0.16 (0.70)
T3	16% (545)	3816 (3447)	0.15 (0.06)	-0.70	2.25	0.03 (0.21)
T4	20% (659)	2306 (1166)	0.17 (0.06)	0.46	2.49	0.01 (0.04)
T5	14% (461)	1459 (600)	0.20 (0.07)	2.72	0.67	0.00 (0.00)

~~**Table 2 Cluster best-fit size distribution and activated fraction parameters are shown, with activated fraction parameters pertaining to the fit model given in the supplemental information.**~~

Table 2 Cluster best-fit number size distribution and activated fraction parameters are shown, with activated fraction parameters a, b, and c pertaining to the fit model equation given in the supplemental information. Clusters with "M" and "T" names refer to marine and terrestrial aerosol population types, respectively.

BML Cluster	Avg Size Dist - Mode 1			Avg Size Dist - Mode 2			Avg Size Dist - Mode 3			Activation Specrum Fit Mode 1			Activation Specrum Fit Mode 2		
	Median (nm)	Geometric Std Dev	Number Fraction	Median (nm)	Geometric Std Dev	Number Fraction	Median (nm)	Geometric Std Dev	Number Fraction	a	b	c	a	b	c
M1	45	1.77	0.69	189	1.54	0.28	561	1.22	0.03	0.12	1.34	0.09	0.38	2.72	0.71
M2	21	1.46	0.07	67	1.47	0.39	185	1.45	0.54	0.1399	1.5613	0.60	0.4331	1.4379	0.30
M3	40	1.99	0.69	182	1.59	0.32	--	--	--	0.1246	1.299	0.09	0.266	2.8918	0.68
T1	28	1.78	0.9	144	1.51	0.10	--	--	--	0.1619	2.0963	0.09	1.528	3	0.57
T2	39	2	0.88	170	1.49	0.12	--	--	--	0.4686	1.2076	0.02	0.5111	2.8612	0.63
T3	59	1.8	0.95	211	1.42	0.04	--	--	--	0.4567	1.2878	0.08	0.63	2.6407	0.70
T4	12	1.72	0.1	76	2	0.90	--	--	--	0.38	2.3426	0.70	0.4228	1.2607	0.08
T5	15	1.53	0.06	94	1.85	0.94	--	--	--	0.28	2.16	0.75	0.42	1.28	0.15

Figures

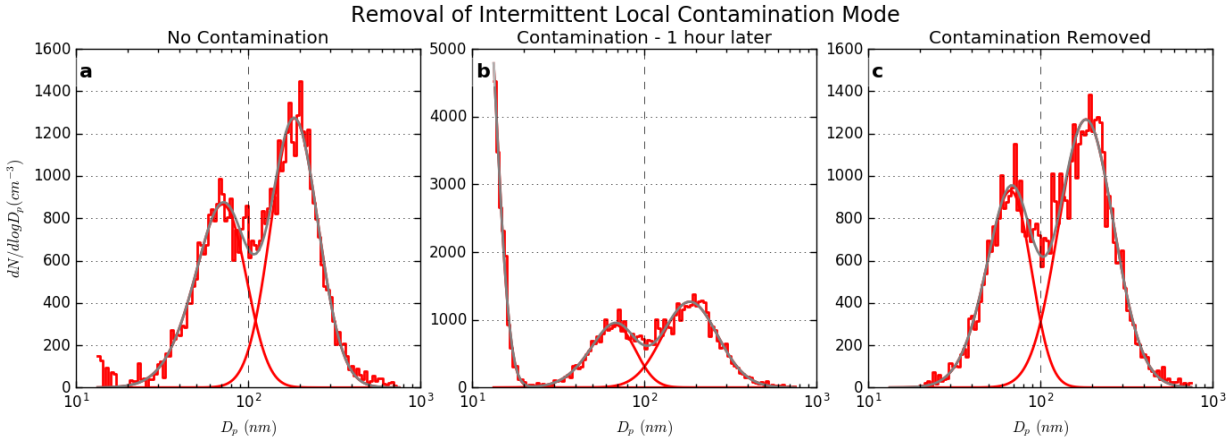


Figure 1 Example removal of intermittent local contamination mode. (a) An observed size distribution with limited impacts from the smallest nucleation mode particles. (b) An observation approximately 1 hour later with contamination from small particles that are not representative of the regional aerosol. (c) The same size distribution as in (b), but with the smallest contamination mode identified and removed as described in the text.

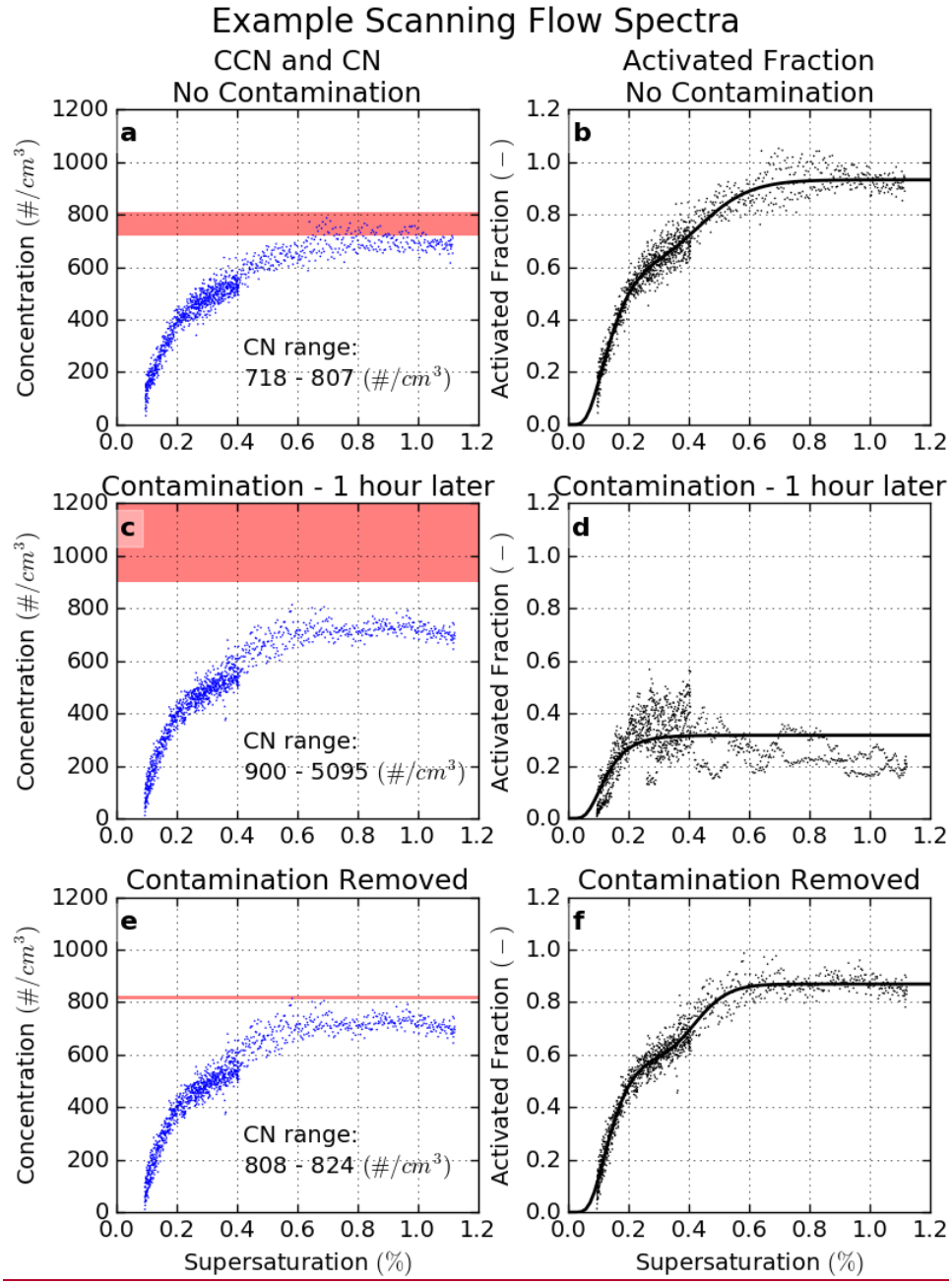


Figure 2 Example spectra measured by the sfCCN system for the same time period as in Figure 1. (a) One second CCN (blue) and the range of CN (red) concentration measurements for a single scan. ~~CCN are shown at the calibrated supersaturations, while CN as measured by the CPC are independent of supersaturation but are shown at 1.3% for comparison.~~ (b) Activated fractions for measured points in (a) along with a best-fit activated fraction spectrum curve (black line). (c) & (d) As in (a) and (b), but for a scan approximately one hour later that was contaminated by an intermittent local ultrafine mode; CN were observed as high as 5000 cm^{-3}. (e) and (f) The same spectra as in (c) and (d), but after correction to remove the contamination mode; corrected CN concentrations are now similar to the range in (a).

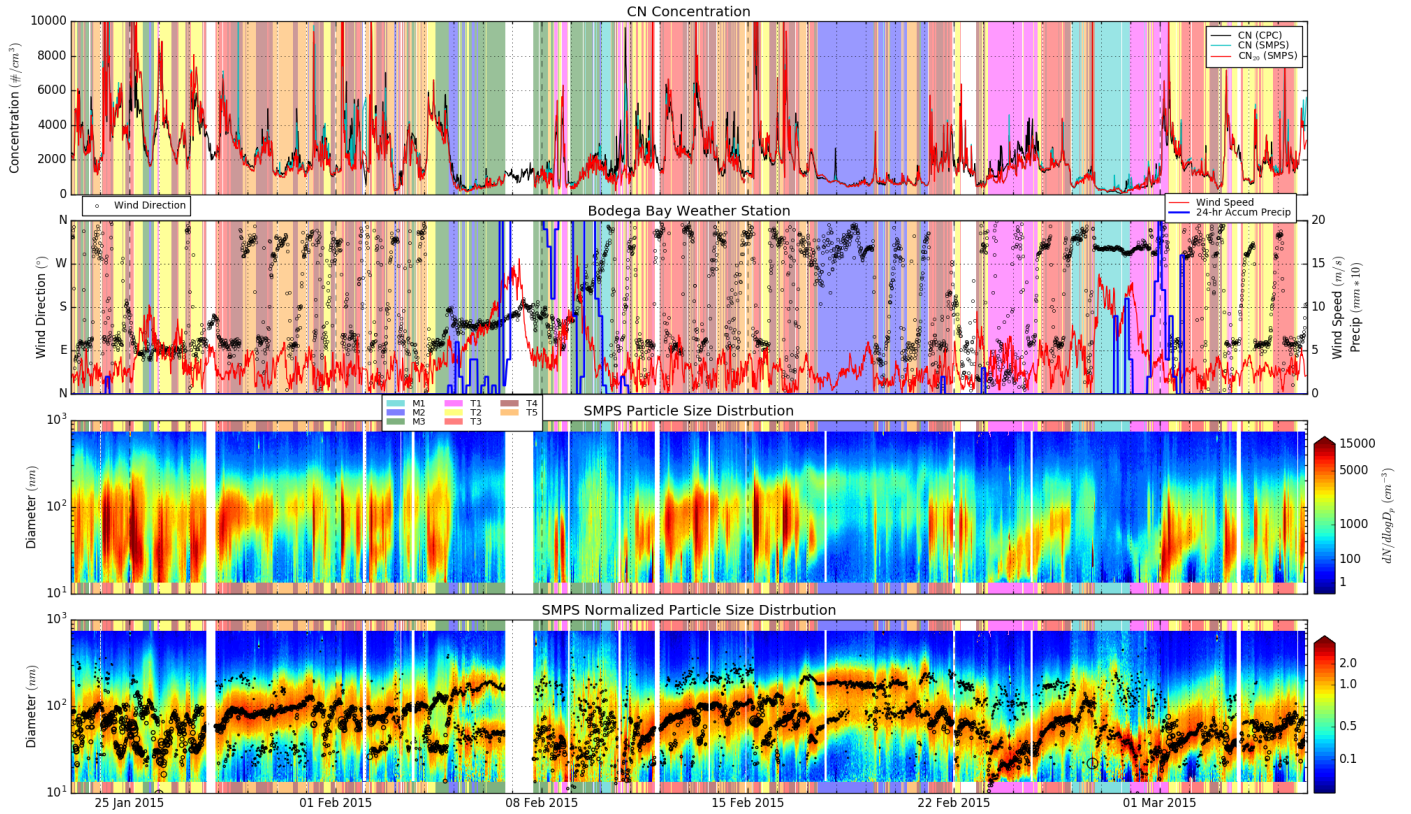


Figure 3 Timelines of study data; all times shown are UTC. (a) Total particle number concentration at BML measured by the CPC (black), reconstructed from SMPS size distributions (blue), and after correction for removal of local sources (red). (b) BML weather station measurements of wind direction (black circles) and wind speed (red lines), with HYSPLIT 24-hour accumulated precipitation along the airmass trajectory (blue). (c) SMPS size distribution as $dN/d\log_{10}D_p$. and (d) normalized size distributions with best fit median diameters (black circles) with the size of the circle proportional to the number of particles in the mode. Colored backgrounds in each panel are shown for time periods classified as each cluster type for the 8-cluster K-means classification.

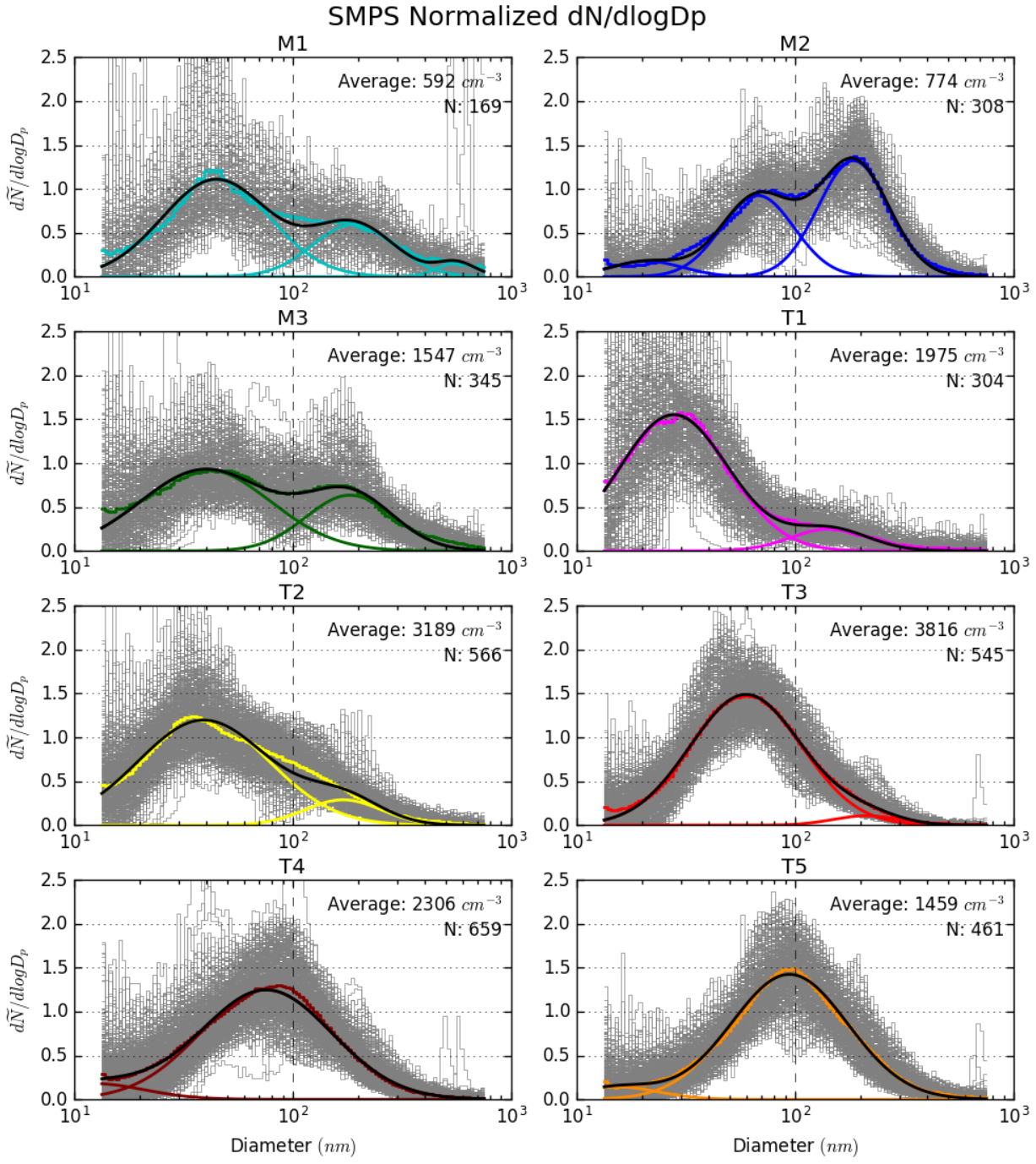
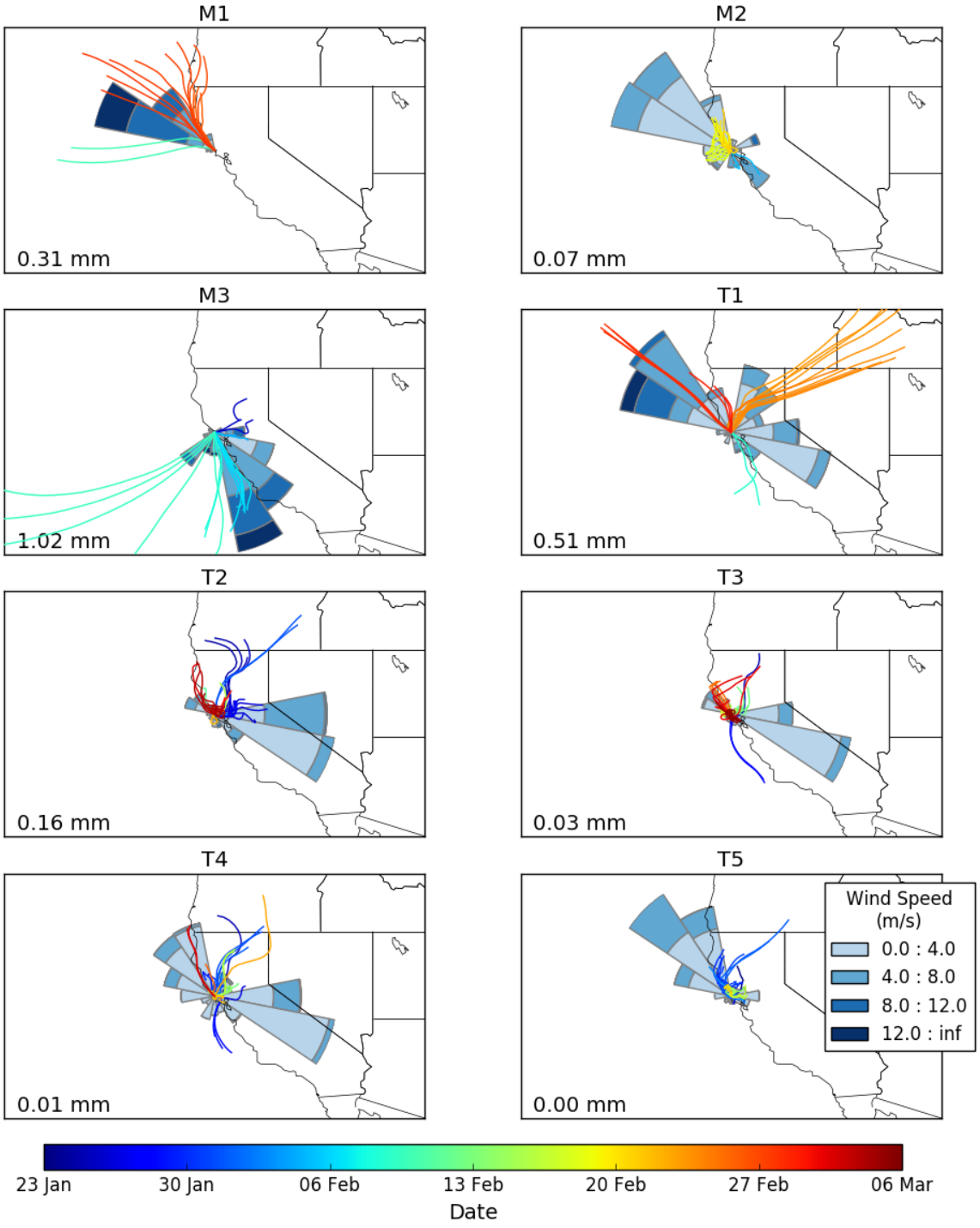


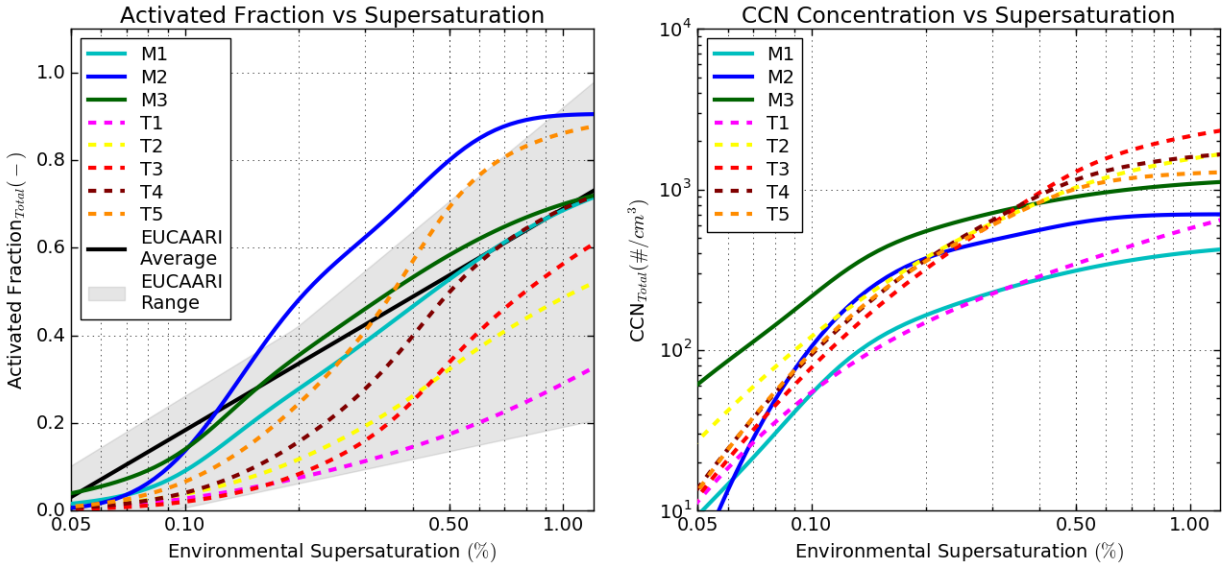
Figure 4 Normalized size distribution for each cluster from the 8 cluster K-means classification. Colored step lines show the average distribution, with best-fit multimodal lognormal fit in black and each mode as colored curves. Observed spectra for each data point in the cluster are shown in grey. Average cluster total particle number concentration and number of data points in each cluster are shown.


Wind Vector and 100m HYSPLIT backtrajectories
M1
0.31 mm
M2
0.07 mm
M3
1.02 mm
T1
0.51 mm
T2
0.16 mm
T3
0.03 mm
T4
0.01 mm
T5
0.00 mm
Wind Speed (m/s)
0.0 : 4.0
4.0 : 8.0
8.0 : 12.0
12.0 : inf
23 Jan
30 Jan
06 Feb
13 Feb
20 Feb
27 Feb
06 Mar
Date

1

Figure 5 Meteorological overview for each cluster from the 8 cluster K-means classification. HYSPLIT 24-hr backtrajectories are shown for each time stamp associated with the cluster and colored by the date of arrival at the receptor. Wind rose plots are shown for all BML local 10m wind direction and velocity measurements for each time stamp associated with the cluster. Mean values for 24-hour accumulated precipitation along the trajectory are included in the lower left for each cluster.

1



2

Figure 6 (a) Best-fit activated fraction spectra for each cluster in the 8 cluster K-means analysis, with marine aerosol population types plotted as solid lines, and terrestrial clusters types as dotted lines. Cluster average spectra are compared against the maximum and minimum values (grey shading) and overall average (black line) reported by Paramonov et al., (2015) from a range of sampling locations in the European EUCAARI network. (b) Parameterized mean CCN concentrations across the same range of supersaturation values for each cluster in the 8 cluster K-means analysis.

3

4

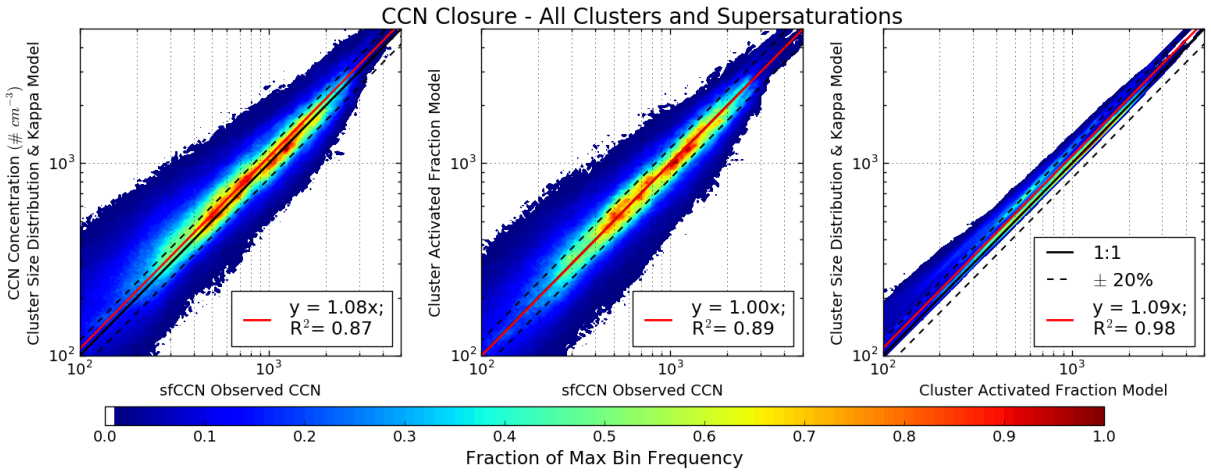


Figure 7 CCN closure comparing predicted and measured CCN number concentrations for (a) the predicted concentration based on the size distribution and ~~kappa~~κ reconstruction against observed concentrations, (b) the predicted concentration based on the best-fit activated fraction reconstruction against observed concentrations, and (c) the comparison between the two reconstruction methods. All cluster types and supersaturations across the full range of observed values are shown. Linear best-fit slope is shown in red with the associated R-squared values. One to one lines and lines at ± 20% are shown. Colors represent the data point density as a fraction of the maximum density in each plot.

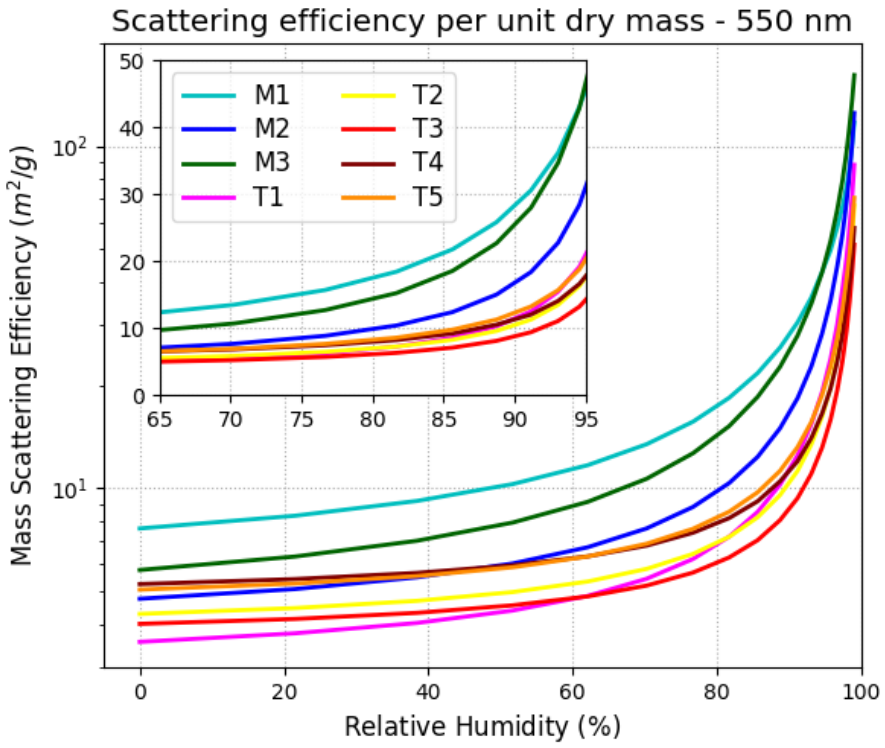


Figure 8 Reconstructed mass scattering efficiencies (per unit dry aerosol mass) for each of the cluster population types across a range of environmental relative humidity values. An assumed dry index of refraction of 1.5 + 0.0i was used for all population types to highlight the differences in aerosol optical properties expected due to differences in population average size distribution, hygroscopicity, and relative humidity.