



1 **Classification of aerosol population type and cloud condensation** 2 **nuclei properties in a coastal California littoral environment using an** 3 **unsupervised cluster model**

4 Samuel A. Atwood¹, Sonia M. Kreidenweis¹, Paul J. DeMott¹, Markus D. Petters², Gavin C. Cornwell³,
5 Andrew C. Martin⁴, Kathryn A. Moore^{1,3}

6 ¹Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA

7 ²Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA

8 ³Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA

9 ⁴Climate Atmospheric Science and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA, USA

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

11 **Abstract.** Aerosol particle and cloud condensation nuclei (CCN) measurements from a littoral location on the northern coast
12 of California at Bodega Bay Marine Laboratory (BML) are presented for approximately six weeks of observations during the
13 CalWater-2015 field campaign. A combination of aerosol microphysical and meteorological parameters was used to classify
14 variability in the properties of the BML surface aerosol using a K-means cluster model. Eight aerosol population types were
15 identified that were associated with a range of impacts from both marine and terrestrial sources. Average measured total
16 particle number concentrations, size distributions, hygroscopicities, and activated fraction spectra between 0.08% and 1.1%
17 supersaturation are given for each of the identified aerosol population types, along with meteorological observations and
18 transport pathways during time periods associated with each type. Five terrestrially influenced aerosol population types
19 represented different degrees of aging of the continental outflow from the coast and interior of California and their appearance
20 at the BML site was often linked to changes in wind direction and transport pathway. In particular, distinct aerosol populations,
21 associated with diurnal variations in source region induced by land/sea-breeze shifts, were classified by the clustering
22 technique. A terrestrial type representing fresh emissions, and/or a recent new particle formation event, occurred in
23 approximately 10% of the observations. Over the entire study period, three marine influenced population types were identified
24 that typically occurred when the regular diurnal land/sea-breeze cycle collapsed and BML was continuously ventilated by air
25 masses from marine regions for multiple days. These marine types differed from each other primarily in the degree of cloud
26 processing evident in the size distributions, and in the presence of an additional large-particle mode for the type associated
27 with the highest wind speeds. One of the marine types was associated with a multi-day period during which an atmospheric
28 river made landfall at BML. The generally higher total particle number concentrations but lower activated fractions of four of
29 the terrestrial types yielded similar CCN number concentrations to two of the marine types for supersaturations below about
30 0.4%. Despite quite different activated fraction spectra, the two remaining marine and terrestrial types had CCN spectral
31 number concentrations very similar to each other, due in part to higher number concentrations associated with the terrestrial
32 type.



1 1 Introduction

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

10 In marine regions impacted by continental outflow, aerosol chemical and microphysical properties, including particle number
11 concentrations and size distributions, are often moderated by impacts from terrestrial sources (Nair et al., 2013; Wex et al.,
12 2016; Zhao et al., 2016; Phillips et al., 2018). For example, freshly emitted sea spray aerosol particles comprise a mixture of
13 salts with generally high hygroscopicities ($\kappa \sim 0.6\text{--}1.2$), and co-emitted organic species with lower hygroscopicities ($\kappa \sim 0\text{--}$
14 0.3) (Prather et al., 2013; Quinn et al., 2014)—classified using the κ hygroscopicity parameter (Petters and Kreidenweis, 2007).
15 Bulk hygroscopicity values above 1 are infrequently observed, but have been reported for some background and precipitation
16 impacted marine aerosol populations (Good et al., 2010; Prather et al., 2013). As the aerosol ages, chlorine replacement by
17 uptake of acidic gases can reduce the hygroscopicity of the salts. Aging and organic components result in reported distribution-
18 averaged κ values in the range of 0.4 to 0.7 for more pristine or background marine aerosol populations (Keene et al., 2007;
19 Bates et al., 2012; Prather et al., 2013; Quinn et al., 2014; Zhang et al., 2014; Forestieri et al., 2016; Atwood et al., 2017;
20 Royalty et al., 2017; Phillips et al., 2018). In contrast, average hygroscopicities for continental organic aerosol populations are
21 typically in the range of 0.1 to 0.3 due to heavy influence of organic and insoluble material, although individual inorganic
22 sulfates and nitrate salts have values in the same range as many remote marine aerosol populations (Andreae and Rosenfeld,
23 2008). The cloud-droplet-nucleating activity of particles in coastal regions therefore depends on the composition, size
24 distribution, and degree of aging of the marine particles, along with the characteristics of particles from non-marine sources
25 that become mixed with the marine aerosol. Thus, observation and analysis of the coastal California aerosol environment will
26 support efforts to better understand implications of aerosol properties on precipitation and cloud development in the region,
27 including aerosol influences on how precipitation develops in landfalling ARs in this region.

28 In this study, observations from the BML ground site during the CalWater-2015 field campaign were used to evaluate the
29 nature and variability of surface (marine boundary layer, MBL) aerosol populations at BML. Observations were classified into
30 periods of similar aerosol and meteorological characteristics using an unsupervised cluster model (Atwood et al., 2017) to
31 derive distinct littoral aerosol population types and link them to source regions. The clusters were assessed with respect to
32 their aerosol characteristics, particularly the contributions of each population type to cloud condensation nuclei (CCN)
33 concentrations that influence the microphysical properties of liquid-phase clouds and the evolution of precipitation in mixed-



1 phase clouds (e.g., Rosenfeld et al., 2008). In prior work analyzing data from the CalWater-2015 / ACAPEX field experiments,
2 Leung (2016) discusses an AR event that was first observed and sampled from aircraft platforms off the coast of California on
3 5 February 2015, during the observational period analyzed here. The AR made landfall near BML on the morning of 6
4 February, and the AR-associated front, producing heavy precipitation along its trajectory, reached the Central Valley and
5 Sierra Nevada in the afternoon. We thus also contrast the aerosol cluster types identified at BML during AR and marine aerosol
6 dominated events to observations outside those periods.

7 **2 Methods and site description**

8 Measurements occurred at Bodega Bay Marine Laboratory (BML; 38.32° N, 123.07° W) and took place between January and
9 March 2015. Data utilized in this study were gathered between 23 January and 5 March, when relevant instruments were
10 operational. All aerosol measurements used here were made from within the Colorado State University/National Park Service
11 mobile lab, located approximately 200 m from the coast, and approximately 150 m north of the main BML lab building. Air
12 was sampled through several ½” stainless steel tube inlets routed through the roof of the mobile lab to a height of approximately
13 5 m, before the flow was split and sent to various instruments described further below. A map and additional description of
14 the site and instrumentation are provided in Martin et al., (2017).

15 **2.1 Meteorological data**

16 Local meteorological measurements were obtained from a 10 m surface met tower located approximately 100 m north of the
17 mobile lab, and operated as part of a NOAA/ESRL observation network (White et al., 2013). Air mass source regions were
18 defined using 24-hour backtrajectories that were initiated every three hours during the study period and computed using the
19 HYSPLIT version 4.9 Lagrangian parcel model (Draxler and Hess, 1997, 1998; Draxler et al., 1999; Stein et al., 2015) with
20 the 40 km x 40 km Edas40 meteorological dataset. All trajectories were generated with 100 m arrival heights at BML to
21 characterize transport in the MBL.

22 **2.2 Aerosol size distribution measurements (SMPS)**

23 Aerosol size distribution measurements made during CalWater-2015 at BML are described in detail in Martin et al., (2017).
24 This study used size distribution measurements for particle diameters between 14 and 730 nm from a TSI 3080 SMPS using a
25 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
26 and were subsequently averaged to approximately 15-minute time periods to match additional coincident measurements.
27 The presence of large numbers of particles smaller than approximately 30 nm in diameter was intermittently observed in the
28 size distribution measurements. These particles were likely from a combination of local sources that included vehicle and other
29 activity at BML, local camp and brush fires, and emissions from the nearby town of Bodega Bay, located to the east of the
30 site. This nucleation mode was generally superimposed on more stable aerosol populations, and thus rather than removing the



1 entire observation during these events, we removed only the contamination mode, as described below and in further depth in
2 the supplemental material.

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

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

14 In the case that an observed size distribution spectrum passed all these criteria, the smallest mode was classified as a
15 contamination mode likely associated with local sources as described above and therefore not representative of the regional
16 aerosol. These modes were removed from the size distributions and total number concentrations while retaining the remaining
17 fitted modes. This was accomplished by multiplying the observed $dN/d\log_{10}D_p$ value by one minus the contamination mode's
18 number fraction of the total distribution for each size bin. An example of the removal is shown in Fig. 1, with an observed
19 distribution without a local mode shown in (a). Approximately one hour later, a distinct small mode was observed with a
20 median diameter less than 14 nm and nearly no contribution to particle concentration at diameters above 20 nm (b), while total
21 number concentration showed a rapid increase (not shown). This mode was no longer present approximately 1.5 hours later.
22 Therefore, the observation passed all four criteria to be considered local contamination and was removed from the reported
23 size distribution and total number concentration to give the final corrected distribution (c).

24 2.3 CCN measurements

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

30 A separate scanning flow CCN system (sfCCN) was used to measure total CCN number concentration as supersaturation was
31 ramped between approximately 0.08% and 1.1% supersaturation. The system used a DMT CCN-100 instrument that had been
32 modified by connecting a voltage modulated proportional flow valve to the bottom of the column to control flow (Suda et al.,
33 2014). The flow rate through the CCN column was increased from 0.2 to 1.2 lpm over 5 minutes while holding the temperature



1 gradient constant, thereby scanning peak column supersaturation as a function of flow rate and column temperature gradient
2 (Moore and Nenes, 2009). A TSI 3010 Condensation Particle Counter (CPC) was placed in parallel to the CCN to measure
3 total particle number concentration (CN). CCN and CN concentrations were recorded at a frequency of one hertz. Each
4 approximately 5-minute scan was repeated three times, after which the temperature gradient was changed to scan a different
5 range of supersaturations. As the column temperatures took approximately 3 minutes to stabilize, the first scan of each three-
6 repetition set was not analyzed. As the residence time in the sfCCN varied with total flow rate when compared to the parallel
7 CPC, the CPC timestamps were empirically adjusted to ensure the 1 Hz data points from both instruments were aligned, as
8 described in more detail in the Supplemental Material. Calibrations of both CCN systems were conducted following the
9 methodology described in Suda et al., (2012).

10 Example sfCCN concentrations and activated fraction spectra are shown in Fig. 2(a&b), respectively, for the same time stamp
11 as the size distribution data in Fig. 1. The range of CN concentrations as measured by the parallel CPC during this period are
12 shown as red data points (placed for comparison to CCN concentrations at 1.3% supersaturation). During the contamination
13 period approximately one hour later (Fig. 2(c&d)) the effect of the small mode noted in Fig. 1(b) was seen via increased CN
14 concentrations (reaching as high as 5000 cm^{-3}) and decreased activated fractions above approximately 0.1% supersaturation.
15 Observed CCN concentrations remained relatively constant indicating the additional particles were too small to activate below
16 1.1% supersaturation. After removal of this small contamination mode the corrected CN concentrations and activation
17 spectrum Fig. 2(e&f) were similar in characteristics to the earlier non-contaminated period.

18 2.4 Classification of aerosol population type

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

27 The K-means clustering methodology involved selection of specific variables that partially defined the state of the aerosol and
28 meteorological environment at the sampling site. The degree of similarity of the state of the environment between any two
29 data points (i.e. specific times) was estimated by the use of a distance function that grouped data points into clusters that had
30 broadly similar values among the input variables. Here, we utilized the *cluster.KMeans* class of the Python scikit-learn package
31 (Pedregosa et al., 2011) to perform the analysis. Selection of the appropriate number of clusters followed a similar
32 methodology to that described in Atwood et al., (2017). A hierarchical cluster analysis was first created using the
33 *cluster.AgglomerativeClustering* class of scikit-learn to identify potential numbers of clusters using a dendrogram. Various



1 internal validity measures of the K-means clusters (Beddows et al., 2009; Baarsch and Celebi, 2012) were then used, along
2 with verification that the results maintained physically distinct and temporally coherent clusters, in order to select the
3 appropriate number of K-means clusters.

4 **2.4.1 Cluster variables**

5 Variables used in this analysis included measurements of aerosol microphysical properties and of meteorological parameters
6 at BML. Aerosol property variables included the normalized size distribution at each time stamp (normalized to an integrated
7 value of 1 cm^{-3} by dividing by the total particle number concentration), after correction for local contamination. The
8 distributions were discretized into 20 logarithmically spaced bins that served as separate variables in the clustering distance
9 function (e.g. Charron et al. (2008)). Activated fraction spectra for each time stamp were divided into 20 linearly spaced bins
10 distributed between 0% and 1.1% supersaturation to incorporate CCN properties into the analysis. Total particle number
11 concentration was included as a separate cluster variable.

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

17 **2.4.2 Distance function**

18 The Karl Pearson Euclidean distance function (Wilks, 2011) used in the cluster model was modified to include a relative
19 weight parameter for each input variable as:

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

20 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
21 independently measured or observed, orthogonal variables at each time stamp), for each variable, k , with relative weight, w_k .
22 Each variable was first standardized, with missing data imputed to values of zero to minimize their impact on the distance
23 function.

24 The relative weights for each variable were included to prevent properties of the aerosol from becoming over-weighted in the
25 cluster model due to having more variables describing them. The size distribution and activated fraction variables, each of
26 which had 20 variables, were given relative weights of $1/20$ such that their total relative weights summed to 1. As these
27 variables were of primary importance to aerosol population microphysical properties, the other variable groups were decreased
28 in relative importance. The backtrajectory and wind vector groups were each assigned a relative weight of 0.5, and the total
29 particle number concentration variable assigned a relative weight of 0.1. As cluster analysis is by its nature an exploratory data



1 analysis technique, these relative weight values were reached by varying their values and assessing the physical interpretability
2 of the results.

3 **2.4.3 Number of clusters**

4 Hierarchical clustering and internal validity measures indicated two, six, eight, and twelve clusters were potentially appropriate
5 for the K-means analysis. Clusters associated with periods of marine aerosol impacts (discussed further in the next section)
6 became temporally coherent and physically meaningful after the number of clusters was increased to eight.

7 In the case of the twelve-cluster analysis, several of the clusters were composed of few or even a single data point, indicating
8 the model had begun to separate outliers into distinct clusters. In addition, several of the temporally consistent clusters were
9 split, indicating that too many clusters had been selected. All potential cluster numbers between seven and eleven were then
10 investigated to determine if physically or temporally coherent population types emerged to a greater degree than the eight-
11 cluster model. As the eight-cluster option was initially identified as potentially appropriate, and the other potential models did
12 not improve physical interpretability of the results, the eight-cluster model was selected as the most appropriate unsupervised
13 classification result.

14 **3 Aerosol population type classification results and discussion**

15 Three of the eight identified clusters were defined as “marine” population types, as backtrajectory data showed evidence of
16 transport pathways primarily over ocean areas. These marine types, denoted as clusters M1–M3, tended to have lower average
17 number concentrations (below approximately 1500 cm^{-3}), while the terrestrial clusters (T1–T5) had typical averages between
18 approximately 2000 and 4000 cm^{-3} and were associated with transport from more terrestrial source regions. The exception to
19 this was cluster T5, which had number concentrations of roughly 1500 cm^{-3} , more oceanic transport pathways, and size
20 distributions with the largest median diameters among the “terrestrial” clusters. Table 1 provides the cluster-averaged number
21 concentrations, wind velocities, HYSPLIT accumulated precipitation along the 24-hour trajectory, hygroscopicity parameters
22 from the srCCN system, and the percentage of all measurements associated with each of the 8 identified clusters. Best-fit size
23 distribution parameters and best-fit CCN-spectrum activated fraction parameters (see Supplemental Material) for each cluster
24 are provided in Table 2.

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

29 Changes between identified cluster types often tracked diurnal changes in wind direction associated with the land/sea-breeze
30 cycle (Fig. 3b; e.g., 28–30 Jan). Several periods occurred during the study during which this diurnal cycle collapsed and BML
31 was ventilated with air masses from marine regions for several days at a time (starting on 4 Feb, 17 Feb, and 26 Feb). Clusters



1 M1, M2, and M3 were selected by the model more consistently during these extended periods and confirmed their
2 characterization as marine aerosol population types.

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

7 **3.1 Marine population types**

8 Further analysis of the marine clusters showed generally distinct meteorological conditions associated with each. Cluster M1
9 primarily occurred toward the end of the study period, during a period of high velocity winds from the northwest.
10 Backtrajectories agreed with local winds and showed generally faster transport velocities. This cluster dominated during 26
11 and 27 February, a period during which the cleanest air masses and lowest number concentrations of the study were observed,
12 reaching as low as 50 cm^{-3} during the height of the event. The normalized size distributions indicated that the Aitken mode
13 dominated the number distribution (Fig. 4), and also suggested the presence of a somewhat larger mode (particles larger than
14 $\sim 400 \text{ nm}$) that may have been associated with generation of sea spray by the higher wind velocities. Backtrajectories indicated
15 that airmasses had passed over the ocean to the northwest of BML, while 24-hour accumulated precipitation along the
16 trajectories (Table 1) indicated rainfall had occurred in the days prior to the airmass' arrival at BML. Each of these findings
17 was consistent with classification of the M1 cluster as a precipitation scrubbed, clean marine aerosol population.

18 The marine cluster with the next highest number concentration was M2, with 78% of its total occurrences between 17 and 21
19 February. The wind rose for this cluster indicated a primarily northwesterly wind, similar to cluster M1 but with much slower
20 velocities, and backtrajectories with oceanic transport pathways. While HYSPLIT does not always simulate sub-synoptic scale
21 transport with high fidelity, this pattern may be indicative of slower transport of air from a marine region just off the coast, as
22 opposed to the direct fast-transport path from more distant ocean regions seen in the M1 type. As observed for M1, the best-
23 fit average normalized particle size distribution (Fig. 4) was primarily bimodal, consistent with many reports of cloud-
24 processed background marine aerosol populations (Hoppel et al., 1986; Bates et al., 2000; Wex et al., 2016; Atwood et al.,
25 2017; Royalty et al., 2017; Phillips et al., 2018), though with a larger accumulation mode number fraction than the M1 type.
26 Minimal rainfall along the transport pathway was evident in the HYSPLIT accumulated precipitation estimates (Table 1),
27 indicating no recent precipitation scrubbing and suggesting that more cloud processing without rainout led to larger numbers
28 of particles in the accumulation mode compared with M1. In contrast to M1, for which a third fitted mode was found above
29 400 nm , the third mode in M2 occurred at very small particle sizes (diameters less than 30 nm).

30 The final marine cluster, M3, shared similarities with the other marine types, including a bimodal normalized size distribution
31 with a minimum near 110 nm and indications of oceanic source regions in local winds and backtrajectories. However, the
32 average total number concentration was nearly double that of the other marine types. The primary period during which this
33 cluster occurred was during 4-9 February, bracketing the time before, during, and after landfall of the AR that impacted the



1 BML region during CalWater-2015. This cluster is therefore interesting as it may be indicative of a unique population type
2 associated with AR meteorological conditions. Some caution is warranted, however, as instrument downtime lead to a gap in
3 the aerosol size distribution dataset during late 6 February through 7 February, during a high-wind and precipitation period
4 that marked the landfall of the AR, and thus some key data that could be used to guide the clustering during this event were
5 missing. When confined to the 5-8 February period noted by Leung (2016) when the AR made landfall at BML, average
6 number concentration was 749 cm^{-3} for time periods associated with the M3 cluster (excluding the data gap on 6-7 February),
7 and 1052 cm^{-3} for the entire 5-8 February AR period (including the intermittent periods classified as terrestrial or anthropogenic
8 aerosol).

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

16 3.2 Terrestrial population types

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

24 Several longer periods (28-31 January and 13-16 February) were observed during which populations T4 and T5 regularly
25 alternated in tandem with the diurnal land/sea-breeze shift. Similar diurnal-shift behavior, but between clusters T2 and T3,
26 occurred during 25-28 January and 1-5 March. During these diurnal shifts between various terrestrial clusters, the cluster with
27 the larger median diameter was typically associated with the sea-breeze and transport from oceanic regions, while the smaller
28 diameter cluster was associated with the land-breeze and transport from terrestrial regions. As aging of terrestrial aerosol
29 typically leads to an increase in the median diameter of the aerosol modes, these four clusters may therefore be indicative of
30 various degrees of aging of regional terrestrial aerosol during “sea-breeze resampling” at BML (Martin et al., 2017).
31 Resampling occurs when terrestrial and marine aerosol populations mix and flow across the coastal boundary at low levels,
32 leading to a region of mixed aerosol populations wherein the larger number concentrations associated with terrestrial types
33 dominate the observed number concentrations in the resulting littoral zone air mass. When this diurnal cycle collapsed and



1 BML was subjected to extended periods of sea-breezes and ventilation by air masses almost exclusively from ocean regions,
2 the cluster model selected marine cluster types, indicative of marine airmasses that had not experienced much mixing with
3 terrestrial air masses.

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

11 3.3 CCN and activated fraction spectra characteristics

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

23 In the CalWater-2015 dataset, marine population types all reached activated fractions of about 0.2 at supersaturations around
24 0.1% to 0.15%, while the terrestrial types did not reach equivalent fractions until supersaturations between approximately
25 0.18% and 0.6% were reached. Terrestrial population types with smaller median diameters tended to have less fractional
26 activation across the full range of measured supersaturations, leading to activated fractions at 1.0% supersaturation that varied
27 from approximately 0.3 to 0.85. However, due to the generally higher total particle number concentrations, and despite lower
28 activated fractions of the terrestrial populations, differences in CCN concentrations between marine and terrestrial types (Fig.
29 6b) were smaller than the differences in the activated fraction spectra. Only at supersaturations above approximately 0.5% did
30 the CCN concentration for the terrestrial types (except T1, associated with many fresh, small particles) consistently exceed
31 those of the marine types (Fig. 6b). Between approximately 0.1% and 0.4%, CCN concentrations were often similar between
32 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 to be an average of the coastal outflow marine cases 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. 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. 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 air mass prior to arrival at BML (Fig. 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 directly-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



1 similar to those noted by Kammermann et al., and potentially explain the larger discrepancies in CCN prediction at lower
2 supersaturations in the marine aerosol types.
3 Further investigation of the closure between predicted and observed CCN concentration was conducted using two prediction
4 models. Hygroscopicity derived predictions using cluster average κ and normalized size distribution values were generated
5 and compared against all observed CCN concentrations by the sfCCN system in Fig. 7(a). Similarly, the predicted CCN
6 concentration using cluster average activated fraction spectra were compared against observations in Fig. 7(b). Both models
7 predicted the activated fraction using the cluster type identified during the observation, which was then multiplied by the
8 observed total number concentration at the observation time. The hygroscopicity and size model showed overprediction
9 compared to both observations and the activated fraction model (Fig. 7c), with best-fit slopes of 1.08 and 1.09 respectively.
10 The activated fraction model predictions did improve on the hygroscopicity model, with a slope of 1.00 and R^2 values
11 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.
12 Nevertheless, the degree of closure between model predictions and observed CCN concentrations is similar to closures
13 previously reported in field studies (e.g. Bougiatioti et al., 2009; Kammermann et al., 2010).

14 3.5 Aerosol optical properties

15 While optical properties of the various aerosol population types were not directly measured, a simple optical reconstruction
16 was conducted to evaluate potential differences between the population types due to differences in size distribution and
17 hygroscopicity. Average particle size distributions and measured average κ values were used to grow particles to equilibrium
18 with relative humidity across a range of values between 0% and 99%, followed by estimation of mass scattering efficiency on
19 a dry mass basis using Mie theory (Bohren and Huffman, 1983). For the purposes of this simple optical comparison,
20 supersaturated κ values used for this analysis were assumed to be sufficient to provide an estimate of subsaturated hygroscopic
21 growth. 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
22 population types in order to estimate the relative effect of differences in size distribution, hygroscopicity, and relative humidity
23 on scattering properties at a wavelength of 550 nm. The index of refraction of each humidified particle was adjusted based on
24 volume mixing with water ($m = 1.33 + 0.0i$). Reconstructed mass scattering efficiencies for each of the population types are
25 shown in Fig. 8.

26 Computed dry mass scattering efficiencies at 550 nm ranged between 4.2 and $7.7 \text{ m}^2 \text{ g}^{-1}$, although actual values would be
27 expected to be lower due to expected particle dry densities higher than 1 g cm^{-3} . Further, these values represent only the
28 contribution to mass scattering efficiencies from the sub-micron aerosol, whereas super-micron aerosol, including particles
29 generated by sea spray in littoral environments, can represent a large fraction of the total light scattering. The M1 and M3
30 marine types, which included the largest fraction of larger accumulation mode particles, had the highest associated dry mass
31 scattering efficiency. The M2 marine type, which occurred at generally lower wind speeds than the other marine types and
32 thus had fewer larger particles associated with wind generated sea spray aerosol (O'Dowd and Leeuw, 2007), had a lower dry
33 mass scattering efficiency than several of the terrestrial types with the largest median particle diameters. However, at relative



1 humidity values above roughly 60%, as would typically be expected in littoral environments such as BML, the marine
2 population types all yielded expected mass scattering efficiencies above those of the terrestrial types. For the marine types at
3 a relative humidity of 95%, the mass scattering efficiencies on a unit-density dry mass basis were between 34 and 49 $\text{m}^2 \text{g}^{-1}$,
4 roughly twice the range of the terrestrial types, 16 to 25 $\text{m}^2 \text{g}^{-1}$. As an additional point of comparison, similar reconstructions
5 with MODIS fine mode ocean aerosol populations (Remer et al., 2006) yielded dry mass scattering efficiencies between 3.4
6 and 5.4 $\text{m}^2 \text{g}^{-1}$, and 33 and 54 $\text{m}^2 \text{g}^{-1}$ at 95% RH, indicating the marine types were within the same range as assumptions used
7 for MODIS marine aerosol populations.

8 **4 Summary**

9 The unsupervised cluster model analysis successfully identified distinct aerosol population types in the littoral zone aerosol at
10 BML during CalWater-2015. The time periods selected by each cluster tended to be both temporally and physically coherent.
11 Clusters also tended to be grouped into periods with physically meaningful microphysical properties that could be associated
12 with meteorological processes and expected sources and transport pathways. For example, the clustering methodology
13 identified regular diurnal swings in aerosol properties associated with land/sea-breeze changes and assigned two distinct,
14 terrestrially-influenced aerosol types during these periods. Overall, the clustering results for the CalWater-2015 dataset
15 produced a reliable set of aerosol population types, and appropriately screened intermittent periods of impacts from various
16 other sources as an outcome of the classification. Both marine and terrestrially influenced aerosol population types were
17 identified by the unsupervised cluster model. Several marine events that persisted for days were identified as distinct in
18 character from each other—differing in the degree of cloud processing and precipitation removal prior to arrival at the
19 measurement site, and in the extent to which high winds contributed larger sea spray particles. About 10% of the observations
20 were associated with a terrestrial population with a large fraction of small particles, indicating it was affected by relatively
21 fresh emissions and/or new particle formation.

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

28 In this littoral region sea-breeze resampling and complex mixing between marine, terrestrial, and free-tropospheric air masses
29 lead to complex aerosol populations. Determination of which aerosol types can become incorporated into AR events and
30 thereby affect cloud and precipitation over California depends strongly on the individual nature of the flow regimes and
31 potential aerosol source regions that they tap into. Nevertheless, the observational data set had cases for which marine and
32 terrestrial CCN concentrations were comparable at supersaturations below approximately 0.4%, and thus changes in aerosol



1 types would result in little change to initial drop populations in the resulting stratus cloud, except at very low supersaturations.
2 Thus, mixed-phase microphysical processes occurring in those clouds might also be expected to be similar whether marine or
3 terrestrial aerosols served as the nuclei for the supercooled droplets. In contrast, for clouds forming at higher maximum
4 supersaturations, terrestrial aerosol populations are expected to yield higher drop concentrations than marine types, with the
5 exception of a terrestrial population characterized by primarily small, recently emitted or newly formed particles. Thus the
6 droplet size distributions formed on terrestrial vs. marine types, for liquid clouds formed in stronger updrafts or at higher
7 cooling rates that reach such higher supersaturations, should be distinctly different.

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

14 **Data availability**

15 Measurement data for the scanning flow CCN used in this study are available at doi:10.25675/10217/180097

16 Additional experimental data used in the analysis and presented in the figures are provided in an online data repository at
17 doi:10.5281/zenodo.XXXXXXX (available upon publication).

18 **Author Contributions**

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

26 **Acknowledgements**

27 *This material is based upon research by the Office of Naval Research under Award Number N00014-16-1-2040. This work*
28 *was supported by NSF award number 1450690 (MDP), NSF award number 1450760 (SAA, SMK, PJD), and NSF award*
29 *number 1451347 (GCC, ACM, KAM). Assistance in operation of the research site was provided by Nicholas E. Rothfuss, Hans*



1 *Taylor, Ezra Levin, Christina McCluskey, Yvonne Boose, Gregg Schill, Camille Sultana, and Kim Prather. The assistance by*
2 *BML staff and Bodega Marine Reserve for lending laboratory space, assisting with measurement site improvements, and*
3 *permission to measure on Reserve land is gratefully acknowledged. The loan of a CCN instrument by Jeff Reid and the U.S.*
4 *Naval Research Laboratory, Monterey is also gratefully acknowledged.*

5



1 References

- 2 Andreae, M. O. and Rosenfeld, D.: Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active
3 aerosols, *Earth-Sci. Rev.*, 89(1–2), 13–41, doi:10.1016/j.earscirev.2008.03.001, 2008.
- 4 Atwood, S. A., Reid, J. S., Kreidenweis, S. M., Blake, D. R., Jonsson, H. H., Lagrosas, N. D., Xian, P., Reid, E. A., Sessions,
5 W. R. and Simpas, J. B.: Size-resolved aerosol and cloud condensation nuclei (CCN) properties in the remote marine South
6 China Sea – Part 1: Observations and source classification, *Atmos Chem Phys*, 17(2), 1105–1123, doi:10.5194/acp-17-1105-
7 2017, 2017.
- 8 Baarsch, J. and Celebi, M. E.: Investigation of internal validity measures for K-means clustering, in Proceedings of the
9 International MultiConference of Engineers and Computer Scientists, vol. 1, pp. 14–16. [online] Available from:
10 http://www.iaeng.org/publication/IMECS2012/IMECS2012_pp471-476.pdf (Accessed 20 November 2015), 2012.
- 11 Bates, T. S., Quinn, P. K., Covert, D. S., Coffman, D. J., Johnson, J. E. and Wiedensohler, A.: Aerosol physical properties and
12 processes in the lower marine boundary layer: a comparison of shipboard sub-micron data from ACE-1 and ACE-2, *Tellus B*,
13 52(2), 258–272, doi:10.1034/j.1600-0889.2000.00021.x, 2000.
- 14 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,
15 S.-M., Hayden, K. L., Nuaaman, I., McLaren, R., Massoli, P., Canagaratna, M. R., Onasch, T. B., Sueper, D., Worsnop, D. R.
16 and Keene, W. C.: Measurements of ocean derived aerosol off the coast of California, *J. Geophys. Res. Atmospheres*,
17 117(D21), D00V15, doi:10.1029/2012JD017588, 2012.
- 18 Beddows, D. C. S., Dall’Osto, M. and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle
19 Size Data, *Environ. Sci. Technol.*, 43(13), 4694–4700, doi:10.1021/es803121t, 2009.
- 20 Bohren, C. F. and Huffman, D. R.: Absorption and scattering by a sphere, *Absorpt. Scatt. Light Small Part.*, 82–129, 1983.
- 21 Bougiatioti, A., Fountoukis, C., Kalivitis, N., Pandis, S. N., Nenes, A. and Mihalopoulos, N.: Cloud condensation nuclei
22 measurements in the marine boundary layer of the Eastern Mediterranean: CCN closure and droplet growth kinetics, *Atmos*
23 *Chem Phys*, 9(18), 7053–7066, doi:10.5194/acp-9-7053-2009, 2009.
- 24 Charron, A., Birmili, W. and Harrison, R. M.: Fingerprinting particle origins according to their size distribution at a UK rural
25 site, *J. Geophys. Res. Atmospheres*, 113(D7), D07202, doi:10.1029/2007JD008562, 2008.
- 26 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.
27 and Cayan, D. R.: Atmospheric Rivers, Floods and the Water Resources of California, *Water*, 3(2), 445–478,
28 doi:10.3390/w3020445, 2011.
- 29 Draxler, R. R. and Hess, G. D.: Description of the HYSPLIT4 modeling system, [online] Available from:
30 <http://warn.arl.noaa.gov/documents/reports/arl-224.pdf> (Accessed 14 April 2015), 1997.
- 31 Draxler, R. R. and Hess, G. D.: An overview of the HYSPLIT_4 modelling system for trajectories, *Aust. Meteorol. Mag.*,
32 47(4), 295–308, 1998.



- 1 Draxler, R. R., Stunder, B., Rolph, G. and Taylor, A.: HYSPLIT4 user's guide, NOAA Tech. Memo. ERL ARL, 230, 35,
2 1999.
- 3 Forestieri, S. D., Cornwell, G. C., Helgestad, T. M., Moore, K. A., Lee, C., Novak, G. A., Sultana, C. M., Wang, X., Bertram,
4 T. H., Prather, K. A. and Cappa, C. D.: Linking variations in sea spray aerosol particle hygroscopicity to composition during
5 two microcosm experiments, *Atmos Chem Phys*, 16(14), 9003–9018, doi:10.5194/acp-16-9003-2016, 2016.
- 6 Frossard, A. A., Russell, L. M., Burrows, S. M., Elliott, S. M., Bates, T. S. and Quinn, P. K.: Sources and composition of
7 submicron organic mass in marine aerosol particles, *J. Geophys. Res. Atmospheres*, 119(22), 2014JD021913,
8 doi:10.1002/2014JD021913, 2014.
- 9 Good, N., Topping, D. O., Allan, J. D., Flynn, M., Fuentes, E., Irwin, M., Williams, P. I., Coe, H. and McFiggans, G.:
10 Consistency between parameterisations of aerosol hygroscopicity and CCN activity during the RHaMBLe discovery cruise,
11 *Atmos Chem Phys*, 10(7), 3189–3203, doi:10.5194/acp-10-3189-2010, 2010.
- 12 Hoppel, W. A., Frick, G. M. and Larson, R. E.: Effect of nonprecipitating clouds on the aerosol size distribution in the marine
13 boundary layer, *Geophys. Res. Lett.*, 13(2), 125–128, doi:10.1029/GL013i002p00125, 1986.
- 14 Hussein, T., Dal Maso, M., Petäjä, T., Koponen, I. K., Paatero, P., Aalto, P. P., Hämeri, K. and Kulmala, M.: Evaluation of an
15 automatic algorithm for fitting the particle number size distributions, *Boreal Environ. Res.*, 10(5), 337–355, 2005.
- 16 Irwin, M., Good, N., Crosier, J., Choularton, T. W. and McFiggans, G.: Reconciliation of measurements of hygroscopic growth
17 and critical supersaturation of aerosol particles in central Germany, *Atmospheric Chem. Phys.*, 10(23), 11737–11752,
18 doi:<https://doi.org/10.5194/acp-10-11737-2010>, 2010.
- 19 Kammermann, L., Gysel, M., Weingartner, E., Herich, H., Cziczo, D. J., Holst, T., Svenningsson, B., Arneth, A. and
20 Baltensperger, U.: Subarctic atmospheric aerosol composition: 3. Measured and modeled properties of cloud condensation
21 nuclei, *J. Geophys. Res. Atmospheres*, 115(D4), doi:10.1029/2009JD012447, 2010.
- 22 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.
23 S., Zhou, X., Smoydzin, L. and Sander, R.: Chemical and physical characteristics of nascent aerosols produced by bursting
24 bubbles at a model air-sea interface, *J. Geophys. Res. Atmospheres*, 112(D21), D21202, doi:10.1029/2007JD008464, 2007.
- 25 Leung, L. R.: ARM Cloud-Aerosol-Precipitation Experiment (ACAPEX) Field Campaign Report, DOE ARM Climate
26 Research Facility, Pacific Northwest National Laboratory, Richland, WA. [online] Available from:
27 <https://www.osti.gov/biblio/1251152> (Accessed 16 October 2018), 2016.
- 28 Martin, A. C., Cornwell, G. C., Atwood, S. A., Moore, K. A., Rothfuss, N. E., Taylor, H., DeMott, P. J., Kreidenweis, S. M.,
29 Petters, M. D. and Prather, K. A.: Transport of pollution to a remote coastal site during gap flow from California's interior:
30 impacts on aerosol composition, clouds, and radiative balance, *Atmos Chem Phys*, 17(2), 1491–1509, doi:10.5194/acp-17-
31 1491-2017, 2017.
- 32 McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M. C., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A.,
33 Lohmann, U., Mentel, T. F., Murphy, D. M., O'Dowd, C. D., Snider, J. R. and Weingartner, E.: The effect of physical and



- 1 chemical aerosol properties on warm cloud droplet activation, *Atmos Chem Phys*, 6(9), 2593–2649, doi:10.5194/acp-6-2593-
2 2006, 2006.
- 3 Moore, R. H. and Nenes, A.: Scanning Flow CCN Analysis—A Method for Fast Measurements of CCN Spectra, *Aerosol Sci.*
4 *Technol.*, 43(12), 1192–1207, doi:10.1080/02786820903289780, 2009.
- 5 Nair, V. S., Moorthy, K. K. and Babu, S. S.: Influence of continental outflow and ocean biogeochemistry on the distribution
6 of fine and ultrafine particles in the marine atmospheric boundary layer over Arabian Sea and Bay of Bengal, *J. Geophys. Res.*
7 *Atmospheres*, 118(13), 7321–7331, doi:10.1002/jgrd.50541, 2013.
- 8 Neiman, P. J., Hughes, M., Moore, B. J., Ralph, F. M. and Sukovich, E. M.: Sierra Barrier Jets, Atmospheric Rivers, and
9 Precipitation Characteristics in Northern California: A Composite Perspective Based on a Network of Wind Profilers, *Mon.*
10 *Weather Rev.*, 141(12), 4211–4233, doi:10.1175/MWR-D-13-00112.1, 2013.
- 11 O’Dowd, C. D. and Leeuw, G. de: Marine aerosol production: a review of the current knowledge, *Philos. Trans. R. Soc. Math.*
12 *Phys. Eng. Sci.*, 365(1856), 1753–1774, doi:10.1098/rsta.2007.2043, 2007.
- 13 Paramonov, M., Kerminen, V.-M., Gysel, M., Aalto, P. P., Andreae, M. O., Asmi, E., Baltensperger, U., Bougiatioti, A., Brus,
14 D., Frank, G. P., Good, N., Gunthe, S. S., Hao, L., Irwin, M., Jaatinen, A., Jurányi, Z., King, S. M., Kortelainen, A.,
15 Kristensson, A., Lihavainen, H., Kulmala, M., Lohmann, U., Martin, S. T., McFiggans, G., Mihalopoulos, N., Nenes, A.,
16 O’Dowd, C. D., Ovadnevaite, J., Petäjä, T., Pöschl, U., Roberts, G. C., Rose, D., Svenningsson, B., Swietlicki, E., Weingartner,
17 E., Whitehead, J., Wiedensohler, A., Wittbom, C. and Sierau, B.: A synthesis of cloud condensation nuclei counter (CCNC)
18 measurements within the EUCAARI network, *Atmos Chem Phys*, 15(21), 12211–12229, doi:10.5194/acp-15-12211-2015,
19 2015.
- 20 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R.,
21 Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É.: Scikit-learn: Machine
22 Learning in Python, *J Mach Learn Res*, 12, 2825–2830, 2011.
- 23 Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation
24 nucleus activity, *Atmos Chem Phys*, 7(8), 1961–1971, doi:10.5194/acp-7-1961-2007, 2007.
- 25 Petters, S. S. and Petters, M. D.: Surfactant effect on cloud condensation nuclei for two-component internally mixed aerosols,
26 *J. Geophys. Res. Atmospheres*, 121(4), 1878–1895, doi:10.1002/2015JD024090, 2016.
- 27 Phillips, B. N., Royalty, T. M., Dawson, K. W., Reed, R., Petters, M. D. and Meskhidze, N.: Hygroscopicity- and Size-
28 Resolved Measurements of Submicron Aerosol on the East Coast of the United States, *J. Geophys. Res. Atmospheres*, 123(3),
29 1826–1839, doi:10.1002/2017JD027702, 2018.
- 30 Prather, K. A., Bertram, T. H., Grassian, V. H., Deane, G. B., Stokes, M. D., DeMott, P. J., Aluwihare, L. I., Palenik, B. P.,
31 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.
32 P., Baltrusaitis, J., Collins, D. B., Corrigan, C. E., Cuadra-Rodriguez, L. A., Ebben, C. J., Forestieri, S. D., Guasco, T. L.,
33 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.,



- 1 Sullivan, R. C. and Zhao, D.: Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol,
2 Proc. Natl. Acad. Sci., 110(19), 7550–7555, doi:10.1073/pnas.1300262110, 2013.
- 3 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.:
4 Contribution of sea surface carbon pool to organic matter enrichment in sea spray aerosol, Nat. Geosci., 7(3), 228–232,
5 doi:10.1038/ngeo2092, 2014.
- 6 Ralph, F. M., Neiman, P. J. and Wick, G. A.: Satellite and CALJET Aircraft Observations of Atmospheric Rivers over the
7 Eastern North Pacific Ocean during the Winter of 1997/98, Mon. Weather Rev., 132(7), 1721–1745, doi:10.1175/1520-
8 0493(2004)132<1721:SACAOO>2.0.CO;2, 2004.
- 9 Ralph, F. M., Prather, K. A., Cayan, D., Spackman, J. R., DeMott, P., Dettinger, M., Fairall, C., Leung, R., Rosenfeld, D.,
10 Rutledge, S., Waliser, D., White, A. B., Cordeira, J., Martin, A., Helly, J. and Intrieri, J.: CalWater Field Studies Designed to
11 Quantify the Roles of Atmospheric Rivers and Aerosols in Modulating U.S. West Coast Precipitation in a Changing Climate,
12 Bull. Am. Meteorol. Soc., 97(7), 1209–1228, doi:10.1175/BAMS-D-14-00043.1, 2015.
- 13 Remer, L. A., Tanre, D., Kaufman, Y. J., Levy, R. and Mattoo, S.: Algorithm for remote sensing of tropospheric aerosol from
14 MODIS: Collection 005, Natl. Aeronaut. Space Adm., 1490 [online] Available from:
15 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.385.6530&rep=rep1&type=pdf> (Accessed 8 January 2017), 2006.
- 16 Rosenfeld, D., Lohmann, U., Raga, G. B., O’Dowd, C. D., Kulmala, M., Fuzzi, S., Reissell, A. and Andreae, M. O.: Flood or
17 Drought: How Do Aerosols Affect Precipitation?, Science, 321(5894), 1309–1313, doi:10.1126/science.1160606, 2008.
- 18 Royalty, T. M., Phillips, B. N., Dawson, K. W., Reed, R., Meskhidze, N. and Petters, M. D.: Aerosol Properties Observed in
19 the Subtropical North Pacific Boundary Layer, J. Geophys. Res. Atmospheres, 122(18), 9990–10,012,
20 doi:10.1002/2017JD026897, 2017.
- 21 Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., Clifford, S. and Morawska, L.: Assessment and
22 application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment,
23 Atmos Chem Phys, 14(21), 11883–11892, doi:10.5194/acp-14-11883-2014, 2014.
- 24 Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D. and Ngan, F.: NOAA’s HYSPLIT Atmospheric
25 Transport and Dispersion Modeling System, Bull. Am. Meteorol. Soc., 96(12), 2059–2077, doi:10.1175/BAMS-D-14-
26 00110.1, 2015.
- 27 Suda, S. R., Petters, M. D., Matsunaga, A., Sullivan, R. C., Ziemann, P. J. and Kreidenweis, S. M.: Hygroscopicity frequency
28 distributions of secondary organic aerosols, J. Geophys. Res. Atmospheres, 117(D4), D04207, doi:10.1029/2011JD016823,
29 2012.
- 30 Suda, S. R., Petters, M. D., Yeh, G. K., Strollo, C., Matsunaga, A., Faulhaber, A., Ziemann, P. J., Prenni, A. J., Carrico, C. M.,
31 Sullivan, R. C. and Kreidenweis, S. M.: Influence of Functional Groups on Organic Aerosol Cloud Condensation Nucleus
32 Activity, Environ. Sci. Technol., 48(17), 10182–10190, doi:10.1021/es502147y, 2014.



- 1 Tunved, P., Ström, J. and Hansson, H.-C.: An investigation of processes controlling the evolution of the boundary layer aerosol
2 size distribution properties at the Swedish background station Aspöreten, *Atmos Chem Phys*, 4(11/12), 2581–2592,
3 doi:10.5194/acp-4-2581-2004, 2004.
- 4 Wegner, T., Hussein, T., Hämeri, K., Vesala, T., Kulmala, M. and Weber, S.: Properties of aerosol signature size distributions
5 in the urban environment as derived by cluster analysis, *Atmos. Environ.*, 61, 350–360, doi:10.1016/j.atmosenv.2012.07.048,
6 2012.
- 7 Wex, H., Dieckmann, K., Roberts, G. C., Conrath, T., Izaguirre, M. A., Hartmann, S., Herenz, P., Schäfer, M., Ditas, F.,
8 Schmeissner, T., Henning, S., Wehner, B., Siebert, H. and Stratmann, F.: Aerosol arriving on the Caribbean island of Barbados:
9 physical properties and origin, *Atmos Chem Phys*, 16(22), 14107–14130, doi:10.5194/acp-16-14107-2016, 2016.
- 10 White, A. B., Anderson, M. L., Dettinger, M. D., Ralph, F. M., Hinojosa, A., Cayan, D. R., Hartman, R. K., Reynolds, D. W.,
11 Johnson, L. E., Schneider, T. L., Cifelli, R., Toth, Z., Gutman, S. I., King, C. W., Gehrke, F., Johnston, P. E., Walls, C., Mann,
12 D., Gattas, D. J. and Coleman, T.: A Twenty-First-Century California Observing Network for Monitoring Extreme Weather
13 Events, *J. Atmospheric Ocean. Technol.*, 30(8), 1585–1603, doi:10.1175/JTECH-D-12-00217.1, 2013.
- 14 Whitehead, J. D., Irwin, M., Allan, J. D., Good, N. and McFiggans, G.: A meta-analysis of particle water uptake reconciliation
15 studies, *Atmospheric Chem. Phys.*, 14(21), 11833–11841, doi:<https://doi.org/10.5194/acp-14-11833-2014>, 2014.
- 16 Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, Academic Press., 2011.
- 17 Zhang, X., Massoli, P., Quinn, P. K., Bates, T. S. and Cappa, C. D.: Hygroscopic growth of submicron and supermicron
18 aerosols in the marine boundary layer, *J. Geophys. Res. Atmospheres*, 119(13), 2013JD021213, doi:10.1002/2013JD021213,
19 2014.
- 20 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
21 marine aerosols over the northern South China Sea: Influence of continental outflow, *Chemosphere*, 153, 332–339,
22 doi:10.1016/j.chemosphere.2016.03.069, 2016.

23
24



Tables

Table 1 Aerosol and meteorological parameters for each of the cluster time periods. 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 N) | Total Number Concentration (# cm ⁻³) (std dev) | k | | Wind Velocity (m/s) | | Trajectory 24-hr Accum Precip (mm) (std dev) | |
|-------------|---|--|-------------|-----------|------------------------|-------|--|-----------|
| | | | Mean (-) | (std dev) | u | v | Mean (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.

| BML Cluster | Avg Size Dist - Mode 1 | | | Avg Size Dist - Mode 2 | | | Avg Size Dist - Mode 3 | | | Activation Spectrum Fit Mode 1 | | | Activation Spectrum 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 |

25

26

27

28

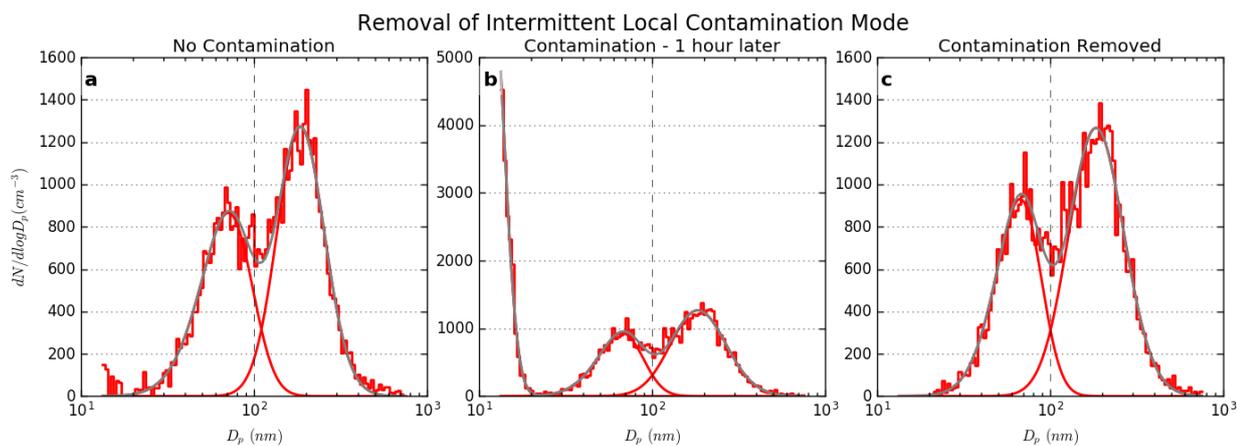
29

30

31



32 **Figures**



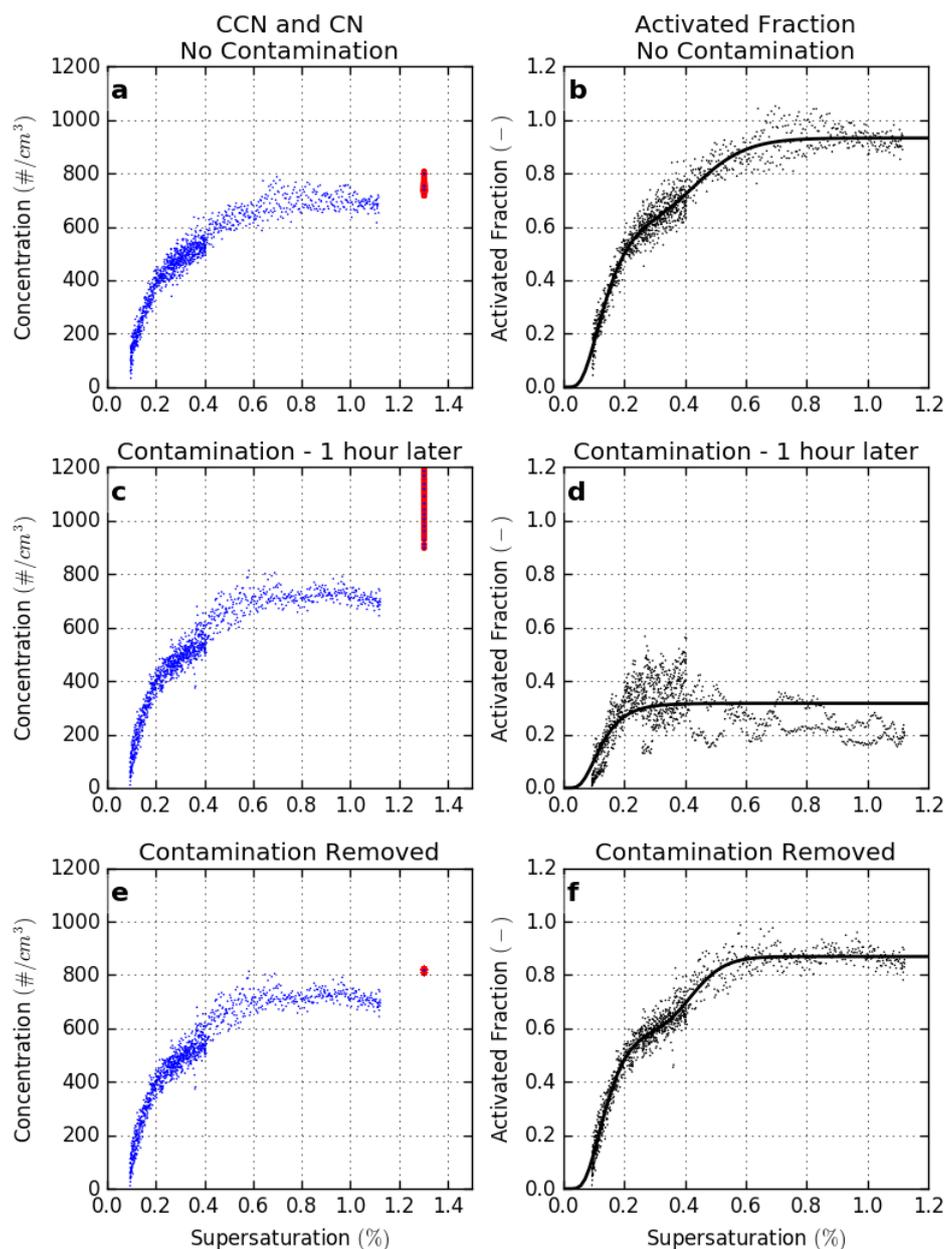
33

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.

34



Example Scanning Flow Spectra



35

Figure 2 Example spectra measured by the sfCCN system for the same time period as in Figure 1. (a) One second CCN (blue) and 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).

36

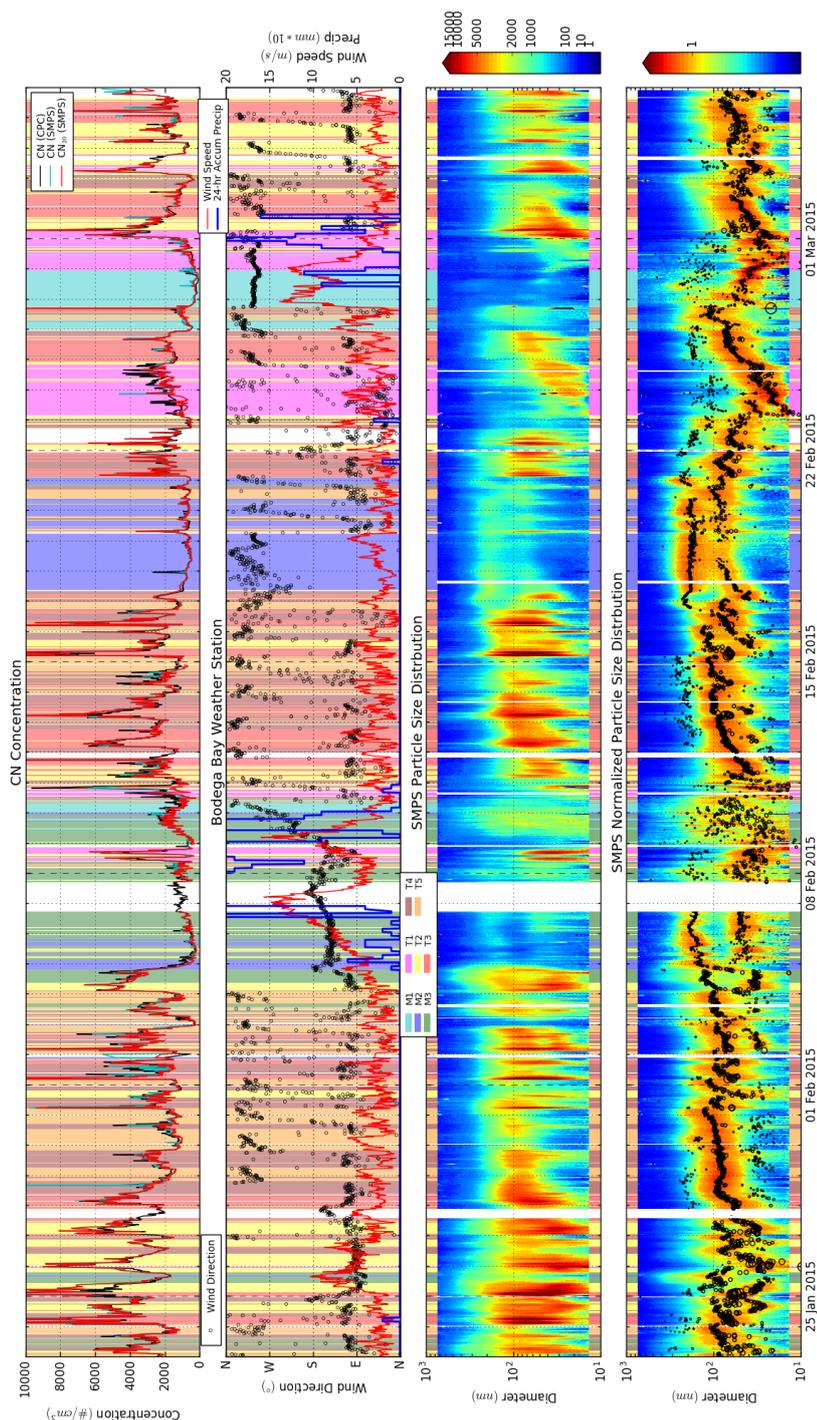
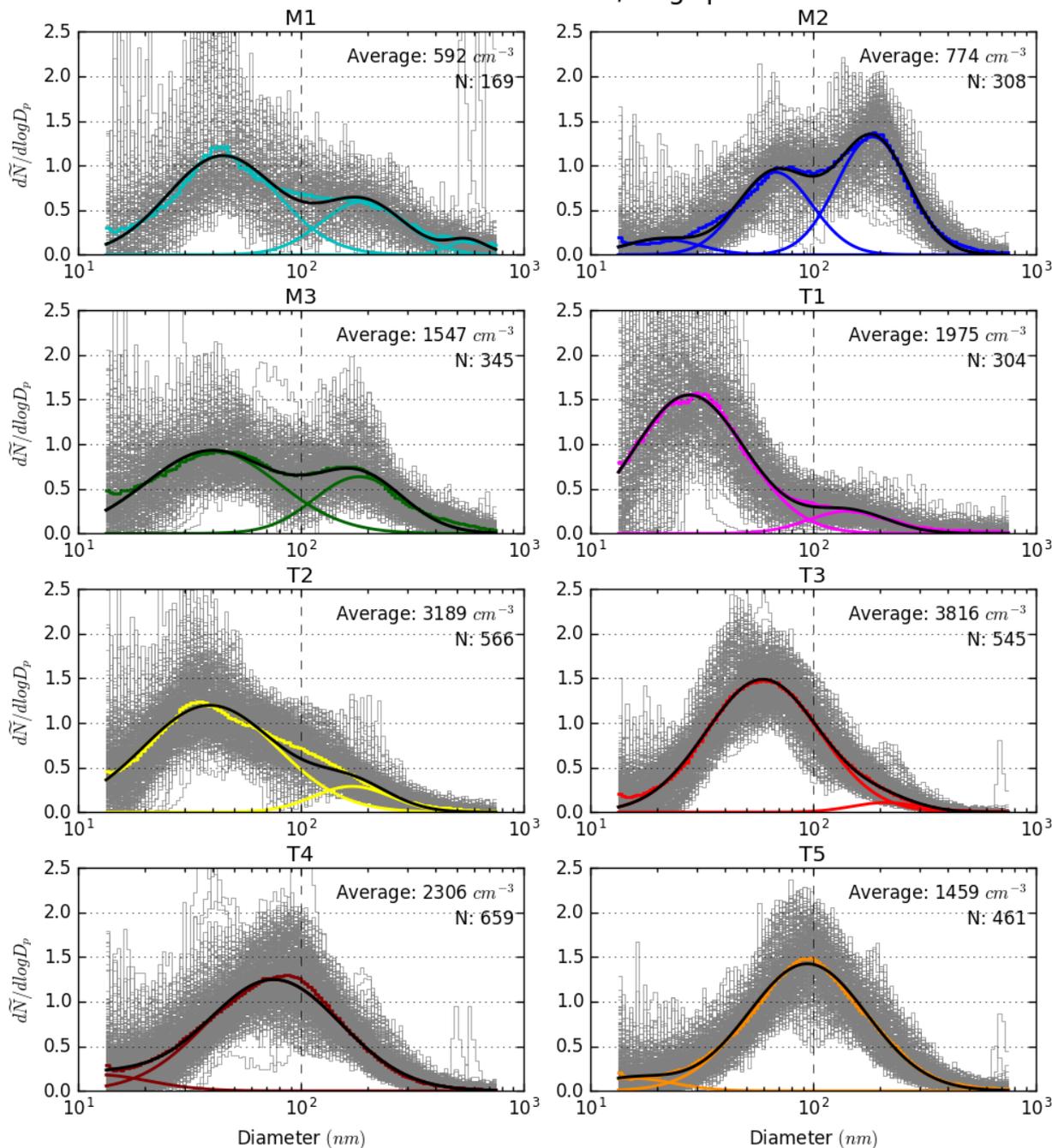


Figure 3 Timelines of study data; all times shown are UTC. (a) Total particle number concentration at BML measured by the CPC (black), reconstructed from SMPS size distributions (blue), and after correction for removal of local sources (red). (b) BML weather station measurements of wind direction (black circles) and wind speed (red lines), with HYSPLIT 24-hour accumulated precipitation along the air mass 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.



SMPS Normalized $dN/d\log D_p$



1

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.

2



Wind Vector and 100m HYSPLIT backtrajectories

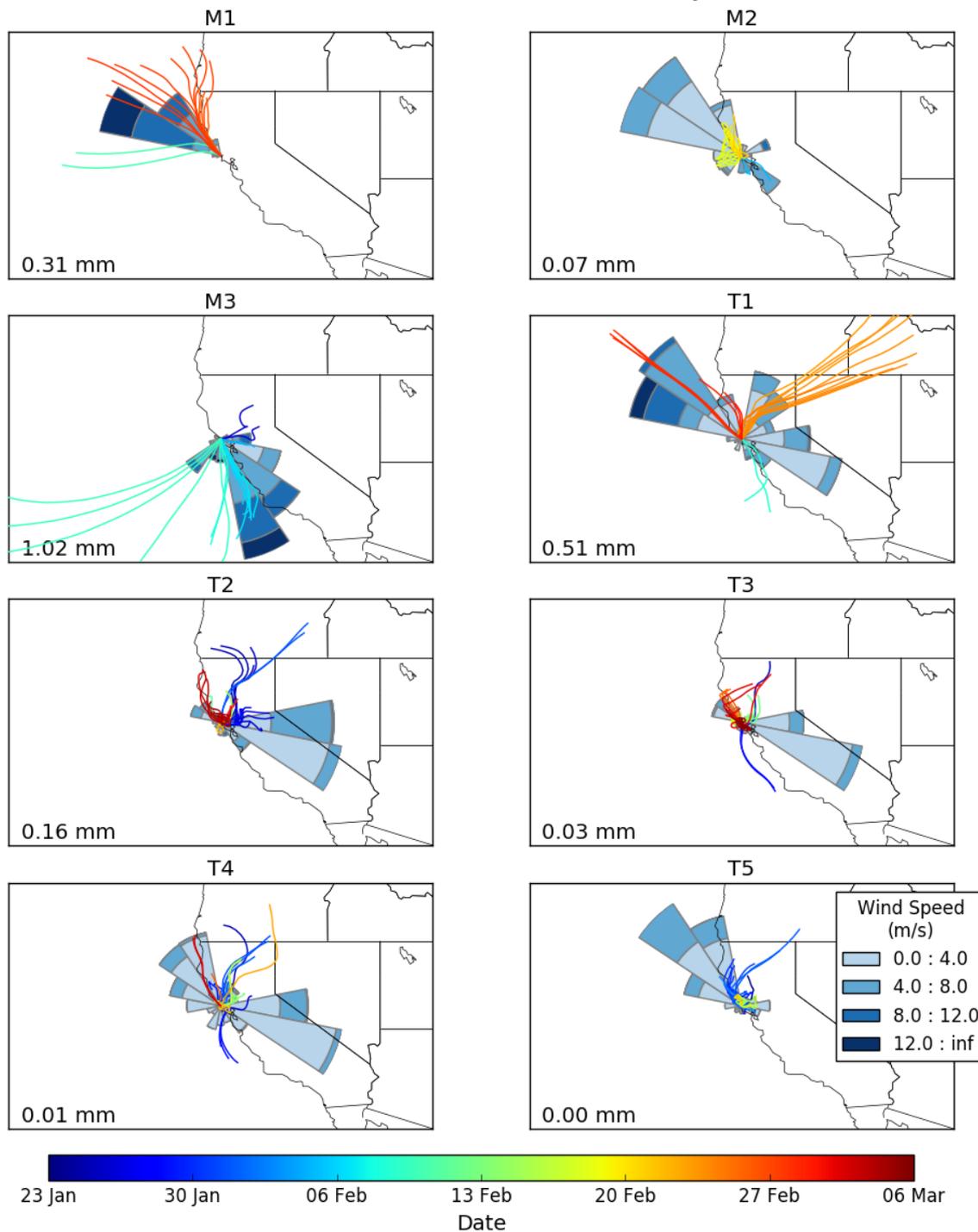
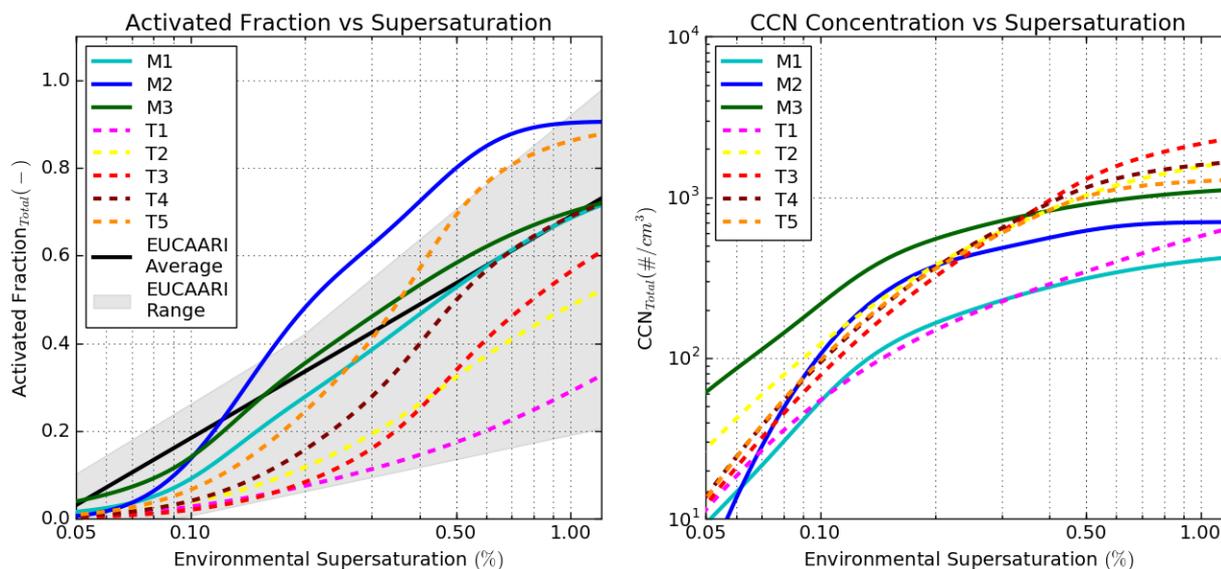




Figure 5 Meteorological overview for each cluster from the 8 cluster K-means classification. HYSPLIT 24-hr backtrajectories are shown for each time stamp associated with the cluster and colored by the date of arrival at the receptor. Wind rose plots are shown for all BML local 10m wind direction and 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

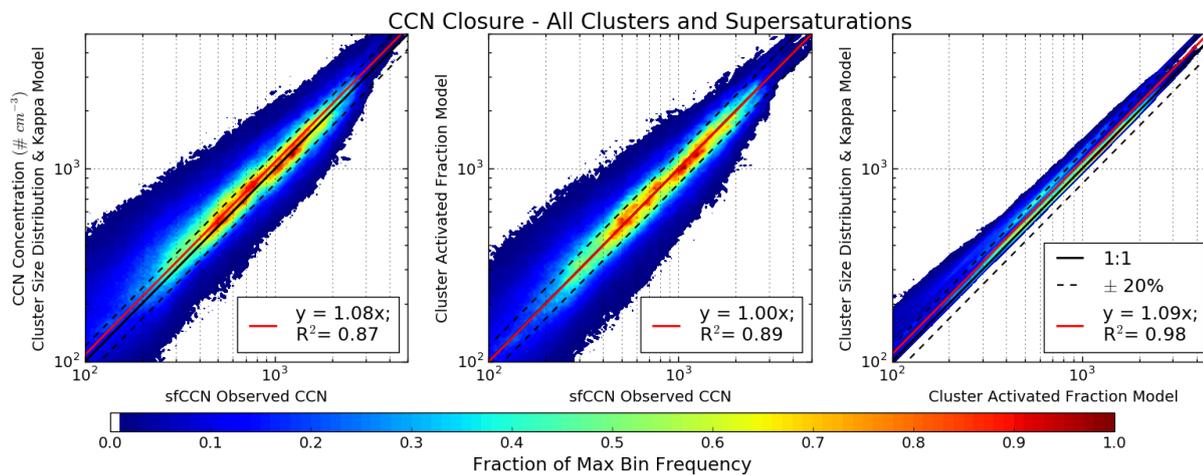


2

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

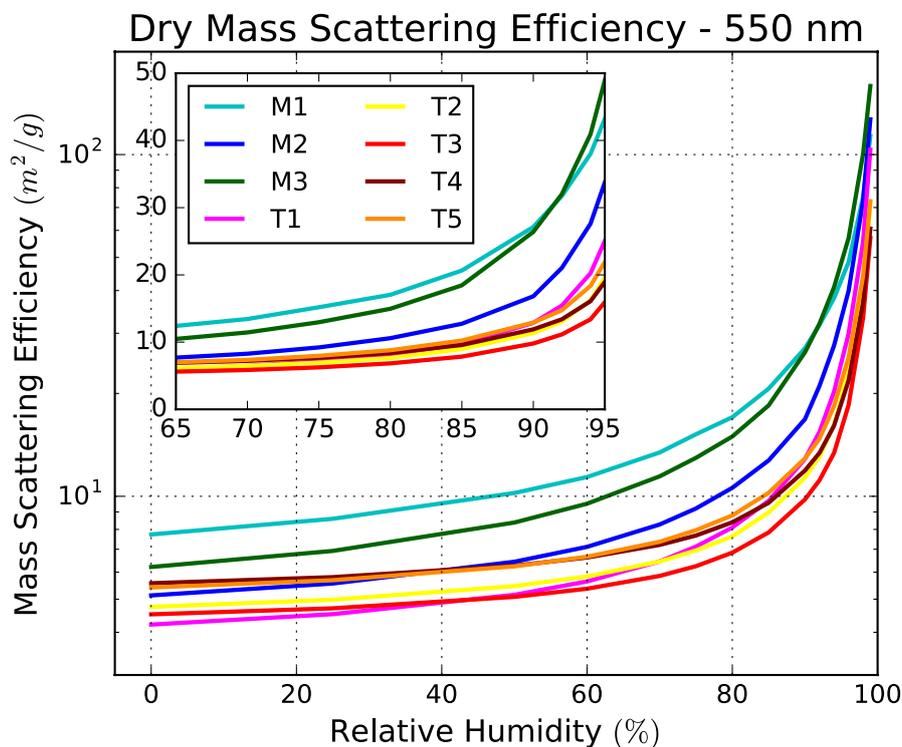
4



1

2 **Figure 7** CCN closure comparing predicted and measured CCN number concentrations for (a) the predicted concentration based
 3 on the size distribution and kappa reconstruction against observed concentrations, (b) the predicted concentration based on the
 4 best-fit activated fraction reconstruction against observed concentrations, and (c) the comparison between the two reconstruction
 5 methods. All cluster types and supersaturations across the full range of observed values are shown. Linear best-fit slope is shown in
 6 red with the associated R-squared values. One to one lines and lines at $\pm 20\%$ are shown.

7



1

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

5