1   Flow Scanning CCN (sfCCN) System Calibration

2   Calibration of the flow scanning CCN column supersaturation for a given flow rate at each temperature gradient

3   setting was conducted following the methodology described in Suda et al., (2012). Briefly, a monodisperse

4   ammonium sulfate aerosol was generated by atomizing an ammonium sulfate solution and size selecting using a TSI

5   3080 Differential Mobility Analyzer (DMA) set to a fixed voltage. Concentrations of CN and CCN were measured

6   as flow rate changed. The flow rate at which 50% of particles had activated (Q_{50}) was calculated by fitting a sigmoid

7   curve to the activation plots for a range of DMA column voltages. Given a κ hygroscopicity parameter of 0.61

8   (Petters and Kreidenweis, 2007) for the ammonium sulfate particles and the parameters of the DMA, calibration

9   curves for column supersaturation as a function of flow rate and temperature gradient were calculated. Calibrations

10   were conducted for column temperature gradients of 4 K and 8 K (referred to as dT4 and dT8, respectively) that

11   were used in these measurements.

12   As residence time in the sfCCN varied with total flow rate, timestamps were empirically adjusted to align the data

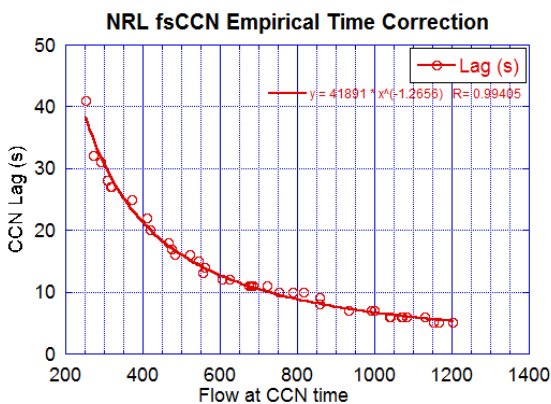
13   with the parallel CPC. The calibrated CPC timestamp adjustment was made as a function of flow by measuring the

14   time delay between narrow peaks of particles introduced at specific flow rates. A filter was placed upstream of the

15   split between instruments and then briefly removed, thereby allowing ambient particles to be sampled by both

16   instruments. The time lag between measurements was then determined for a number of flow rates for both

17   measurement and calibration samples. A calibration curve was fit to the aligned time stamps (Fig. S1) and used to

18   adjust the timestamp of the CPC data points for both ambient and calibration scans.

Supplemental Figure 1   Equivalent CPC time correction function for the sfCCN system at various flow rates (lpm).

19

20

21 **Aerosol Size Distribution Correction**

22 A raw timeline of particle number concentration and size distribution is shown in Supplemental Figure 2(a&b). The

23 corrected number concentrations and size distributions after removal of the contamination modes are shown in

24 (a&c). Both the generally short duration and small particle size of the identified contamination modes are evident in

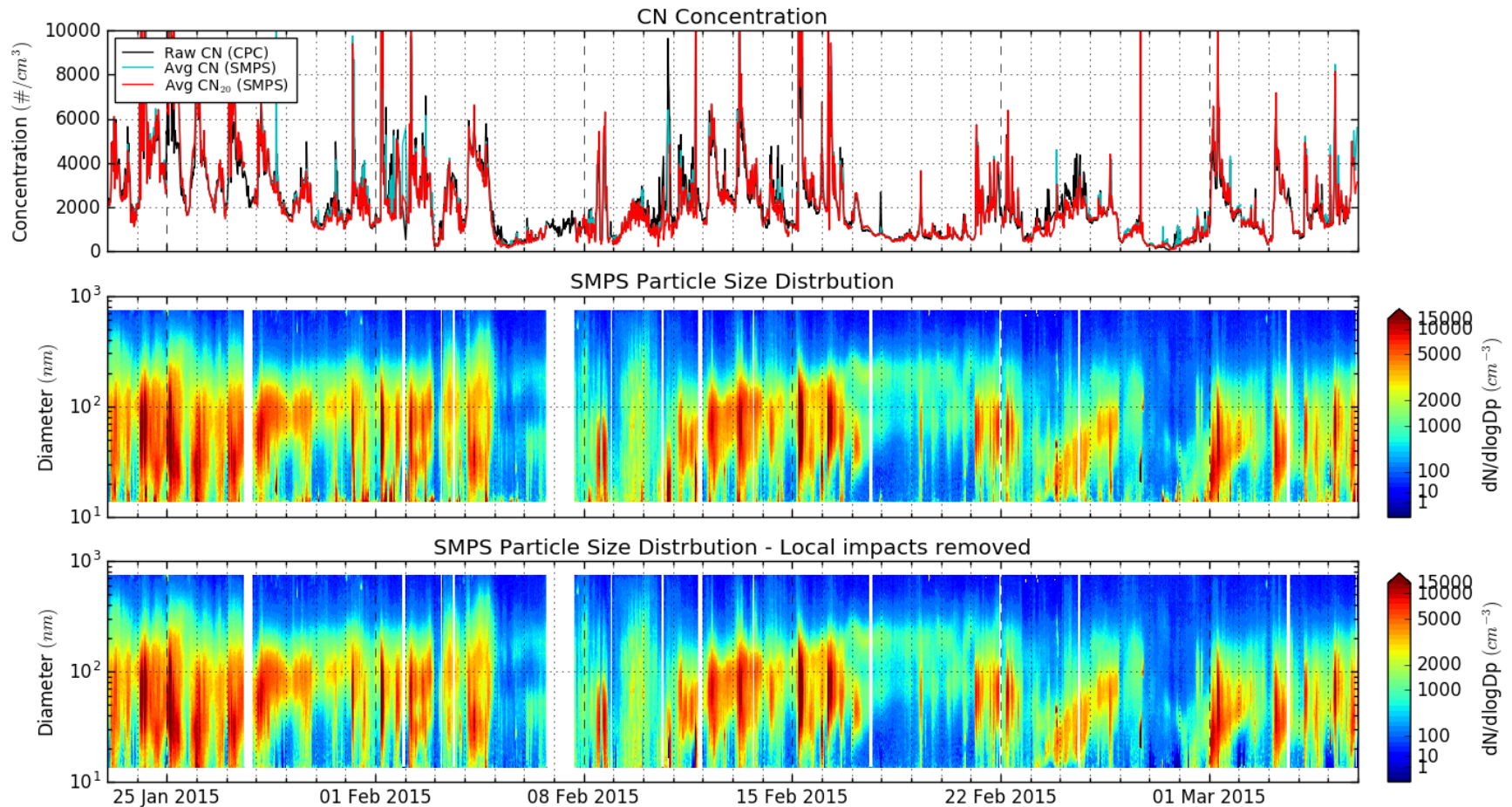
25 these figures. Narrow peaks in the light blue line in (a) show that the correction to number concentration occurred

26 over generally short time periods, while the predominant differences in the corrected size distribution of (c) were for

27 the smallest particle sizes. Relatively little change in the overall nature of the shape of the overall distributions

28 resulted from this correction, but it allowed for markedly improved performance of the cluster model.

29

30

Supplemental Figure 2 (a) Study timeline with CN concentrations from the bulk CPC (black), reconstructed CN from the SMPS (light blue), and reconstructed CN corrected for intermittent local contamination modes (red). SMPS particle size distributions are shown (b) before and (c) after correction to remove contamination.

33

34   **Activated Fraction Spectra Fitting**

35   Parameterizations of activated fraction spectra have been presented using various methodologies. The
36   parameterizations of Paramonov et al., (2015) for various locations in the EUCAARI network, shown in **Error!**
37   **Reference source not found.**, used a two parameter fit of the form $AF = a \times \ln(S_{eff}) + b$, where S_{eff} is the
38   effective environmental supersaturation. However, this parameterization did not adequately reflect the differences
39   between spectra in the BML dataset. In order to provide a more accurate parameterization of observed activation
40   spectra, a lognormal cumulative mixture distribution function was used of the form,

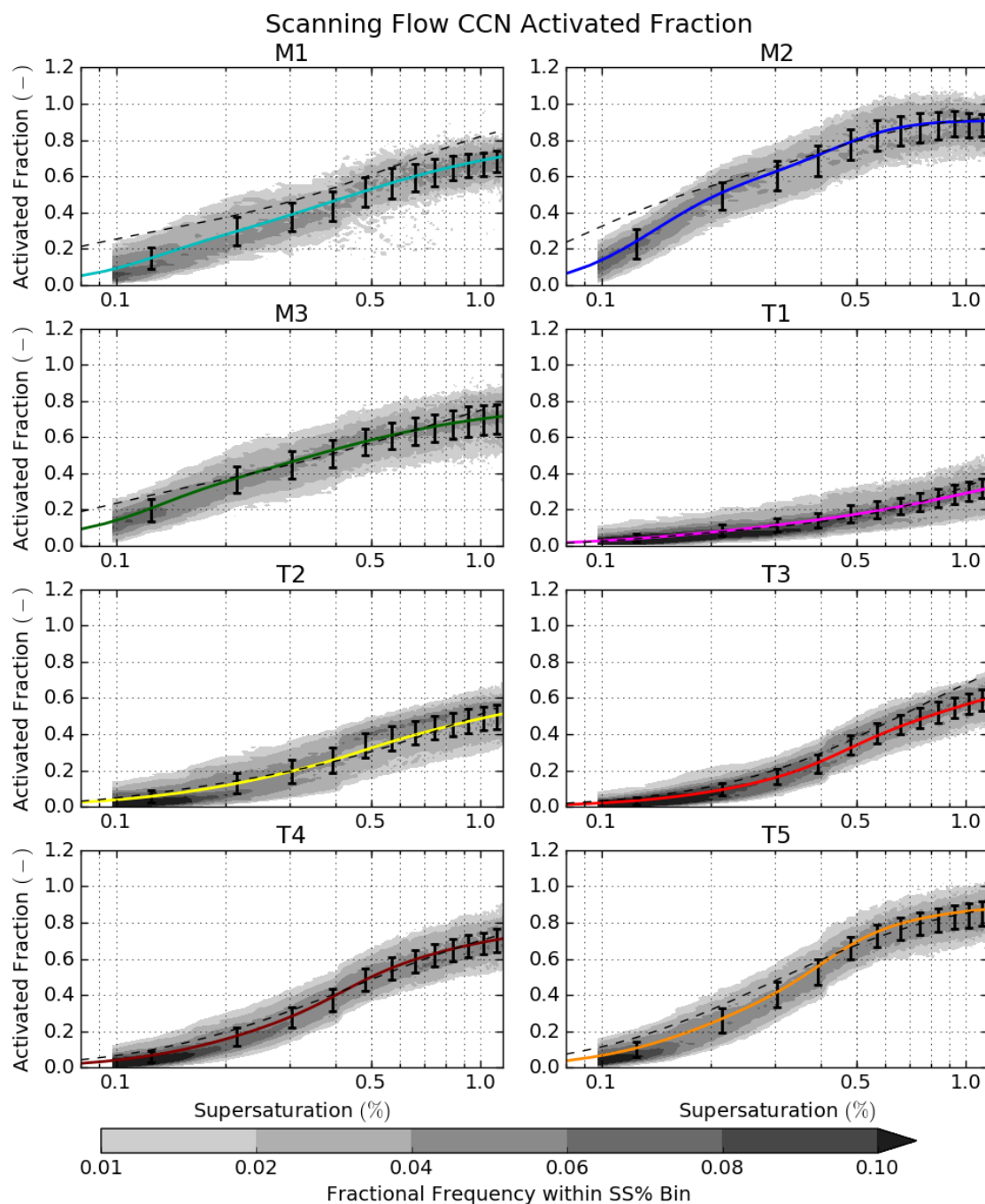
41    $$AF(ss)_{pop} = \sum_{i}^{FN} \frac{c_i}{2} + \frac{c_i}{2} erf\left(\frac{\ln(ss/a_i)}{\sqrt{2}\ln b_i}\right)$$

42   where AF is the parameterized activated fraction for the population, ss is the environmental supersaturation
43   expressed as a percentage ($ss = (S - 1) * 100\%$), FN is the total number of fit modes for each mode i, and a, b,
44   and c are the fit parameters for each lognormal mode. As most cluster average size distributions were bimodal with
45   only a small fraction of the total number in an additional third identified mode, a bimodal activation spectrum
46   parameterization was used for all population types here. The associated parameters for each of the activation spectra
47   shown in **Error! Reference source not found.** are given in **Error! Reference source not found.**.

48   **Additional Cluster Figures**

49   Similar to presentation of cluster measurements for size distribution and meteorological data, the activated fraction
50   spectra for each cluster are shown in Fig. S3. The best-fit distributions are shown as colored lines with activated
51   fraction data point density contours shown in greyscale around them.

52

53

Supplemental Figure 3 Activated fraction spectra for each cluster from the 8 cluster K-means analysis. Best-fit bimodal lognormal cumulative distribution functions are shown in colors for each cluster, with black error bars plotted at +/- 1 standard deviation of the mean for data split into 10 linear-spaced supersaturation bins with roughly equal numbers of data points. Grey contours show relative density of valid cluster data points in each supersaturation bin. Dashed black lines show the reconstructed activated fraction spectra based on average measured size distribution parameters and median measured kappa hygroscopicity parameter.

54

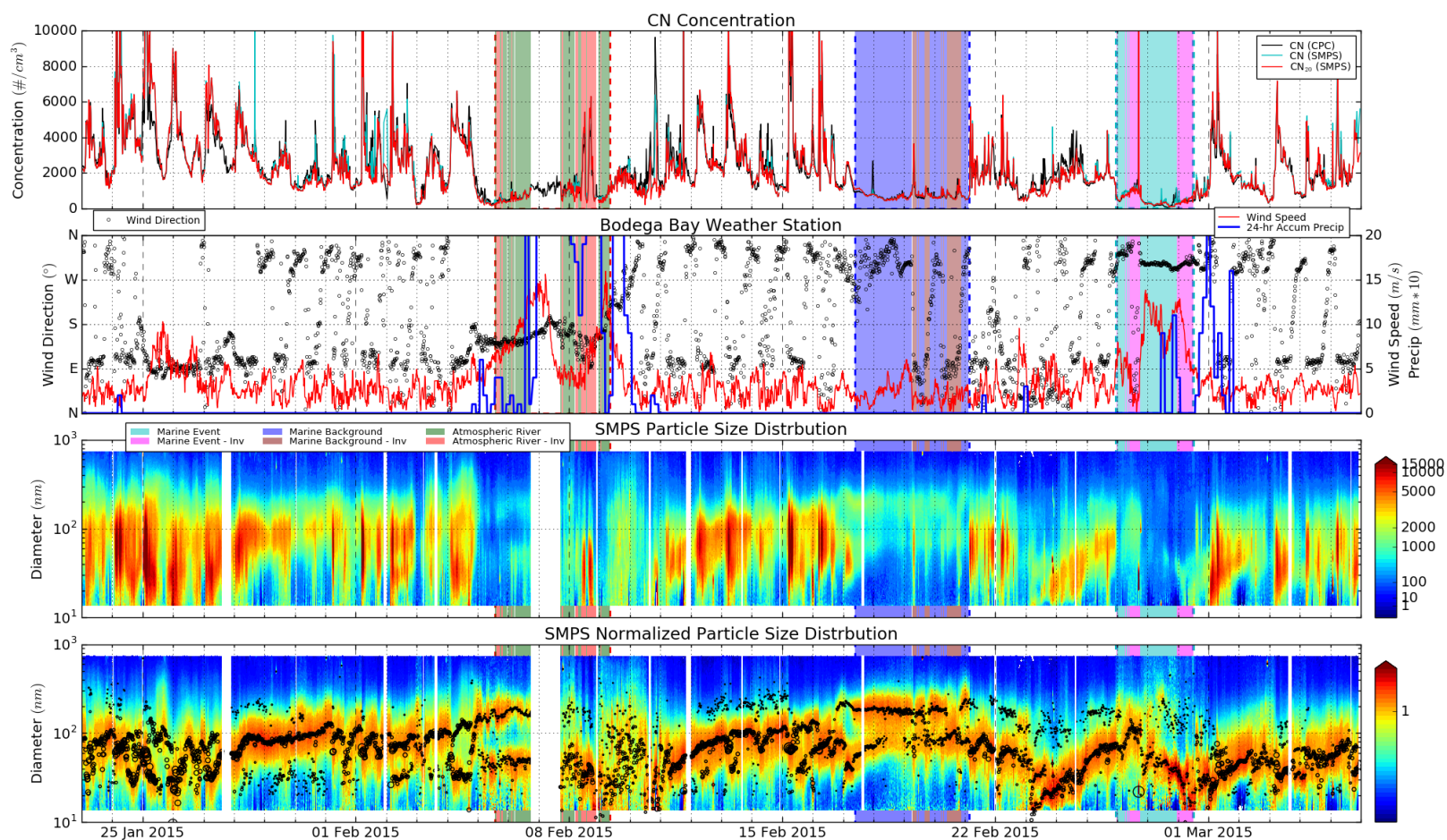
Validation of clustering methodology during marine aerosol dominated periods

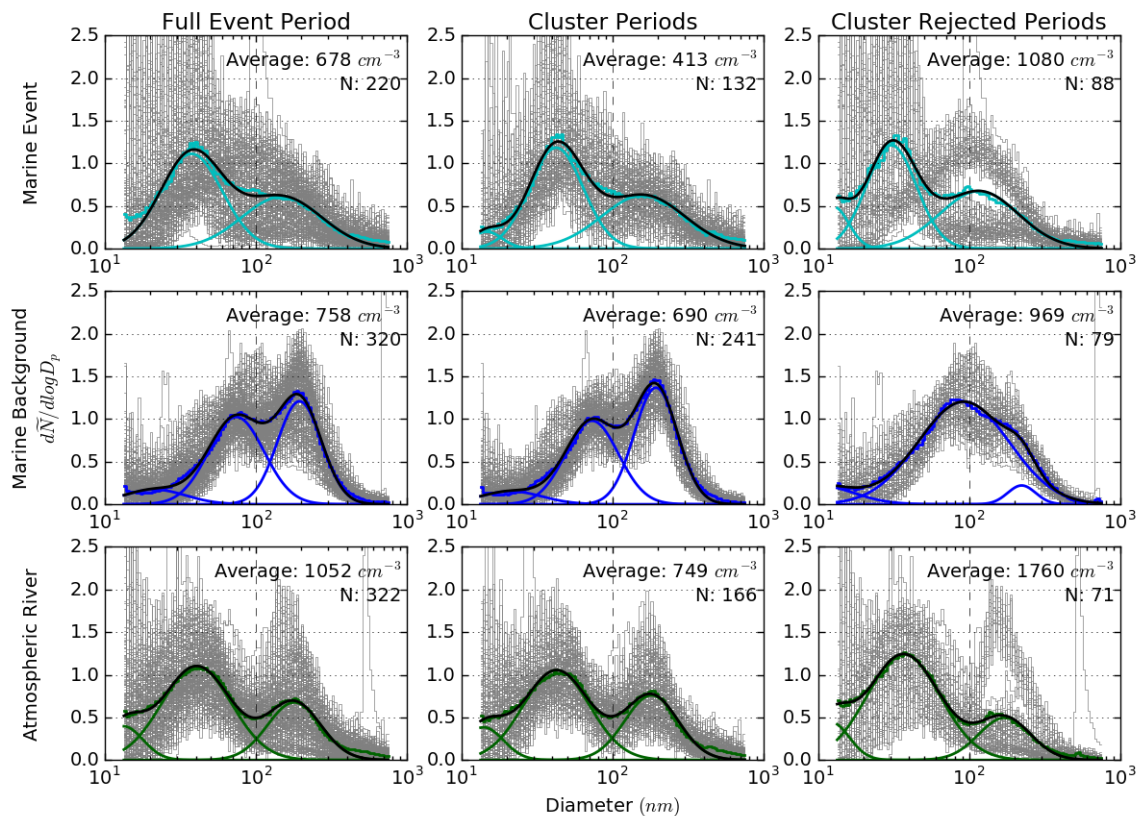
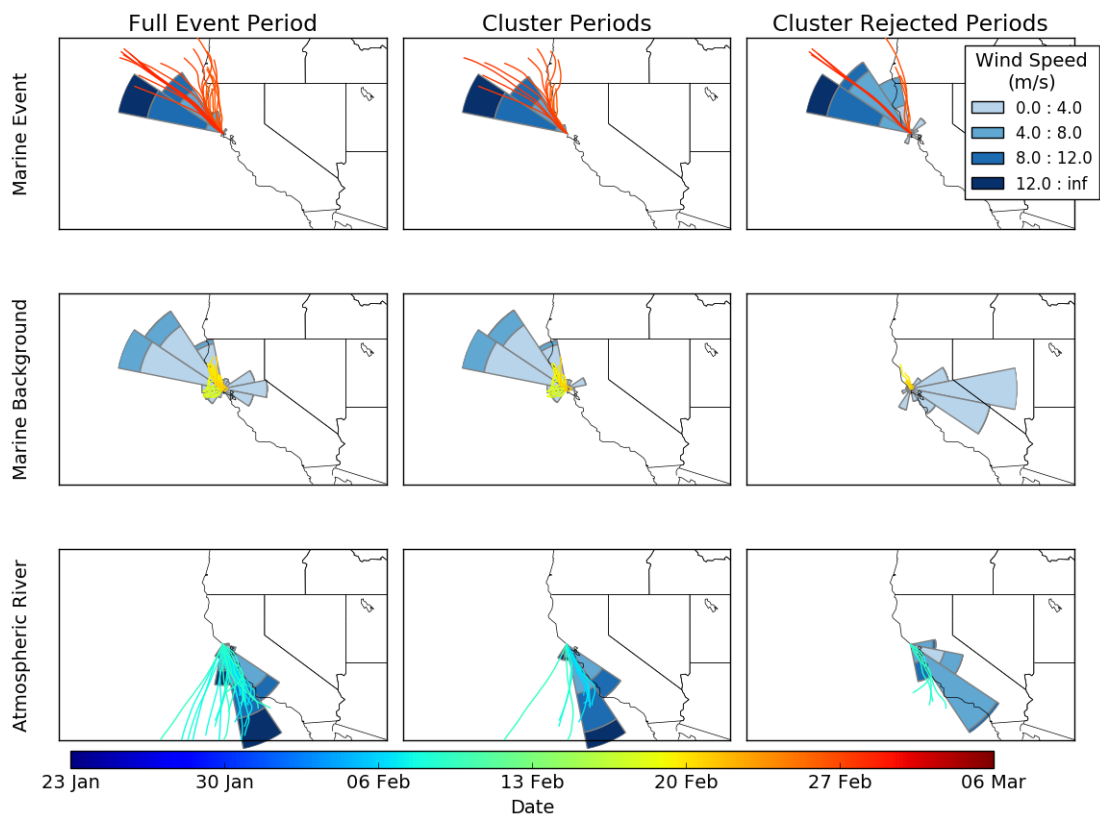
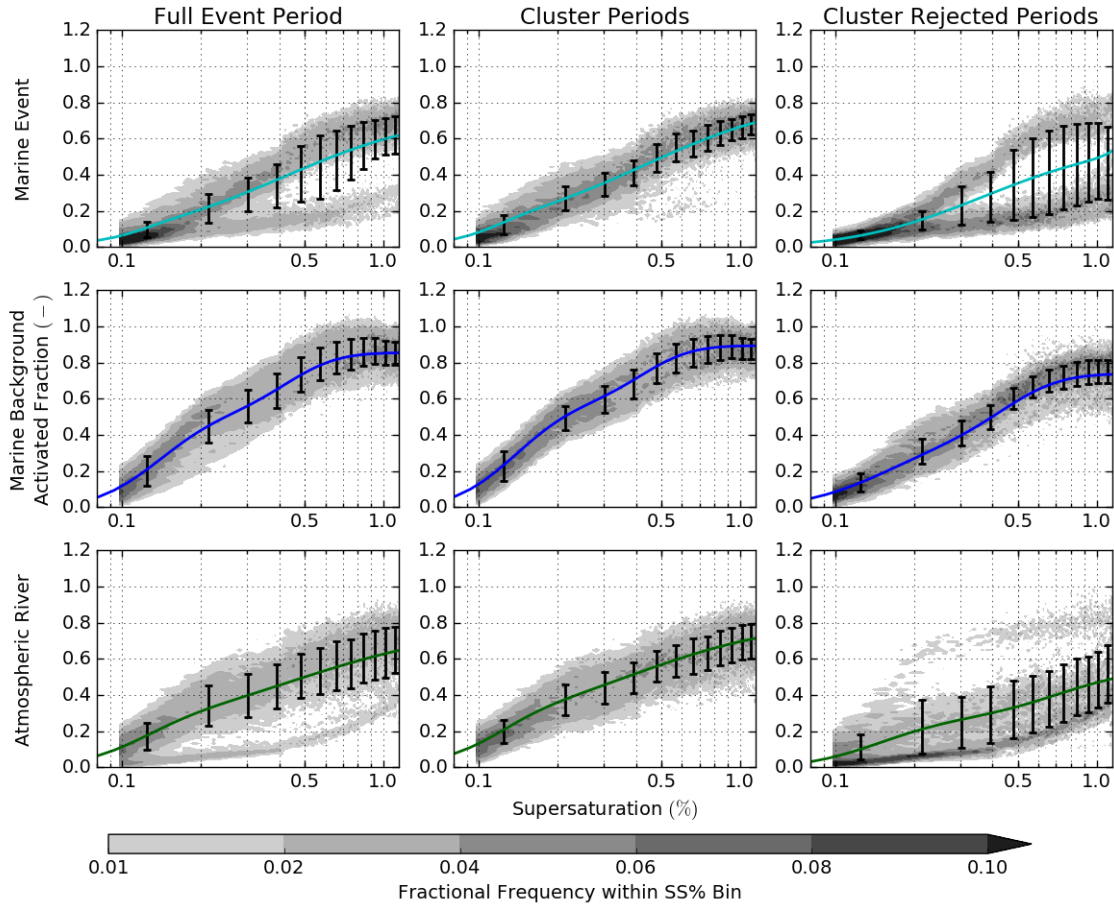
In order to validate that the ability of the clustering model to improve characterization of distinct aerosol population types, further analysis of the cluster results was performed for three multi-day time periods that were associated with marine aerosol dominated air masses impacting BML. As noted in the main text, these marine dominated periods were the precipitation scrubbed clean marine event (25 Feb 23:00 UTC – 28 Feb 12:00 UTC), the cloud-processed background marine event that involved ventilation of BML from oceanic regions (17 Feb 09:00 UTC – 21 Feb 03:00 UTC), and the atmospheric river (5 Feb 13:00 UTC – 9 Feb 08:00 UTC). These periods are identified in Fig. S4 for the periods associated with the cluster type associated with the marine event, along with periods of time classified as other cluster types.

Normalized size distributions for all data from each of the event time periods were generated and are shown in the left column of Fig. S5, all showing typical bimodal size distributions expected of cloud processed marine aerosol populations. The data points for these time periods were then segregated into only those corresponding to the marine cluster identified with each event (center column), and all data points from other cluster types that were excluded from that cluster (right column). In all three cases, noise in the range of distributions associated only with the cluster periods was reduced and average particle number concentration decreased. The data excluded from the marine clusters included some distributions that were largely monomodal and similar to the aged terrestrial distributions discussed in the previous section. Local wind vectors shown in Fig. S6 showed that the marine cluster data points were more generally associated with northwesterly winds from ocean regions, while the excluded data points occurred during periods of more easterly winds from inland regions.

The consequences to predicted CCN characteristics of removing non-marine data points are obvious in the resulting activated fraction spectra shown in Fig. S7. For the marine event associated with cluster M1 (top row), averaging all time points (left column) resulted in two distinct activated fraction curves, with the lower curve resulting from the inclusion of a period with small particles that had low activated fractions. Excluded data (right column) removed not only those data points, but also screened time periods with higher activated fractions that belonged to a different population type from M1. The resulting activated fraction spectrum associated with this event for only cluster periods (center column) had considerably smaller standard deviations in the result. Similar results were seen in the analysis of the M3 cluster event, with lower activated fractions associated with the terrestrial or anthropogenic contamination periods being effectively removed in the clustering activated fraction spectrum. While the M2 marine background event did not see as large of a change between time-averaging and cluster-averaging, the standard deviations were smaller in the clustering results, indicating increased precision in the activated fraction parameterization. Additionally, the excluded data points were themselves generally consistent and had low standard deviations, indicating that that a single separate population with distinct microphysical properties was likely intermittently observed during this event.

As a result, it was concluded that the clustering methodology using both properties of the aerosol and meteorological values was able to successfully segregate time periods associated with various aerosol population impacts. Additionally, it was able to provide improved identification of aerosol population types with less noise in the resulting parameterizations.

92

Supplemental Figure 4 Timeline comparison for three marine aerosol events, as identified by the reference classification (periods within dotted lines), and the K-means cluster analysis. Atmospheric river (green), marine background (dark blue), and precipitation impacted clean marine (light blue) events are shown. The invalid periods from the cluster classification are shown in red, brown, and pink, respectively.

93

Supplemental Figure 5 Normalized size distribution comparison for the three marine aerosol events (rows), averaged by the entire event time period (left column), only those times associated with the marine cluster associated with the event (center column), and those times that were excluded from the marine cluster (right column).

Supplemental Figure 6 Meteorological comparison for the three marine aerosol events as in Fig. S5.

Supplemental Figure 7 Activated fraction spectra comparison for the three marine aerosol events as in Fig. S5.