Long- and Short-Term Temporal Variability in Cloud Condensation Nuclei Spectra over a Wide Supersaturation Range in the Southern Great Plains Site

Russell J. Perkins,[1] Peter J. Marinescu,[1,2] Ezra J. T. Levin,[1,3] Don R. Collins,[4] and Sonia M. Kreidenweis[1]

[1]Colorado State University, Fort Collins, CO, 80523, USA
[2]Cooperative Institute for Research in the Atmosphere, Fort Collins, CO, 80523, USA
[3]Handix Scientific, Boulder, CO, 80301, USA
[4]University of California Riverside, Riverside, CA 92521

Correspondence to: Russell J. Perkins (rperkins@colostate.edu)

Abstract.

When aerosol particles seed formation of liquid water droplets in the atmosphere, they are called cloud condensation nuclei (CCN). Different aerosols will act as CCN under different degrees of water supersaturation (relative humidity above 100%) depending on their size and composition. In this work we build and analyze a best-estimate CCN spectrum product, tabulated at ~45 min resolution, generated using high quality data from ~~eight~~ seven independent instruments at the US Department of Energy Atmospheric Radiation Measurement (ARM) Southern Great Plains site. The data product spans a large supersaturation range, from 0.0001 to ~30%, and time period, 5 years from 2009-2013 and is available on the ARM data archive. We leverage this added statistical power to examine relationships that are unclear in smaller datasets. Our analysis is performed in three main areas. First, ~~Pp~~robability distributions of many aerosol and CCN metrics are found to exhibit skewed log-normal distribution shapes. Second, ~~Cc~~lustering analyses of CCN spectra reveal that the primary drivers of CCN differences are aerosol number size distributions, rather than hygroscopicity or composition, especially at supersaturations above 0.2%, while also allowing for simplified understanding of seasonal and diurnal variations in CCN behaviour. The predictive ability of using limited hygroscopicity data with accurate number size distributions to estimate CCN spectra is investigated and uncertainties of this approach are estimated. ~~Finally~~Third, the dynamics of CCN spectral clusters and concentrations are examined with cross-correlation and autocorrelation analyses.~~, which assist in determining the~~ We find that ~~time scales of changing~~ CCN concentrations change rapidly on the timescale of 1-3 hours, with some conservation beyond that which is greatest for the lower supersaturation region of the spectrum~~at different supersaturations and are important for cloud modelling studies~~.

1 Introduction

The interactions between atmospheric aerosol particles and ambient water vapor are key drivers of the formation of haze and clouds. Particles are the nuclei upon which liquid water first condenses to form haze and cloud droplets, thereby affecting visibility, cloud microphysical properties, and precipitation. Water uptake by particles depends upon their size and chemical

composition as well as on ambient environmental conditions, and on their rates of change. Particles that are identified as cloud condensation nuclei (CCN) are typically those that are predicted to form cloud drops at water supersaturations (SS_w) of 1% or lower, conditions believed typical of most clouds that are formed in weak to moderate updrafts. Based on average aerosol characteristics, and typical atmospheric aerosol number size distributions (shortened to size distributions hereafter), number concentrations of particles active as CCN are thus generally assumed to correspond to those particles in the 80-300 nm dry diameter size range. Instruments designed to directly measure number concentrations of activated particles at fixed SS_w are also typically limited to 2% SS_w as an upper bound on the measurement range (Uin, 2016).

However, supersaturations and particle sizes outside of these ranges are also of atmospheric interest. In deep convection with intense updrafts, and in regions of very low existing particle or droplet surface area concentrations, SS_w of water vapor can build rapidly to high levels as the condensation sink rates are so low relative to the rate of supersaturation generation (Pinsky et al., 2012). Those high SS_w may be sufficient to allow particles smaller than 40 nm to serve as cloud condensation nuclei. Thus, despite their relatively short lifetimes in the atmosphere compared with larger accumulation-mode particles, high concentrations of small particles can potentially influence cloud microphysical processes leading to precipitation formation or evaporation. It is now recognized that nucleation of new particles from the gas phase, generating particles on the order of 10 nm in diameter which subsequently grow, occurs in many regions of the troposphere and is an important control on global atmospheric aerosol number concentrations (Hodshire et al., 2016; Bianchi et al., 2016; Venzac et al., 2008; Pierce et al., 2014; Nieminen et al., 2018).

At the other end of atmospherically-relevant supersaturations, droplet formation may occur at very low SS_w where CCN concentrations are very difficult to probe. For example, the slow cooling rates in radiation fogs allow vapor scavenging to effectively compete with the generation of supersaturation, and thus maximum SS_w conditions reached in fogs can be below 0.05% (Gerber, 1991; Low, 1975; Shen et al., 2018), suggesting that only larger and more hygroscopic particles can participate in fog droplet formation. The cloud physics community has had a longstanding interest in elucidating the microphysical roles of "giant CCN" (GCCN), that is, relatively large particles that activate at very low SS_w, ~0.01% or less. Specifically, GCCN are hypothesized to control the initiation of drizzle and precipitation in shallow clouds (Cohard et al., 1998; Johnson, 1982; Feingold et al., 1999; Cheng et al., 2009; Hudson et al., 2011; Posselt and Lohmann, 2008; Levin and Cotton, 2009; Gantt et al., 2014; Jung et al., 2015).

Their potentially controlling roles in fog, cloud, and precipitation formation has motivated interest in direct measurements of the number concentrations of CCN active over a range of atmospherically-relevant SS_w, with modern instrumentation making possible long-term, unattended monitoring. Here, we analyse observations of CCN spectra from the United States Department of Energy's Atmospheric Radiation Measurement's (ARM) Southern Great Plains (SGP) site located in north central Oklahoma for the five-year period from 2009-2013. The CCN measurement instrumentation deployed at this site is

2

typically limited to stable operation over the 0.1 to 1% SS_w range. As described further below, we extend those observations to a broader supersaturation range of interest using ancillary aerosol observations, described briefly in Table 1, creating a CCN estimate relevant for clouds ranging from fog through to intense deep convective updrafts. These observations are especially important for modelling studies of aerosol impacts on clouds in this region that use CCN number concentrations as a basis for determining their aerosol initial conditions (Saleeby et al., 2016; Marinescu et al., 2017; Glenn et al., 2020). This extensive CCN dataset is subsequently analyzed using methods to leverage its range and statistical power to characterize and understand the statistical distributions, seasonal and diurnal variations, and dynamics of CCN spectra at this site.

From Reviewer 2: Add a few sentences to state the scientific goals of this study and the organization of this manuscript to the end of Sec.1

2 Methods

This work builds upon prior work reported in Marinescu et al. (2019). In that study, for the same 2009-2013 time period studied herein, aerosol data from three instruments deployed at the SGP site were averaged over ~45-minute intervals and merged into dry aerosol number size distributionsize distributions, $n(D_p)$, spanning 7 nm < particle dry diameter, D_p < 14 µm. We build upon that work by combining those merged size distributions with information on aerosol hygroscopicity, κ (Petters and Kreidenweis, 2007), taken over the same ~45-minute intervals. κ values were obtained via measurements of diameter growth factors (GF) at 90% relative humidity measured by a humidified Tandem Differential Mobility Analyzer (H-TDMA; Collins, 2010; Mahish and Collins, 2017). The procedure for integrating these data into a CCN spectrum is described more fully in Appendix A, but is reviewed briefly here. All base instrument measurements used in data construction are shown in Table 1, along with a brief description of their associated measurement and aerosol size range probed.

Table 1: Base Measurements Used in Data Set Construction

Instrument	Measurement	Measurement Size Range (µm)
SMPS Scanning Mobility Particle Sizer	Aerosol Size Distribution	0.012 - 0.750 (mobility diameter)
APS Aerodynamic Particle Sizer	Aerosol Size Distribution	0.5 – 20* (aerodynamic diameter)
CPC Condensation Particle Counter	Total Aerosol Number	>0.007*
HTDMA Humidified Tandem Differential Mobility Analyzer	Aerosol Growth Factor / Hygroscopicity	0.012 - 0.750 (mobility diameter)
CCNC Cloud Condensation Nuclei Counter	CCN Spectrum	Dependent on operating SS_w
ACSM	Aerosol Mass Composition	0.040 – 1

3

Aerosol Chemical Speciation Monitor		(vacuum aerodynamic diameter)
Nephelometer	Aerosol Light Scattering	

*upper size cut will depend on instrument inlet losses.

The critical saturation ratio, S_c, at which a particle with dry particle diameter D_p can be activated into a cloud drop is determined by finding the maximum of the following equation (Petters and Kreidenweis, 2007):

$$S(D) = \frac{D^3 - D_p^3}{D^3 - D_p^3(1 - \kappa)} \exp\left(\frac{A}{D}\right) \qquad \text{Eq. 1}$$

$$A = \frac{4\sigma_{s/a}M_w}{RT\rho_w} \qquad \text{Eq. 2}$$

where D is the droplet diameter, ρ_w is the density of water, M_w is the molecular weight of water, $\sigma_{s/a}$ is the surface tension of the solution/air interface, R is the universal gas constant, and T is temperature. Saturation ratio, S, and water supersaturation percent, SS_w are related by:

$$SS_w = (S - 1) * 100 \qquad \text{Eq. 3}$$

The functional relationship between D_p, κ, and S_c can be more readily illustrated by the approximate relationship that is valid for $\kappa > 0.2$ (Petters and Kreidenweis, 2007) where $SS_{w,c}$ is the critical water supersaturation percent:

$$\frac{SS_{w,c}}{100} \cong \left\{\frac{4A^3}{27D_p^3\kappa}\right\}^{1/2} \qquad \text{Eq. 4}$$

From Eq. 4 it can be seen that information on the aerosol ~~number size distribution~~ size distribution, $n(D_p)$, and the variation of κ with dry diameter, can be used to compute the critical SS_w corresponding to each selected dry size used in Eq. 4. It should be noted that the approximation in Eq. 4 is only used for demonstration here, with full methods described in Appendix A. In practice, the size distribution is discretized to obtain total number concentrations in each selected dry diameter bin, 222 logarithmically spaced bins in this work, and a constant κ is assumed across each selected bin. To produce the size-dependent κ distribution, measurements of aerosol hygroscopic growth were made for 7 different sizes (Figure A2), which were subsequently processed to obtain a single weighted κ value for each size. In this study, these κ values are then interpolated linearly between measured values, with invariant κ beyond the largest and smallest sizes (below ~10 nm and above ~600 nm). The cumulative spectrum of CCN concentration, CCN(SS_w), can then be constructed, which defines the total number of

4

particles that can be activated at a particular SS_w. Figure 1a shows an example of the measured $n(D_p)$ and $\kappa (D_p)$, and the resulting CCN spectrum is shown in Figure 1b.
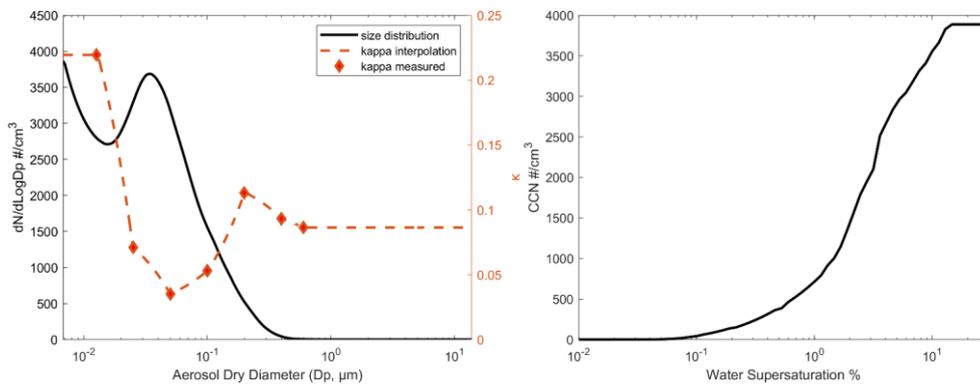
Available data from the measurement suite include direct, concurrent, and collocated observations of cumulative CCN number concentrations at selected supersaturations, which were used to check the accuracy of the initial CCN spectra, reconstructed from size distributions and hygroscopicity measurements discussed above, between 0.1% and 1%
110 supersaturation and adjusted as needed. Two additional instruments provide separate, independent, continuous observations that were used to constrain the reconstructed spectra: a nephelometer measured total particle scattering coefficients, and an Aerosol Chemical Speciation Monitor (ACSM) measured nonrefractory, speciated submicron mass concentrations. ~~Available data from the measurement suite include direct, concurrent, and collocated observations of cumulative CCN number concentrations at selected supersaturations, which can be used to check the accuracy of the reconstructed spectra. Two~~
115 ~~additional instruments provide separate, independent, continuous observations that were also used to assess the reconstructed spectra: a nephelometer measured total particle scattering coefficients, and an Aerosol Chemical Speciation Monitor (ACSM) measured nonrefractory, speciated submicron mass concentrations.~~ Both of these observations emphasize larger particle sizes (generally > 300 nm) and thus served as constraints on the particles contributing strongly at the lowest supersaturations, for which no direct observations exist. At the other end of the size distribution, the smallest particles are expected to require the
120 highest supersaturations for activation, but neither the size distribution nor κ are well constrained observationally. Reasonable assumptions are applied to extrapolate the CCN spectrum beyond SS_w=1%, the approximate upper limit of the CCN counter.

5

In total, data from ~~six~~ five instruments are merged then constrained with observations from two additional instruments to produce a best-estimate CCN spectrum for each ~45 minute interval, as described more fully in Appendix A. Resulting size distributions range from 0.0068 to 13.8 µm (bin centers), with corresponding CCN spectra generated for SS_w from 0.0001 to ~30 percent SS_w (100.0001 to 130 percent RH). The SS_w range is chosen to span the entire range of particle activations: the largest and most hygroscopic measured particles activate at 0.0001% SS_w while the smallest and least hygroscopic at ~30% SS_w. As noted in the Data Availability statement below, the final merged data~~, with estimates of uncertainty,~~ are available in the DOE ARM archive. This dataset is subsequently analyzed using several different methods. Results of these analyses are discussed in section 3, below, while the details of analytical methods are found in Appendix B, for skewed log-normal fitting procedures, Appendix C, for clustering analysis, and Appendix D, for non-periodic autocorrelation and fits.

~~Reviewer 2: Augment the method section – At least move some of the basic descriptions of the three statistical analyses from the appendix to the method section.~~

<div style="color:gray; font-size:smaller; float:right;">Formatted: Indent: Left: 0", First line: 0"</div>

<div style="color:gray; font-size:smaller; float:right;">Formatted: Indent: First line: 0"</div>

3 Results and Discussion

3.1 Distribution Characteristics

While we expect the final merged data over the entire range to be useful, all observations are not equally reliable for several reasons. In the size distributions, the lowest size bins are generated using a fitting procedure previously described by Marinescu et al. (2019). This fitting procedure is constrained by instrument data, and produces good agreement with direct observations of the number concentrations of particles in the ultrafine mode (Marinescu et al., 2019), but the shape of the aerosol spectrum at the smallest particle sizes, especially below 12 nm, is more uncertain than at larger particle sizes. κ values are assumed to be invariant outside of the measured range (below ~10 nm and above ~600 nm) and are thus more uncertain in those ranges. Additionally, at larger aerosol sizes, growth factor distributions can be bi-modal (Figure A2) which is not captured in this approach, and results in additional uncertainty. Because the CCN spectra are generated by combining size and κ distributions, uncertainties in each are inherited in certain regions. The high SS_w region of the CCN spectrum is dominated by smaller particles, and uncertainty in this region is increased due to uncertainties in aerosol distribution shape. Beyond a certain SS_w that is sufficient to activate all particles regardless of size and composition, a CCN spectrum must level off. In cases where there are particles present in the smallest size bins, this region has increased uncertainty due to the effects of very small particles (below ~8 nm) that were not measured. This occurs at SS_w greater than 10%. Interestingly, uncertainties in κ in this region are largely irrelevant as high SS_w is required for activation regardless of hygroscopicity (see discussion in section 3.1.2, Figure 4). The region of lowest SS_w, less than 0.1%, is the other region subject to additional uncertainties, coming from uncertainties in κ and ~~number size distribution~~size distributions. Particles >13.8 µm were not measured~~,~~ and are not

6

included in our size distributions, and these particles this lowest SS_w region. Additionally, particles larger than several microns are rare and subject to significant shot noise, even over 45-minute sampling intervals. In the low SS_w region, the kappa values of these large particles have significant impacts on critical SS_w which was not measured above 600 nm. Compared with aerosol counting uncertainties, however, this is a lesser issue. For example, an error in kappa from 0.1 to 0.4 for a 5 µm particle shifts critical SS_w from 0.001 to 0.0005, producing error only in that region, which is not propagated outside of it. On the other hand, undercounting large aerosol produces a downward shift in CCN number concentration across the entire SS_w spectrum (greater than the critical SS_w value), which can be quite significant at lower SS_W values. While the abundances of large particles which activate in this low SS_w regime are quite low, they can be important in controlling further activation under some conditions. Ultimately, the regions of increased uncertainty are under SS_w conditions where few measurements of CCN concentrations exist, which adds value to these data despite these uncertainties. Although further study is needed to fully constrain CCN concentrations in the high and low SS_w regions, this work provides best-estimate values that can be used for analysis or modelling of strong updraft (high SS_w) or precipitation initiation (low SS_w) scenarios.

3.1.1 Probability Fitting

We present statistical descriptions for key parameters related to the aerosol and CCN populations to describe their variability over the study period. Figure 2 shows occurrence probability (y axis) of the number concentration of CCN active at 1% SS_w. Similar distribution shapes are observed for all variables examined (total distribution number concentrations, total distribution volume concentrations, number concentrations in 100 and 1000 nm individual bins, CCN concentrations at 0.01% and 0.1% supersaturation; Appendix B).

7

Formatted: Not Superscript/ Subscript

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

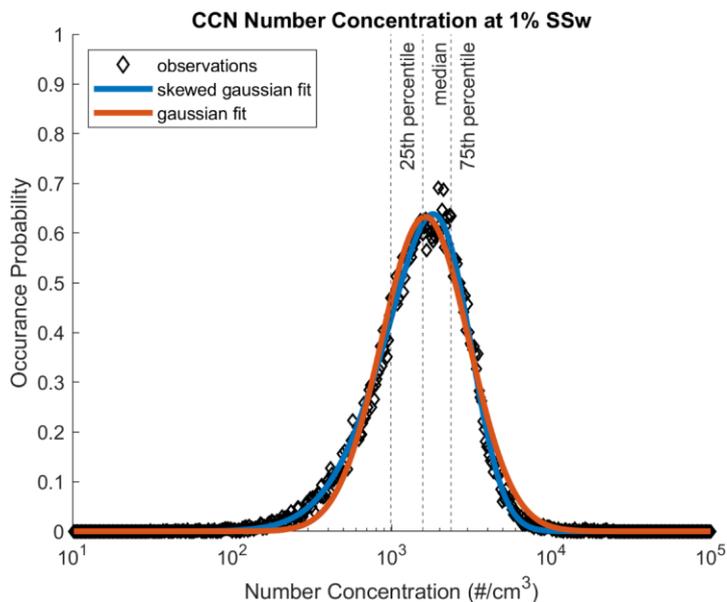**Formatted:** Font: Italic

Formatted: Not Highlight

Figure 2: Distribution of CCN number concentrations active at 1% SS_w over the course of the study. The distribution fits a skewed Gaussian model exceptionally well, and an un-skewed Gaussian model with moderate fidelity.

The frequency distributions of observations fit exceptionally well to skewed log-normal distributions (described in Appendix B), with low degrees of skew such that log-normal distributions remain a fair approximation in many cases. Aerosol data are seldom fit in this manner, simply reporting median and percentile bounds instead. On the other hand, log-normal distributions have been noted and used for Aerosol Optical Depth (AOD), either for spatial or temporal variations (Alexandrov et al., 2004, 2016; Anderson et al., 2003; Sayer and Knobelspiesse, 2019). While the methods used for AOD treatments have not been widely adopted for aerosol distributions or CCN spectra, they could be, and hopefully supplying parameterizations in the same form makes further work that is focused on impacts of variability more accessible. Additionally, better fits for probability distribution functions could be incorporated into microphysical modelling studies or other efforts interested in the likelihood of given aerosol conditions occurring. When using these fit data, it is important to keep in mind that neighbouring size bins are statistically correlated with each other – the probability of finding 100 particles in the 10 nm size bin is not independent of the probability of finding 100 particles in the 15 nm size bin. Because of this, the simplest way to calculate combined or correlated quantities (for example, the number concentrations of all particles between 10 and 20 nm) is through

8

185 our archived distributions across time points of interest, rather than utilizing our fit parameters. A similar skewed log-normal distribution could be fit for the combined data if desired. It should also be noted that distribution shapes may not be well conserved across all time or length scales. Variations are most likely to occur at small time (less than 2 hours) or length (less than 0.5 km) scales based on the analysis discussed in Section 3.2 and previous works (Alexandrov et al., 2004; Anderson et al., 2003). Finally, it is important to emphasize that the uncertainties in the CCN spectra discussed in Section 3.1 are not

190 necessarily reduced by this statistical fitting approach, due to their potentially systematic rather than random nature.

3.1.2 Clustering Analysis

Clustering analysis is used to simplify and seek relationships in the rather large and complicated data set. K-means clustering is performed using a vector-based distance metric. Details of cluster analysis can be found in Appendix C. From the K-means clustering applied to the CCN spectra, three distinct clusters are identified that achieve good separation in both CCN

195 and ~~number size distribution~~size distribution characteristics, as shown in Figure 3. The systematic uncertainties in our CCN distributions discussed in Section 3.1 are expected to be inherited into the characteristic cluster spectra shown here.
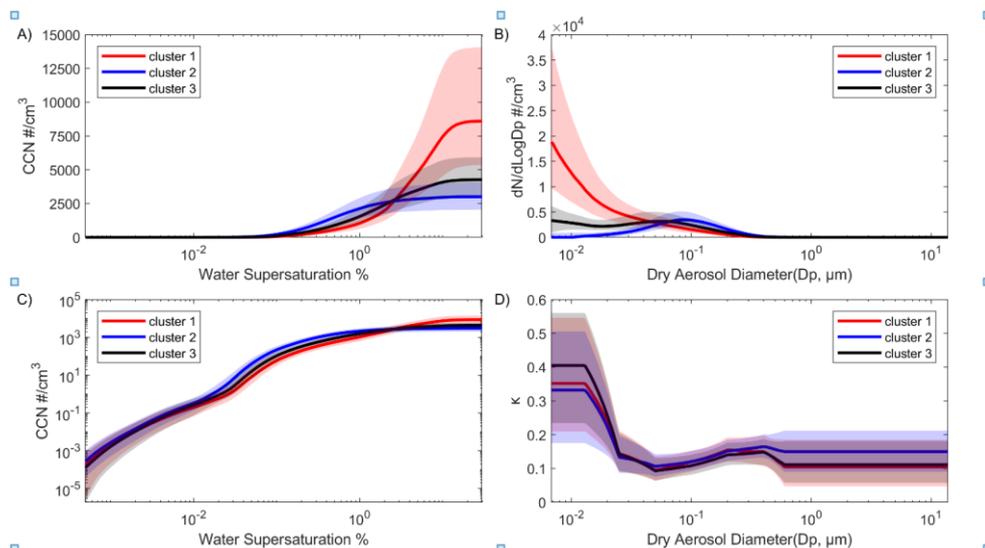

Figure 3: Clusters generated by k-means clustering procedure. Cluster centres are shown by solid lines, with shaded region representing 25th and 75th percentile bounds for spectra associated with each cluster. Panels A) and C) show cluster CCN spectra,
200 **panel B) shows cluster size distributions, and panel D) shows cluster hygroscopicity parameter κ.**

9

Clustering is carried out based on CCN spectra – that is, each spectrum was assigned to a cluster based on its shape and magnitude. Even though the clustering procedure had no direct information about particle ~~number size distribution~~size distributions, the size distributions associated with each cluster are well resolved (Fig. 3b). On the other hand, hygroscopicity parameter distributions are similar for all three clusters (Fig. 3d). This indicates that particle ~~number size distribution~~size

205 distributions have a greater influence on the resulting CCN spectra than do κ distributions, consistent with other analyses (Patel and Jiang, 2021). As a result, estimates of CCN spectra using size distribution data and either estimated or median κ values are expected to be reasonable approximations, although deviations of approximate CCN spectra from observed CCN spectra can still be quite large for any given timepoint. We estimate the error introduced using a "median κ" to compute CCN spectra, as follows. For this "median κ" estimate, we calculated a median κ value based on the entire data set, then used this median

210 value in combination with all individual size distributions to generate estimated CCN spectra. These are then compared to the CCN spectrum products (using concurrent κ and size distribution data) to calculate error estimates, with the results shown in Figure 4. Estimates are generally least reliable for lower supersaturations, with estimates below 0.2% SS_w having a 95% confidence interval broader than ±50% of the estimated value. Therefore, caution should be taken interpreting estimated CCN spectra in this low SS_w region. This highlights the region of the CCN spectrum that is most sensitive to observed variations in

215 κ (Mahish and Collins, 2017), and~~This result additionally highlights~~ the uncertainty in our data product below about 0.03% SS_w. In this region, particles significantly larger than 600 nm are expected to activate but we do not have accurate κ measurements for these sizes, as discussed in Section 2. Above 0.2% SS_w, the "median κ" estimate works quite well, with uncertainties decreasing as SS_w increases. This region of the spectrum is likely a good candidate for generation of CCN spectra from observations of particle distributions where high-quality κ measurements are available for only a limited time period. It

220 is important to note that this approach will only work for accurate ~~number size distribution~~size distribution data extending to diameters larger than 500 nm. For distributions ending at 500 nm, many CCN activating at or below ~0.2% SS_w will not be directly counted, and due to the cumulative nature of the distributions, this gap can introduce large errors for all SS_w values. We expect this median κ estimation method to be especially applicable for the SGP site and similar environments, but it may apply elsewhere as well.
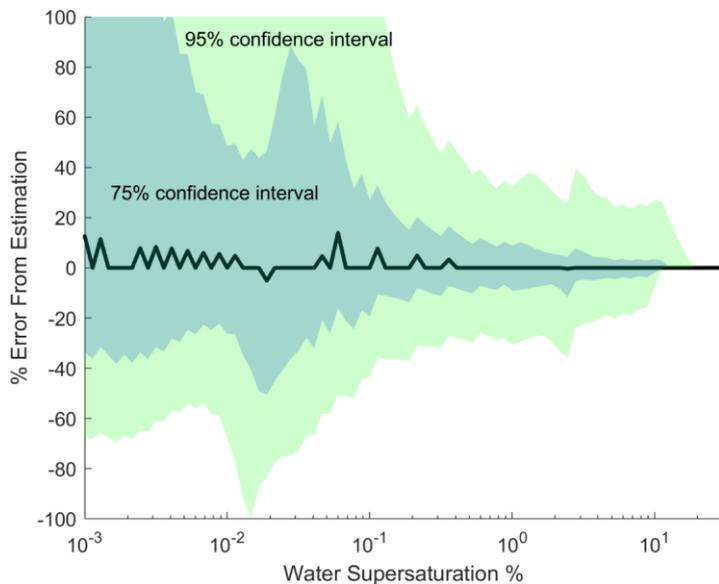
10

225

Figure 4: Error estimation of CCN product constructed from median hygroscopicity data, as compared with that computed for size-dependent κ. The black line depicts the median error, while the dark and light green shaded regions depict 75 and 95% confidence intervals, respectively.

The clustering of CCN spectra into distinct groups highlights the contributions from particle ~~number size distribution~~size distributions as the distinguishing factor. This analysis does not suggest that κ distributions for small particles are invariant, random, or unimportant, only that their contribution to a final CCN spectrum is small compared to the contribution of the particle ~~number size distribution~~size distribution. Mahish and Collins (2017) provide a more complete analysis of the κ measurements at SGP during this time period, which is consistent with the data we use here.

Because of the factors discussed above, the different CCN clusters represent different characteristic particle ~~number size distribution~~size distributions. Cluster 1 represents cases where the nucleation mode particles, associated with NPF events, dominate the ~~number size distribution~~size distributions. Cluster 1 also has lower absolute number concentrations of accumulation mode particles than found in the other clusters. Cluster 2 represents the opposite case – the absence of small particles and higher accumulation mode number concentrations combined with a shift of the accumulation mode to larger sizes. Cluster 3 is the intermediate case, with some nucleation mode particles and a substantial accumulation mode. The three clusters represent approximately equal portions of the total number of observations. Because these clusters represent three

11

different scenarios quite well, we will use them to simplify further discussion. Cluster 1 will subsequently be referred to as the nucleation cluster, cluster 2 the accumulation cluster, and cluster 3 the intermediate cluster.

3.1.~~3~~2 Seasonal and ~~Hourly~~ Diurnal Trends
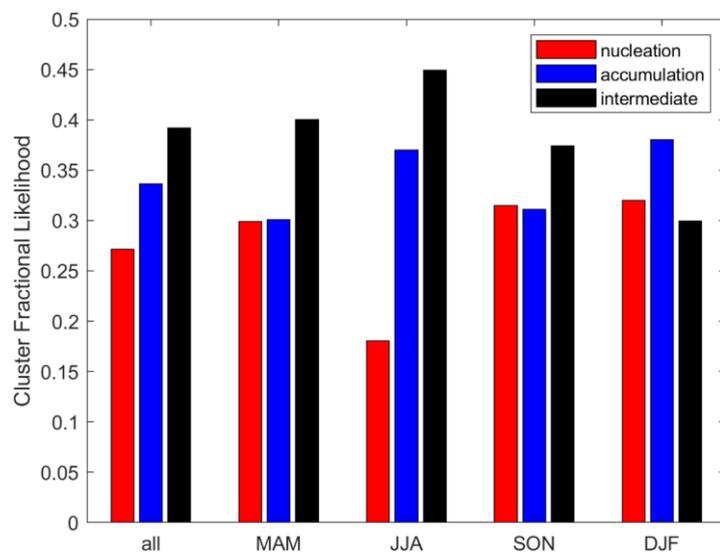
245   **Figure 5: Cluster fractional likelihood for each season.**

The clustered data are examined for seasonal variations in particle and CCN characteristics. Due to the differential nature of comparisons between clusters, the effect of uncertainties discussed in Section 3.1 is likely minimized. Cluster prevalence shows some seasonal dependence, although all clusters are still found a significant portion of the time for all seasons (Fig. 5). Summer (JJA) and winter (DJF) seasons show the highest prevalence of accumulation clusters, but significant differences in
250   fractions of the intermediate and nucleation clusters. Summer has the highest prevalence of the intermediate clusters, while winter has the highest prevalence of nucleation clusters. This suggests that during the summer significant particle concentrations are more likely to coexist in both the accumulation and nucleation modes, or perhaps that growth of nucleation mode particles to larger sizes (i.e., transfer of particles to the accumulation mode) is more likely to occur. These trends are not obvious from looking at seasonal particle data alone (Marinescu et al., 2019). An important consideration for reconciling

12

255 seasonal particle data, discussed in (Marinescu et al., 2019), and seasonal cluster trends is the fact that the distributions within a cluster will have seasonal dependence as well, as shown in Figure C1.
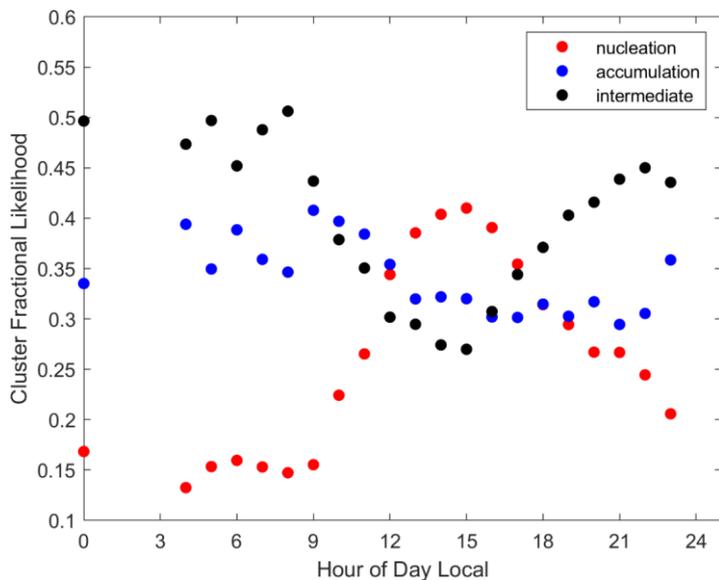
13
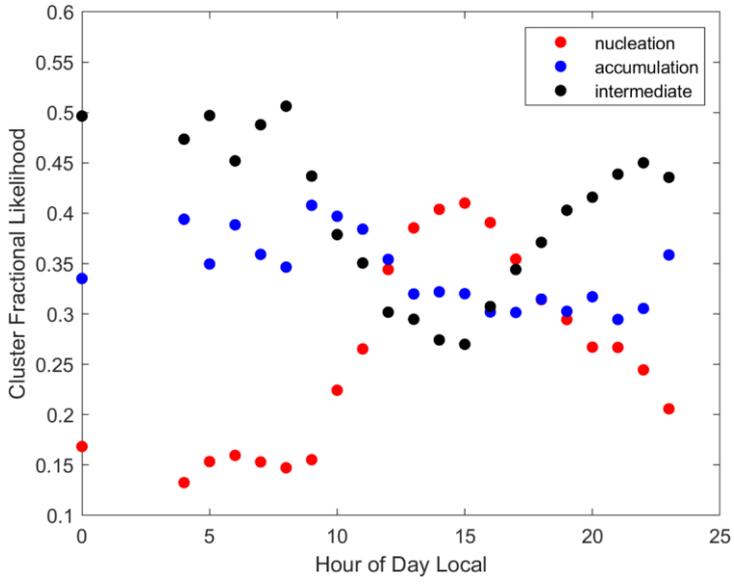
Figure 6: Diurnal variability of cCluster fractional likelihood for each local time of day.

Figure 6 depicts how cluster prevalence changes as a function of the time of day. Nucleation mode clusters are most common during daylight hours, with intermediate clusters most likely at night. Interestingly, accumulation clusters show the least time-dependence over the course of the day. Analysis of CCN concentrations at several supersaturations, as shown in Figure 7, shows no hourly dependence in median values at 1% SS_w or lower. At 10% SS_w, the hourly trend in CCN is similar to the hourly trend in the nucleation cluster. These data combined suggest that the hourly changes that occur are due to addition of nucleation mode particles, rather than changes in the particle concentrations of other modes. Seasonal variability in the nucleation mode time-of-day dependence is shown in Figure C1 and reflects the same overall time dependence within a day, alongside seasonal changes shown in Figure 5.

15
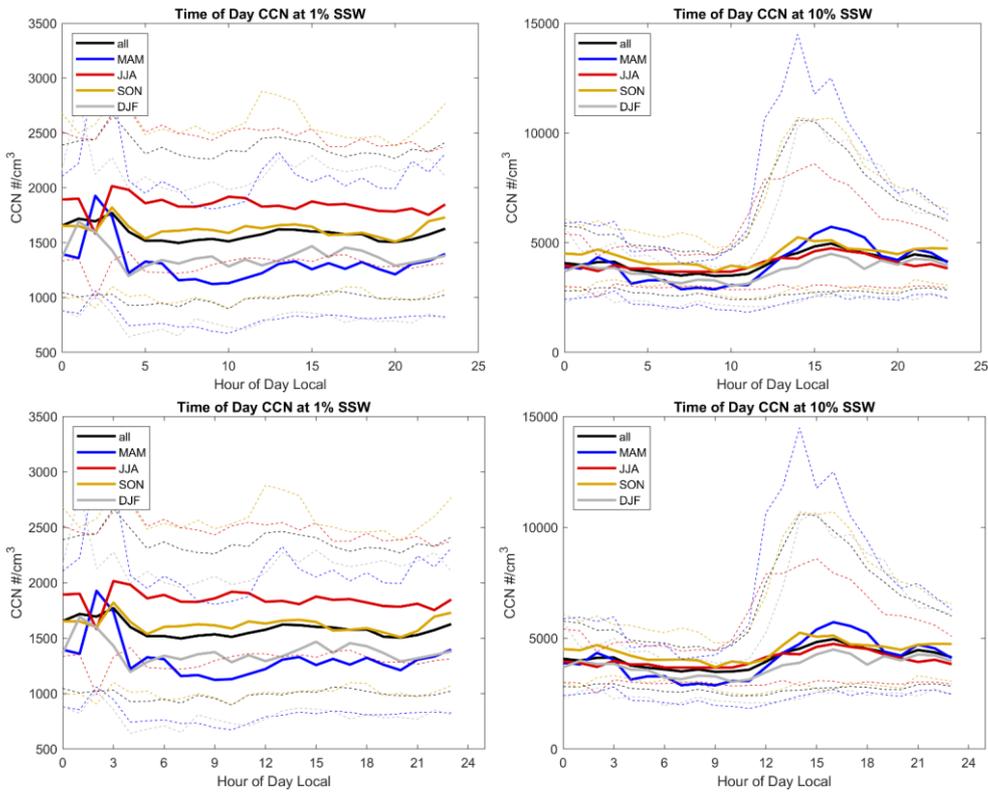
Figure 7: Median (solid lines) and 75% confidence intervals (dotted lines) seasonally averaged CCN concentrations at 1% (left) and 10% (right) SS_w, as a function of local time of day.

Another way to examine seasonal changes is through comparison of occurrence probabilities at a single supersaturation, as shown in Figure 8. Figure 8 shows this information in two similar ways: the skewed log-normal fits from Section 3.1.1 and Appendix B, and more traditional box plots. Both methods of parameterizing the data require the same number of parameters (3 coefficients, a zero fraction and a correlation coefficient for the fit; 5 points for the box and whiskers), but the fits convey more information. Seasonal differences are somewhat obscured by the box plot, but for cumulative CCN active at 1% supersaturation, a clear difference between summer months and the rest of the year is observed with the fits, where CCN concentrations are more tightly grouped at higher values in the summer. Fit parameters for all supersaturations and seasons

16

can be found in the Supporting File CCN_fit_coeffs.txt. The fits derived for cumulative CCN active at high supersaturations (1% and higher) are relevant to cases of deep convection, whereas those derived for very low supersaturations (below 0.1%) may be helpful for estimates of the abundances of particles in special populations such as giant CCN.
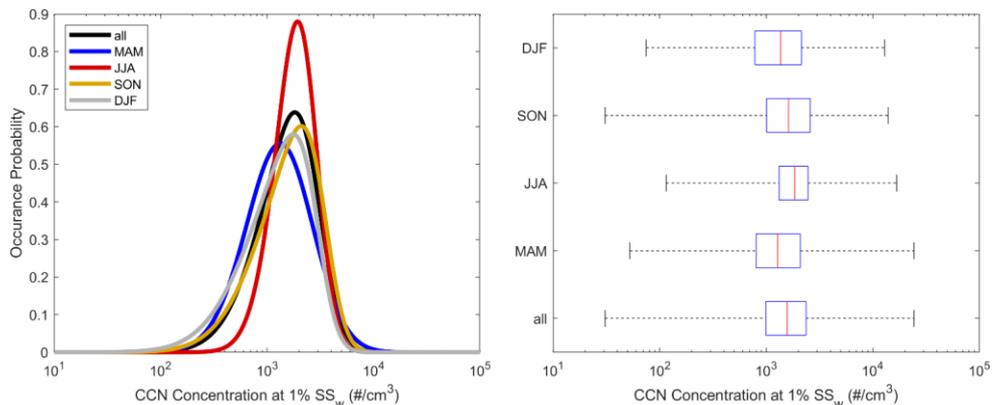
Figure 8: Distributions of CCN concentrations at 1% SS_w separated seasonally using skewed log-normal fits (left) and box plots (right).

3.2 Time Evolution of Clusters

Our large data set allows for additional statistical analysis to examine evolution of CCN spectra over time. This gives insight into underlying processes that are obscured when examining single cases or shorter data periods. Figure 9 shows the evolution of cluster classifications over time, examining all clusters starting in a given classification (nucleation, accumulation or intermediate). Cluster classification changes for all three clusters on the timescale of hours. Nucleation clusters are most likely to transition to intermediate clusters, rather than going directly to accumulation clusters. This could be through any or all of: growth of nucleation mode aerosol into larger sizes, coagulation scavenging/deposition/evaporation induced loss of the nucleation mode, and changes in air mass. Similarly, accumulation mode clusters are also more readily transitioned to intermediate clusters than to nucleation clusters. The role of intermediate clusters as the pathway of conversion between accumulation and nucleation clusters is further reinforced by the fact that they are equally likely to transition to either cluster type. In terms outside of the clustering perspective, it appears most likely that transitions from aerosol distributions dominated

17

by nucleation mode particles to ones dominated by accumulation mode particles (or vice versa) occur smoothly through intermediate cases where both modes are of similar magnitude, rather than abruptly. However, the analysis cannot distinguish the specific role of meteorology in these transitions.
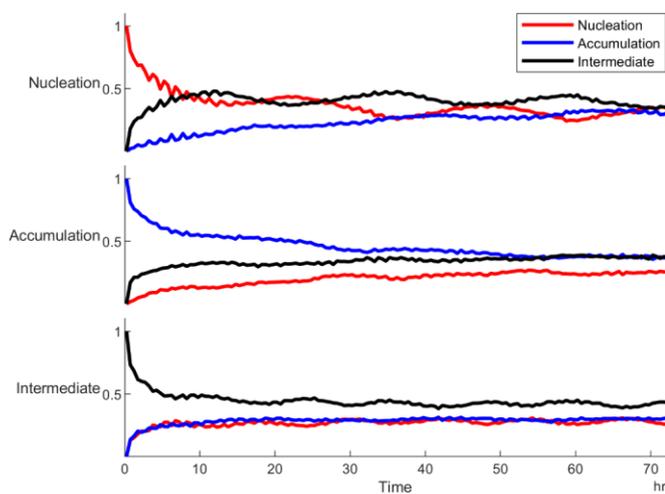

Figure 9: Cluster evolution over time (hours after appearance of cluster). Y axis for each plot shows the likelihood of the cluster transitioning to a new cluster type after a specified time, X axis, has elapsed. The first collection of traces labelled "Nucleation" shows evolution of nucleation cluster spectra into the other categories, with the middle collection depicting evolution of accumulation cluster initial states and the bottom showing evolution of intermediate cluster initial states. Traces indicate final states (after the specified time lag) of the nucleation cluster (red), accumulation cluster (blue) or intermediate cluster (black).

There is some periodicity observed in the cluster evolution, which we examine more closely alongside fluctuations in CCN number concentrations using autocorrelations. Autocorrelation coefficients are calculated for several different SS_w conditions, shown in Figure 10, using the methods described in Appendix D. Autocorrelation coefficients can be interpreted similarly to other correlation coefficients – they describe that portion of the variance that can be explained by the observation at a previous timepoint. Because of the differential nature of these comparisons, uncertainties in CCN distribution discussed in Section 3.1 are unlikely to propagate into this analysis. Further, the regions of highest uncertainty are avoided here. The higher the value of an autocorrelation coefficient, the more stable that quantity is over time, so that an autocorrelation

18

315 coefficient of 1 implies no change in state at a specified time lag, while 0 implies that a previous data point (separated by the specified time lag) has no influence on a current one. From Figure 10, a great deal of variability is observed in the first several hours of the computed time lags, which is an unexpected finding. Autocorrelation coefficients are expected to be highest for the first several timepoints, but the oscillating nature of these points implies aerosol processes with some periodicity in the 2-3 hour range. Natural processes that might produce such variability throughout the day and over all seasons seem unlikely, so

320 the oscillation may be an artefact, for example, introduced by sampling schedules. We apply a bi-exponential fit as an approximate way to smooth the data for time lags of up to twelve hours, removing the effect of these oscillations. Single exponential fits are poor approximations of the shape of the autocorrelation functions for 0.1 and 1% SS_w cases. Because a single decay pathway is expected to produce relatively consistent decay rates, the appearance of multi-exponential decays suggests multiple decay pathways. In this case decay pathways for autocorrelation can be interpreted as pathways for changes
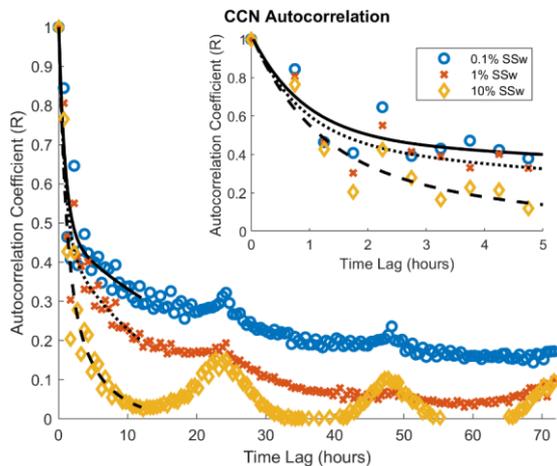
325 in CCN number concentrations.

Figure 10: Autocorrelation functions for CCN number concentrations at variable SS_w. Bi-exponential fits are shown for the first 12 hours of lag time for each SS_w, with solid, dotted, and dashed lines corresponding with 0.1%, 1%, and 10% SS_w, respectively.

Autocorrelation decays much more quickly for larger (10%) SS_w, with greatly increased values appearing at 24-hour

330 intervals. At higher SS_w, the CCN number concentration is often dominated by the smallest particles, associated with the nucleation mode. This interpretation fits well with the autocorrelation data, which indicate short-lived events tied to diurnal cycles. At moderate (1%) and low (0.1%) SS_w, there is a pronounced fast initial decay in autocorrelation, followed by a period of slower decay. Fit constants are described in appendix D and shown in Table D1. The fast initial decay rate is comparable

19

(within the large uncertainties, Table D1) for low and moderate SS_w cases, but the slow decay rate is significantly slower for the low SS_w case. The diurnal peak (at 24-hour time lag) is also significantly weaker for the lowest SS_w case, suggesting that there is less variability in these lowest SS_w CCN observations, as compared to higher SS_w. This is consistent with that fact that number concentrations at larger particle sizes are much less variable than those at smaller particle sizes (Marinescu et al., 2019). Previous work on AOD spatial and temporal autocorrelation at the SGP site (Alexandrov et al., 2004) and elsewhere (Anderson et al., 2003) suggest that the fast decay can be attributed to 3D micro-scale turbulent fluctuations, while the slow decay is due to 2D large-scale turbulence. The aerosol data used here are obtained only at the surface, to the influence of the three-dimensional nature of the atmosphere may be present but cannot be distinguished. Given the role of NPF events, there is also potentially a chemical (non-turbulence-driven) source of variability for CCN, contributing to autocorrelation decay for CCN concentrations at high SS_w. Ultimately these data illustrate that CCN spectra change rapidly over 1-3 hour timescales, with some conservation at longer timescales for the lower end of the supersaturation range. The granularity of our data (in ~45-minute increments) makes it somewhat difficult to resolve the exact timescales, but it is clear that the period of rapid change is in the 1-3 hour range. The role of variability in CCN concentration is something that should be considered in modelling studies that focus on the impacts of aerosol, especially those that use fixed concentrations of aerosol particles or those that do not capture the comprehensive processes that cause aerosol concentrations and properties to evolve. For example, using fixed CCN concentrations for a given short-term (<2 hour) simulation of shallow clouds (i.e., lower supersaturations) is more justifiable than for a longer-term simulation of the development of deep convective clouds (i.e., higher supersaturations), based on the faster autocorrelation decay rates of CCN at higher supersaturations. The autocorrelation results can also help define the time scales for data assimilation to ensure models are updated frequently enough to allow for accurate simulations.

4 Conclusions

We have developed, described and examined a long term CCN spectrum data product for the SGP site in Oklahoma. The data product builds on merged ~~number size distribution~~size distributions (Marinescu et al., 2019) and hygroscopicity measurements (Mahish and Collins, 2017) to create a best estimate of CCN spectra across a wide supersaturation range from ~0.0001% to 30%. It has been generated and verified by combining high quality data from ~~eight~~ seven different instruments. It has ~45 minute temporal resolution across five years of data, from 2009 to 2013, which has allowed for analyses not normally possible for smaller data sets.

We have determined that skewed log-normal distributions provide excellent fits to occurrence probabilities of CCN concentrations at any given supersaturation, as well as to occurrence probabilities of a wide range of other aerosol quantities. These types of distributions have been observed for AOD measurements previously, but not widely used. They provide more information than traditional box plots, while requiring a comparable number of parameters. For established occurrence

20

distribution shapes, shorter timescale measurements could likely take advantage of these fit parameters to fill in data gaps to
365 estimate data over longer periods. They also serve as useful inputs to models that include the expected variability in input parameters in model predictions.

CCN spectra are controlled primarily by particle ~~number size distribution~~size distributions, especially at larger SS_w values (above ~0.2%). In this high SS_w region of the spectrum, it appears possible to generate estimated CCN spectra using only median κ values, rather than concurrent measurements of κ and size distribution. However, this estimation relies on accurate
370 size distribution data that extend beyond 500 nm. Approximations of uncertainties introduced by this "median κ" estimation have been investigated and should hold for data from the SGP site during different time periods. This estimation method is also likely applicable for other sites, especially in similar environments, but is beyond the scope of this analysis.

Clustering analysis also highlights ~~number size distribution~~size distributions as the driving force behind changes in CCN spectra. Three distinct clusters are found for cases dominated by nucleation mode particles, accumulation mode particles, or
375 similar amounts of each. These are analyzed seasonally and hourly, finding all clusters in significant quantities across all seasons and times. Intermediate clusters are more likely during summer months, while accumulation clusters are abundant in the winter. Fall and spring appear similar in this view, falling between summer and winter. Nucleation mode clusters are most likely during daylight hours, corresponding with decreased intermediate clusters but nearly invariant accumulation clusters.

Time evolutions are examined in this data set to try to understand the dynamics of CCN spectra. Analysis of transitions
380 between clusters reveals that the most likely path is for nucleation and accumulation mode clusters to transition to an intermediate cluster first, rather than direct transitions occurring between the two. Autocorrelation analyses probe the evolution of a given CCN SS_w bin over time. A relatively quick decay is found for all SS_w values, with the bulk of correlation decaying in several hours, indicating that relatively large changes in CCN spectra can be expected over that time period. An additional slow decay is observed for smaller SS_w values, indicating that the CCN number is better conserved at longer timescales (>2-3
385 hours) in lower SS_w regions of the CCN spectrum, corresponding to particles in the coarse mode.

Appendix A: Merging of Distributions and CCN Spectra

The initial ~~number size distribution~~size distributions used in this work are generated from a combination of scanning mobility particle sizer (SMPS), aerodynamic particle sizer (APS), and condensation particle counter (CPC) data, as described previously by Marinescu et al., (2019), which are available in the DOE ARM archive (Marinescu and Levin, 2019). This initial
390 dataset, here referred to as size distribution (SD) 1, is processed to take into account additional instrument data utilizing humidified tandem differential mobility analyzer (HTDMA), CCN counter (CCNC), nephelometer, and aerosol chemical speciation mass spectrometer (ACSM) instrument data as outlined in Figure A1. SD1 data is available in approximately ~45
21

minute time intervals, where additional instrument data are available in higher time resolution, and are subsequently averaged over the time period of SD1 data for comparison. In order to compare to CCNC measurements, SD1 must be converted to CCN spectra using the hygroscopicity parameter (κ) that is derived from HTDMA measurements.
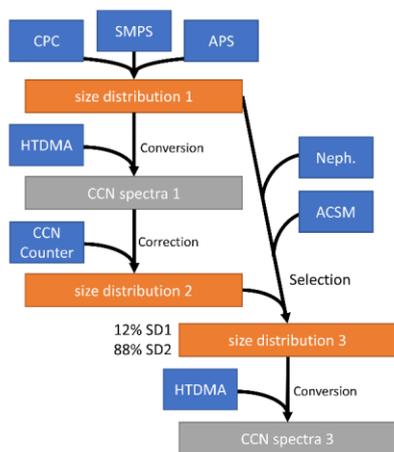

Figure A1: Flowchart illustrating processing of data to various distributions, with input instrument data shown in blue boxes, ~~number size distribution~~size distributions in orange, and CCN spectra in grey.

Size resolved aerosol hygroscopicity was measured with a HTDMA (Collins, 2010b), which first selected dried, monodisperse aerosol at seven diameters and then exposed them to a humidified (relative humidity, RH ~ 90%) growth region. The humidified aerosol ~~number size distribution~~size distribution was then measured and the change in particle diameter between the selected dry particle diameter (D_p) and the resulting humidified size (D_w) is termed the growth factor (GF = D_w/D_p). An example of HTDMA measured growth factor distributions is shown in Figure A2. The orange lines indicate the selected dry size for each GF measurement. The top part of the figure shows the ~~number size distribution~~size distribution measured by the SMPS at the same time.
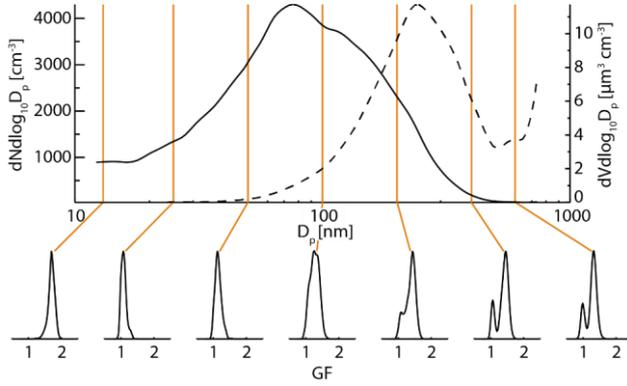
Figure A2: Example aerosol number distribution (solid line, upper plot) and growth factor (at 90% RH) distributions (lower plots) measured by an HTDMA. The dashed line in the upper plot is the corresponding volume distribution.

These GF data at a given RH (properly written e.g. GF(90%), but here abbreviated to GF for convenience) can be used to calculate the hygroscopicity of the particles as expressed via the hygroscopicity parameter, κ (Carrico et al., 2010; Petters and Kreidenweis, 2007):

$$\kappa = \frac{(GF^3 - 1)(1 - a_w)}{a_w}$$

Eq. 25

where

$$a_w = \frac{RH}{exp\left(\frac{4\sigma_w M_w}{\rho_w RT D_d GF(RH)}\right)}$$

Eq. 36

and σ_w, M_w and ρ_w are the surface tension, molecular weight, and density of water, respectively, T is the absolute temperature, and R is the ideal gas constant. After calculating κ distributions from each measured GF distribution at the diameters selected by the HTDMA we averaged κ for each selected D_d and linearly interpolated between selected sizes to generate a continuous distribution of aerosol hygroscopicity across the entire ~~number size distribution~~size distribution. Given the uncertainties introduced by interpolation and extension of kappa data beyond measurement bounds, we believe any added uncertainty introduced by using average kappa rather than a kappa distribution is relatively minor.

23

CCN spectra were generated using these κ values, derived from the HTDMA growth factor data described above using
either the SD1 (initial) or SD3 (final) size distributions (described below). For each time period with concurrent κ and size
distribution data, critical SS_w was calculated for each size bin. This was accomplished using Eq. 1 assuming the following
constant conditions: temperature of 25 °C, water density of 1 g/mL, and surface tension of 72 mJ/m^2. This calculation is
accomplished numerically, by calculating water SS_w for a logarithmically-spaced array of wet diameters, with the largest SS_w
chosen as the critical value. Calculated errors from this method were less than 1% of the calculated values using this method
(e.g. a 1% error in a 0.01% SS_w being ±0.0001% SS_w). A CCN spectrum was then generated by adding up activated particle
populations at each SS_w value, making no assumptions about the order particles activated in; smaller particles with higher κ
could activate before large particles with low κ if appropriate.

In order to compare spectra between CCN spectrum 1 and measured CCNC values averaged across the same ~45-minute
time intervals, interpolation of the CCNC data across the calculated SS_w bins was performed using MATLAB's built in
Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) function for each CCNC spectrum. These data were used to create
a distribution corrected for the CCNC data using a similar method to that described for the CPC corrections in SD1 (Marinescu
et al., 2019). The only difference between the algorithm described in Appendix A of Marinescu et al. (2019) and the one used
for generation of SD2 here occurs in step 3. We calculated a two-week rolling median percent difference between the CPC
and SMPS + APS distribution and used this as a scaling factor across the entire distribution in this step. Times between 1200
and 1800 local are excluded from the rolling median since new particle formation events are common during those times and
large differences between the SMPS+APS integrated number concentration and the CPC number concentration are expected.
To generate SD2, the average difference between CPC and CCNC (in total particles and SS_w specific CCN numbers,
respectively) is used instead of solely CPC data. For example, if the comparison with the CPC suggested there should be 25%
more particles in the SMPS+APS ~~number size distribution~~size distributions and the comparison with the CCNC suggested
there should be 15% fewer particles in the SMPS+APS ~~number size distribution~~size distributions, then the SMPS+APS ~~number
size distribution~~size distribution data are scaled up by 5% (the average of +25% and -15%). The remaining steps described in
(Marinescu et al., 2019) are performed unchanged on this distribution. The remaining steps in the algorithm can change the
shape of the size distribution, so SD1 and SD2 are not simply scaled versions of each other. If no quality CCNC data are
available for a given timepoint, the SD1 and SD2 spectra are identical.

The resulting SD2 was then compared to ACSM and nephelometer data to examine whether the CCNC correction was
warranted. ACSM comparisons were accomplished by generating total particle mass concentration for each distribution
assuming spherical particles and a density of 1.77 mg/mL, which is that of ammonium sulfate. Additionally, the ACSM cut-
off of 1~~0~~ μm and the volume equivalent diameter (DeCarlo et al., 2004) were accounted for to produce calculated aerosol mass
for comparison. The density chosen is within the region of best agreement between the ACSM and distribution data for both

24

450 the SD1 and SD2 and is chosen for consistency with the nephelometer comparison. The nephelometer comparison was accomplished by generating single-particle scattering cross-sections for all size bins in the distributions assuming the optical properties of ammonium sulfate, which was again within the region of best agreement in Figure A3. Both of these comparisons produce excellent agreement for many time points, as shown in Figure A4. There was evidence of systematic bias for some time periods but the bias was relatively low for periods outside of 10-Mar-2011 through 1-Nov-2011 for the nephelometer

455 data, which were not used in distribution selection, below.
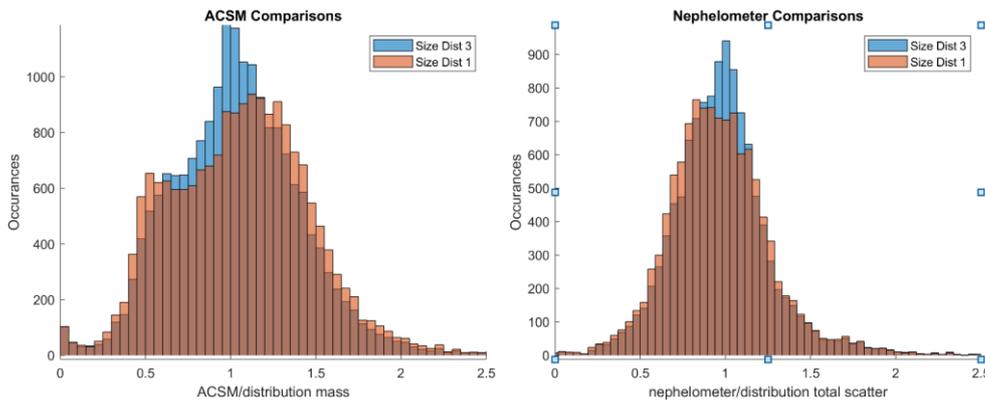
Figure A3: Statistical comparisons of the agreement between measured quantities (left, ACSM aerosol mass concentrations; right, PM10 nephelometer scattering coefficients) and those estimated from either the SD1 or SD3 ~~number size distribution~~size distributions. The x axes show the comparisons as ratios, where a ratio of 1 indicates perfect agreement.

460   SD2 generally produced better agreement with the nephelometer and ACSM than SD1 did, although this was not true for all time periods. In order to construct a final data set including the nephelometer and ACSM comparisons, the ratios between distribution-calculated values and measured values were used to select between SD1 and SD2 at each point where data were available. Given the better general agreement for SD2, it was used as default if there were no ACSM or nephelometer data available, or if there was disagreement between the two instruments. Through this process, 4711 distributions were selected

465 from SD1, and 16203 from SD2, with 19407 points defaulting to SD2. The resulting distribution, SD3 compared with ACSM and nephelometer measurements in Figure A3, is considered the final product distribution, and analyzed throughout the manuscript alongside the CCN spectra generated from it.
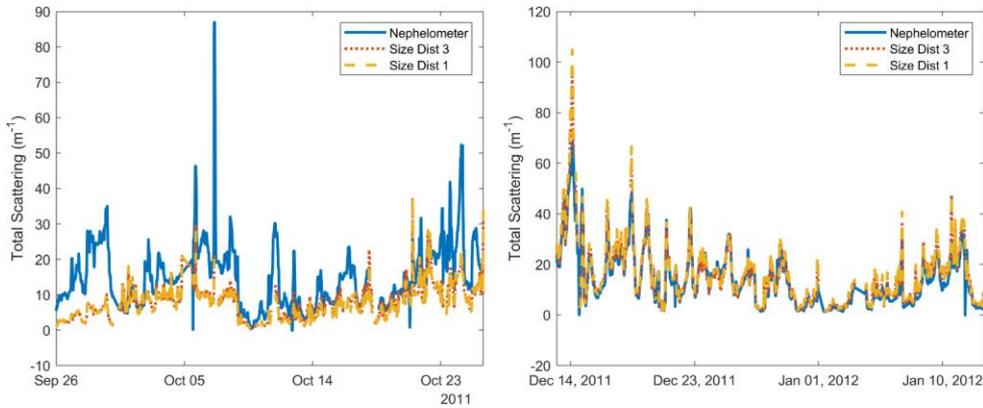
25

Figure A4: Examples of excluded (left) and included (right) nephelometer data (blue traces) compared to total scattering coefficients calculated from the indicated estimated ~~number size distribution~~size distributions (SD1 or SD3).

Appendix B: Skewed Log-Normal Fits

The log-normal probability density function is defined as:

$$\varphi(\ln x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

Eq. ~~4~~7

where x is a number concentration (bin), μ is the median value of log(x), and σ^2 is the variance in log(x). The log-normal cumulative distribution function is defined as:

$$\Phi(\ln x) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left[\frac{\ln x - \mu}{\sigma\sqrt{2}}\right]$$

Eq. ~~5~~8

where erf is the error function. A skewed log-normal probability density function can subsequently be defined as:

$$f(\ln x) = 2\varphi(\ln x)\Phi(\alpha \ln x)$$

Eq. ~~6~~9

where α is a parameter representing the degree of skew, such that when α = 0, the log-normal distribution is recovered. When these functions are used to fit data, μ, σ, and α are used as the fit parameters.

26

Two issues arise using skewed log-normal fits. The first is that there is no closed-form expression for the median value of a skewed log-normal distribution. The median can of course still be evaluated numerically. The second is that x values of zero cannot be represented within the distribution due to the logarithm. In this work, we address these issues by simply reporting median values and fractions of the data where zeros occur, alongside fit parameters. It should be noted μ is no longer the median value of log(x) for the skewed log-normal distribution. This can be observed in the data in Supporting File CCN_fit_coeffs.txt where μ decreases at high SS_w while median value increases monotonically, as expected for a cumulative distribution. Fits are generally very good, as shown in Figure B1 for several different aerosol quantities. At very low SS_w, or for very large particle bins, the quality of the fits degrades due to the large amount of noise in the data. This noise occurs largely due to detection limits of the instruments involved at low particle concentrations (for very large particles). If concentrations are so low that particle detections are not guaranteed in the sampling period (45 minutes), a large amount of shot noise is introduced.
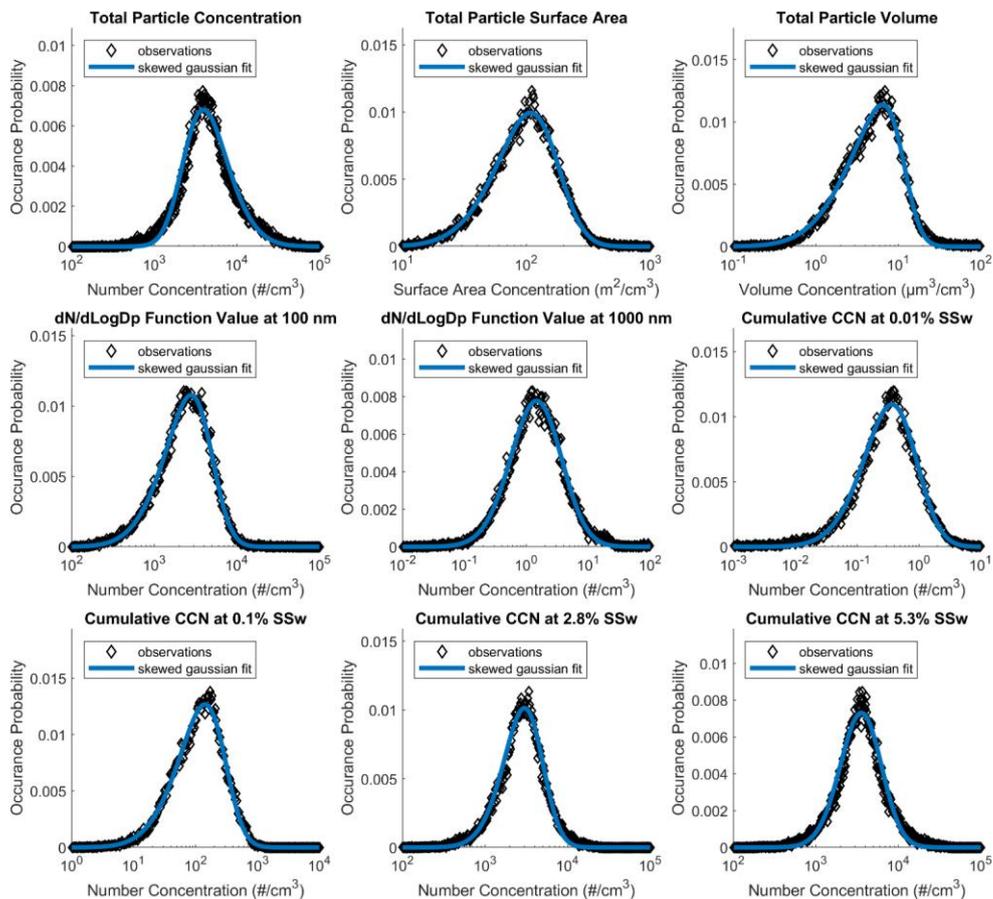
Figure B1: Examples of skewed log-normal fits to occurrence probabilities for aerosol metrics, as labelled: total particle number, total particle volume, dN/dLogDp function at 100 nm, dN/dLogDp function at 1000 nm, particle number concentrations active as CCN at 0.01%, 0.1%, 2.8%, and 5.3% SS_w.

28

Appendix C: Clustering Analysis

Clustering was done after evaluating several different methods and options. The primary parameters that were varied were the number of clusters used, and the distance metric used to distinguish clusters. All analysis was performed using built-in Matlab functions. Three distance metrics were evaluated using the "kmeans" function options: 1) squared Euclidian, 2) sum of absolute differences (cityblock), and 3) cosine. Each defines cluster distance, d, as follows:

$$d_1(x,c) = (x-c)(x-c)'$$ Eq. 7~~10~~

$$d_2(x,c) = \sum_j |x_j - c_j|$$ Eq. 8~~11~~

$$d_3(x,c) = 1 - \frac{xc'}{\sqrt{(xx')(cc')}}$$ Eq. 9~~12~~

where x is input data (CCN spectrum for a given timepoint), and c is a cluster centroid. Both x and c are arrays, with subscripts indicating a single array element and apostrophes indicating a transpose operation. It was found that metrics 1 and 2 produced separation of spectra based solely on total particle number, whether clustering was applied to aerosol ~~number size distribution~~size distributions or to CCN spectra. Distance metric 3, however, produced well-resolved clusters based on distribution shape (how the CCN spectrum changes with SS_w), and was selected for final cluster designations. Mathematically, distance metric 3 is a measure of the included angle between points treated as vectors, which provides some effective normalization, so the result of clusters based on distribution shape rather than total aerosol number is not surprising. Clustering in CCN space also produced well resolved clusters in size distribution space, and vice versa. CCN space was ultimately chosen for clustering, due to the focus of the current work, but differences from the alternative are expected to be very minor.

Next, optimal number of clusters was explored. While this is often accomplished somewhat arbitrarily, intuitively, or based on external models for a given process, statistical methods have been developed to guide the process. We chose to use the "gap statistic" (Tibshirani et al., 2001), a built in Matlab functionality through the "evalclusters" function. This method was too computationally intensive to use on the entire data set, so a subset of 500 randomly selected spectra were used instead. Three clusters were suggested to be the optimal number to use, based on this approach. These clusters all appeared physically distinct, as discussed in Section 3.1.2, and addition of a 4th cluster simply resulted in the splitting of two adjacent clusters. Three clusters were thus chosen for use in further analysis.

Clusters are generally similar year round, but there is some seasonal dependence within a given cluster, as shown in Figure C1.
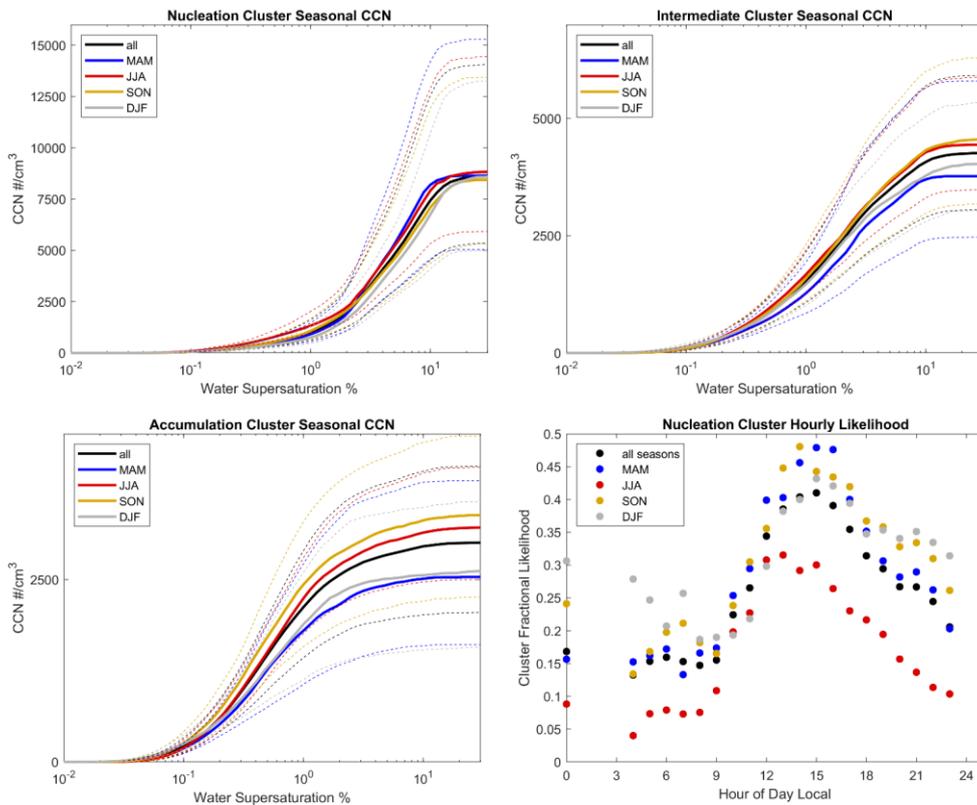
29
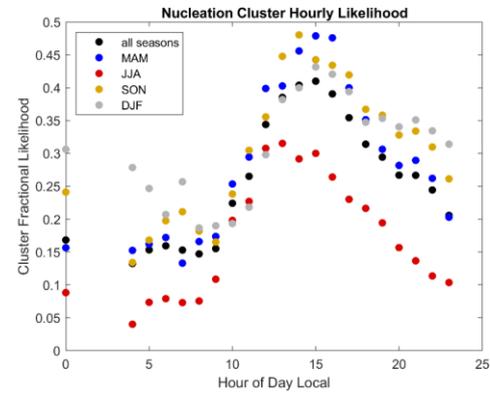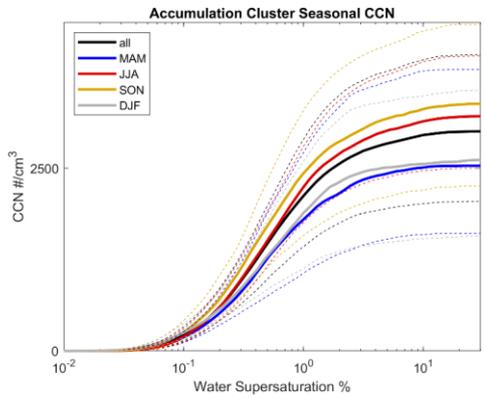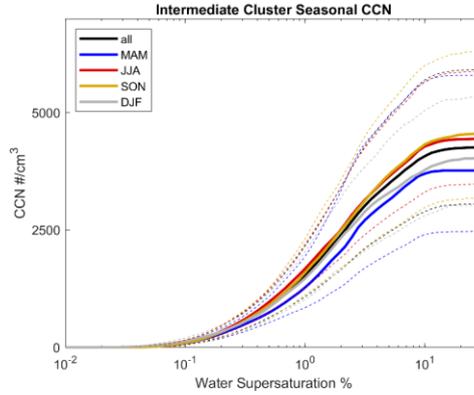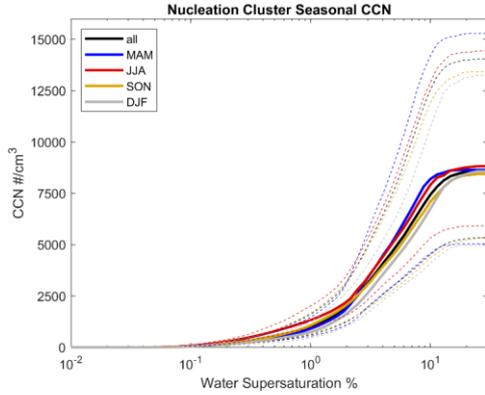
Figure C1: Median (solid lines) and confidence intervals containing 75% of the data (dotted lines) for each cluster as a function of season, along with seasonal variations in hourly cluster likelihood for the nucleation mode cluster (bottom right).

520 **Appendix D: Non-Periodic Autocorrelation and Fits**

Autocorrelation coefficients (Box et al., 2015) were generated by comparing adjacent points in a timeseries to determine the portion of variance that can be explained by the adjacent points. For a timeseries with equally spaced measurements, the autocorrelation function is defined as:

31

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}$$

Eq. 10~~13~~

where r_k is the autocorrelation coefficient at time lag k, N is the total number of timepoints, Y_i is the measurement value (in our case CCN number concentration) at timepoint i, and $\bar{Y}$ is the mean measurement value. For timepoints that are not evenly spaced, the same coefficient can be produced with a few extra steps. The way we have accomplished this is to 1) calculate differences between all adjacent timepoints, for a fixed number of integer time lags, before 2) sorting all of these data into time bins based on how much time elapsed between any given set of measurements.

Autocorrelation functions were subsequently fit to bi-exponential decays for the first 12 hours of time lag data, using the following form:

$$f(x) = c_1 e^{-\tau_1 x} + c_2 e^{-\tau_2 x}$$

Eq. 11~~14~~

This is accomplished using the "fit" function in Matlab, which provides 95% confidence interval information. Best fit parameters and 95% confidence intervals are reported in Table D1 below.

Table D1: Bi-exponential Fit Coefficients

	0.1% SS_w	1% SS_w	10% SS_w
c_1	0.48 (0.34, 0.61)	0.46 (0.31, 0.61)	0.39 (-0.09, 0.88)
τ_1	0.036 (0.001, 0.071)	0.070 (0.027, 0.113)	0.215 (0.026, 0.404)
c_2	0.55 (0.37, 0.72)	0.57 (0.38, 0.75)	0.63 (0.15, 1.11)
τ_2	1.11 (0.35, 1.86)	1.18 (0.39, 1.96)	0.99 (0.08, 1.90)

Data Availability

All data are publicly available via the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) user facility data archive, including instrument data (Salwen et al., 1990, https://doi.org/10.5439/1025152; Hageman et al., 1996, https://doi.org/10.5439/1025259; Collins, 2005, https://doi.org/10.5439/1025303, 2010a, https://doi.org/10.5439/1150275; Koontz et al., 2012, https://doi.org/10.5439/1228051; Zawadowicz and Howie, 2021, https://doi.org/10.5439/1763029) ~~(Atmospheric Radiation Measurement (ARM), 1995, 2001, 2007, 2010, 2011, 2015) (https://doi.org/10.5439/1025220, https://doi.org/10.5439/1095386, https://doi.org/10.5439/1025259, https://doi.org/10.5439/1150275,~~

32

~~https://doi.org/10.5439/1025152, and https://doi.org/10.5439/1025303), including~~the initial (Marinescu and Levin, 2019~~,~~ https://doi.org/10.5439/1511037) merged aerosol size distributions used here, ~~the~~and the final~~(Perkins, 2009)~~ merged aerosol ~~number~~ size distribution and CCN data ~~((Perkins, 2009,~~ https://doi.org/10.5439/1832908).

Supplement link

A supplemental text document containing skewed log-normal fit coefficients for all CCN data, named CCN_fit_coeffs.txt, is available at <mark>XXX.</mark>

Author Contribution

RJP performed the analyses presented. All authors assisted in interpretation of raw instrument data and construction of the merged CCN product. RJP and SMK prepared the manuscript with contributions from all co-authors.

Competing Interests

The authors declare that they have no conflicts of interest.

Acknowledgments

All data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. We would also like to acknowledge Jeffrey Pierce for helpful discussions about new particle formation events.

Financial Support

This research has been supported by the U.S. Department of Energy's Atmospheric System Research, and Office of Science, Office of Biological and Environmental Research program (grant no. DESC0016051).

References

Alexandrov, M. D., Marshak, A., Cairns, B., Lacis, A. A., and Carlson, B. E.: Scaling Properties of Aerosol Optical Thickness Retrieved from Ground-Based Measurements, J. Atmospheric Sci., 61, 1024–1039, https://doi.org/10.1175/1520-0469(2004)061<1024:SPOAOT>2.0.CO;2, 2004.

33

Alexandrov, M. D., Geogdzhayev, I. V., Tsigaridis, K., Marshak, A., Levy, R., and Cairns, B.: New Statistical Model for
565 Variability of Aerosol Optical Thickness: Theory and Application to MODIS Data over Ocean, J. Atmospheric Sci., 73, 821–
837, https://doi.org/10.1175/JAS-D-15-0130.1, 2016.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale Variations of Tropospheric Aerosols,
J. Atmospheric Sci., 60, 119–136, https://doi.org/10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2, 2003.

Bianchi, F., Tröstl, J., Junninen, H., Frege, C., Henne, S., Hoyle, C. R., Molteni, U., Herrmann, E., Adamov, A., Bukowiecki,
570 N., Chen, X., Duplissy, J., Gysel, M., Hutterli, M., Kangasluoma, J., Kontkanen, J., Kürten, A., Manninen, H. E., Münch, S.,
Peräkylä, O., Petäjä, T., Rondo, L., Williamson, C., Weingartner, E., Curtius, J., Worsnop, D. R., Kulmala, M., Dommen, J.,
and Baltensperger, U.: New particle formation in the free troposphere: A question of chemistry and timing, Science, 2016.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M.: Time Series Analysis: Forecasting and Control, 5th Edition.,
John Wiley and Sons Inc., Hoboken, New Jersey, 712 pp., 2015.

575 Carrico, C. M., Petters, M. D., Kreidenweis, S. M., Sullivan, A. P., McMeeking, G. R., Levin, E. J. T., Engling, G., Malm, W.
C., and Collett, J. L. J.: Water uptake and chemical composition of fresh aerosols generated in open burning of biomass,
Atmospheric Chem. Phys., 10, 5165–5178, https://doi.org/10.5194/acp-10-5165-2010, 2010.

Cheng, W. Y. Y., Carrió, G. G., Cotton, W. R., and Saleeby, S. M.: Influence of cloud condensation and giant cloud
condensation nuclei on the development of precipitating trade wind cumuli in a large eddy simulation, J. Geophys. Res.
580 Atmospheres, 114, https://doi.org/10.1029/2008JD011011, 2009.

Cohard, J.-M., Pinty, J.-P., and Bedos, C.: Extending Twomey's Analytical Estimate of Nucleated Cloud Droplet
Concentrations from CCN Spectra, J. Atmospheric Sci., 55, 3348–3357, https://doi.org/10.1175/1520-
0469(1998)055<3348:ETSAEO>2.0.CO;2, 1998.

Collins, D.: ARM: Tandem Differential Mobility Analyzer: size-resolved concentrations, Atmospheric Radiation
585 Measurement (ARM) Archive, https://doi.org/10.5439/1025303, 2005.

Collins, D.: ARM: Tandem Differential Mobility Analyzer Aerosol Particle Sizer, Atmospheric Radiation Measurement
(ARM) Archive, https://doi.org/10.5439/1150275, 2010a.

Collins, D.: Tandem Differential Mobility Analyzer/Aerodynamic Particle Sizer (APS) Handbook, PNNL; Richland, WA,
https://doi.org/10.2172/982072, 2010b.

590 DeCarlo, P. F., Slowik, J. G., Worsnop, D. R., Davidovits, P., and Jimenez, J. L.: Particle Morphology and Density
Characterization by Combined Mobility and Aerodynamic Diameter Measurements. Part 1: Theory, Aerosol Sci. Technol.,
38, 1185–1205, https://doi.org/10.1080/027868290903907, 2004.

Feingold, G., Cotton, W. R., Kreidenweis, S. M., and Davis, J. T.: The Impact of Giant Cloud Condensation Nuclei on Drizzle
Formation in Stratocumulus: Implications for Cloud Radiative Properties, J. Atmospheric Sci., 56, 4100–4117,
595 https://doi.org/10.1175/1520-0469(1999)056<4100:TIOGCC>2.0.CO;2, 1999.

Gantt, B., He, J., Zhang, X., Zhang, Y., and Nenes, A.: Incorporation of advanced aerosol activation treatments into
CESM/CAM5: model evaluation and impacts on aerosol indirect effects, Atmospheric Chem. Phys., 14, 7485–7497,
https://doi.org/10.5194/acp-14-7485-2014, 2014.

34

Gerber, H.: Supersaturation and Droplet Spectral Evolution in Fog, J. Atmospheric Sci., 48, 2569–2588, https://doi.org/10.1175/1520-0469(1991)048<2569:SADSEI>2.0.CO;2, 1991.

Glenn, I. B., Feingold, G., Gristey, J. J., and Yamaguchi, T.: Quantification of the Radiative Effect of Aerosol–Cloud Interactions in Shallow Continental Cumulus Clouds, J. Atmospheric Sci., 77, 2905–2920, https://doi.org/10.1175/JAS-D-19-0269.1, 2020.

Hageman, D., Behrens, B., Smith, S., Uin, J., Salwen, C., Koontz, A., Jefferson, A., Watson, T., Sedlacek, A., Kuang, C., Dubey, M., Springston, S., and Senum, G.: ARM: Aerosol Observing System (AOS): aerosol data, 1-min, mentor-QC applied, Atmospheric Radiation Measurement (ARM) Archive, https://doi.org/10.5439/1025259, 1996.

Hodshire, A. L., Lawler, M. J., Zhao, J., Ortega, J., Jen, C., Yli-Juuti, T., Brewer, J. F., Kodros, J. K., Barsanti, K. C., Hanson, D. R., McMurry, P. H., Smith, J. N., and Pierce, J. R.: Multiple new-particle growth pathways observed at the US DOE Southern Great Plains field site, Atmospheric Chem. Phys., 16, 9321–9348, https://doi.org/10.5194/acp-16-9321-2016, 2016.

Hudson, J. G., Jha, V., and Noble, S.: Drizzle correlations with giant nuclei, Geophys. Res. Lett., 38, https://doi.org/10.1029/2010GL046207, 2011.

Johnson, D. B.: The Role of Giant and Ultragiant Aerosol Particles in Warm Rain Initiation, J. Atmospheric Sci., 39, 448–460, https://doi.org/10.1175/1520-0469(1982)039<0448:TROGAU>2.0.CO;2, 1982.

Jung, E., Albrecht, B. A., Jonsson, H. H., Chen, Y.-C., Seinfeld, J. H., Sorooshian, A., Metcalf, A. R., Song, S., Fang, M., and Russell, L. M.: Precipitation effects of giant cloud condensation nuclei artificially introduced into stratocumulus clouds, Atmospheric Chem. Phys., 15, 5645–5658, https://doi.org/10.5194/acp-15-5645-2015, 2015.

Koontz, A., Flynn, C., Uin, J., and Jefferson, A.: AOS humidified nephelometer, harmonized, Atmospheric Radiation Measurement (ARM) Archive, https://doi.org/10.5439/1228051, 2012.

Levin, Z. and Cotton, W. R. (Eds.): Aerosol Pollution Impact on Precipitation: A Scientific Review, Springer Netherlands, https://doi.org/10.1007/978-1-4020-8690-8, 2009.

Low, R. D. H.: Microphysical and meteorological measurements of fog supersaturation, Tellus, 27, 507–513, https://doi.org/10.3402/tellusa.v27i5.10177, 1975.

Mahish, M. and Collins, D.: Analysis of a Multi-Year Record of Size-Resolved Hygroscopicity Measurements from a Rural Site in the U.S., Aerosol Air Qual. Res., 17, 1489–1500, https://doi.org/10.4209/aaqr.2016.10.0443, 2017.

Marinescu, P. and Levin, E.: SGP Merged Aerosol Size Distribution (CPC+SMPS+APS), Atmospheric Radiation Measurement (ARM) Archive, United States, https://doi.org/10.5439/1511037, 2019.

Marinescu, P. J., Heever, S. C. van den, Saleeby, S. M., Kreidenweis, S. M., and DeMott, P. J.: The Microphysical Roles of Lower-Tropospheric versus Midtropospheric Aerosol Particles in Mature-Stage MCS Precipitation, J. Atmospheric Sci., 74, 3657–3678, https://doi.org/10.1175/JAS-D-16-0361.1, 2017.

Marinescu, P. J., Levin, E. J. T., Collins, D., Kreidenweis, S. M., and Heever, S. C. van den: Quantifying aerosol size distributions and their temporal variability in the Southern Great Plains, USA, Atmospheric Chem. Phys., 19, 11985–12006, https://doi.org/10.5194/acp-19-11985-2019, 2019.

Nieminen, T., Kerminen, V.-M., Petäjä, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Hõrrak, U., Kivekäs, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Németh, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmospheric Chem. Phys., 18, 14737–14756, https://doi.org/10.5194/acp-18-14737-2018, 2018.

Patel, P. N. and Jiang, J. H.: Cloud condensation nuclei characteristics at the Southern Great Plains site: role of particle size distribution and aerosol hygroscopicity, Environ. Res. Commun., 3, 075002, https://doi.org/10.1088/2515-7620/ac0e0b, 2021.

Perkins, R.: Southern Great Plains Merged and Extended Cloud Condensation Nuclei Data, Atmospheric Radiation Measurement (ARM) Archive, https://doi.org/10.5439/1832908, 2009.

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

Pierce, J. R., Westervelt, D. M., Atwood, S. A., Barnes, E. A., and Leaitch, W. R.: New-particle formation, growth and climate-relevant particle production in Egbert, Canada: analysis from 1 year of size-distribution observations, Atmospheric Chem. Phys., 14, 8647–8663, https://doi.org/10.5194/acp-14-8647-2014, 2014.

Pinsky, M., Khain, A., Mazin, I., and Korolev, A.: Analytical estimation of droplet concentration at cloud base, J. Geophys. Res. Atmospheres, 117, https://doi.org/10.1029/2012JD017753, 2012.

Posselt, R. and Lohmann, U.: Influence of Giant CCN on warm rain processes in the ECHAM5 GCM, Atmospheric Chem. Phys., 8, 3769–3788, https://doi.org/10.5194/acp-8-3769-2008, 2008.

Saleeby, S. M., van den Heever, S. C., Marinescu, P. J., Kreidenweis, S. M., and DeMott, P. J.: Aerosol effects on the anvil characteristics of mesoscale convective systems, J. Geophys. Res. Atmospheres, 121, 10,880-10,901, https://doi.org/10.1002/2016JD025082, 2016.

Salwen, C., Boyer, M., Springston, S., Kuang, C., and Andrews, E.: ARM: AOS: condensation particle counter, Atmospheric Radiation Measurement (ARM) Archive, https://doi.org/10.5439/1025152, 1990.

Sayer, A. M. and Knobelspiesse, K. D.: How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth, Atmospheric Chem. Phys., 19, 15023–15048, https://doi.org/10.5194/acp-19-15023-2019, 2019.

Shen, C., Zhao, C., Ma, N., Tao, J., Zhao, G., Yu, Y., and Kuang, Y.: Method to Estimate Water Vapor Supersaturation in the Ambient Activation Process Using Aerosol and Droplet Measurement Data, J. Geophys. Res. Atmospheres, 123, 10,606-10,619, https://doi.org/10.1029/2018JD028315, 2018.

Tibshirani, R., Walther, G., and Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, J. R. Stat. Soc. Ser. B Stat. Methodol., 63, 411–423, https://doi.org/10.1111/1467-9868.00293, 2001.

Uin, J.: Cloud Condensation Nuclei Particle Counter Instrument Handbook, DOE ARM Clim. Res. Facil., https://doi.org/DOE/SC-ARM-TR-168, 2016.

36

Venzac, H., Sellegri, K., Laj, P., Villani, P., Bonasoni, P., Marinoni, A., Cristofanelli, P., Calzolari, F., Fuzzi, S., Decesari, S., Facchini, M.-C., Vuillermoz, E., and Verza, G. P.: High frequency new particle formation in the Himalayas, Proc. Natl. Acad. Sci., 105, 15666–15671, https://doi.org/10.1073/pnas.0801355105, 2008.

670

Zawadowicz, M. and Howie, J.: Aerosol Chemical Speciation Monitor, mentor processed, .c2, Atmospheric Radiation Measurement (ARM) Archive, https://doi.org/10.5439/1763029, 2021.