Perkins et al. ACP Submission
https://doi.org/10.5194/acp-2021-1008
Long- and Short-Term Temporal Variability in Cloud Condensation Nuclei Spectra over a Wide Supersaturation Range at the Southern Great Plains Site

RESPONSES TO REVIEWS

Reviewer 1

This manuscript introduced a best-estimate CCN spectrum product across a wide range of supersaturation range from 0.0001 to 30%, using multiple independent instruments at ARM SGP site. Cluster analysis is used to identify the key driving element (aerosol number size distribution) in CCN spectra. The seasonal and diurnal cycles of the clusters are analyzed as well as the time evolution of these clusters. This dataset is useful in aerosol and aerosol-cloud interaction study. It also has potential to benefit modeling study in high-SS or low-SS conditions. Overall, the manuscript is well written and I recommend it be published in ACP, with some comments listed below.

Response: We thank the Reviewer for these remarks and for the comments that have helped improve this manuscript. Our point-by-point responses are below.

One aspect I think the authors can improve this manuscript is by adding more analysis and discussions on the uncertainty of the reconstructed CCN, especially in the high- and low-SS ends. Extending the CCN spectra into a wider range is a key point of this dataset. However, the authors also admit the data limits in the high and low ends of SS. This makes me wonder how good we can trust the data in the two ends. As the authors mentioned the uncertainty is included in the dataset, it would be helpful to add a plot of CCN uncertainty across all SS range and discuss how much we can trust the data.

Response: As the Reviewer notes, the data are highly uncertain at the two extremes. Unfortunately our biggest sources of uncertainty in the CCN distributions are due to uncertainties in the underlying size distributions, and methods of estimating uncertainties in composite aerosol size distributions created from multiple instruments are not well developed, making this very difficult to do quantitatively. As such, we have settled for a qualitative discussion of the uncertainties, and where we think they become most problematic.

It would also be great if some discussions can be made on what can we learn from CCN in these extremely high (low) SS conditions.

Response: This is a great point. We have added some additional discussion at the beginning of Section 3.1. In particular, uncertainties at low SS conditions could be reduced by more attention to the routine measurement of particles in the 500 nm – 20 micron size range, which we now point out explicitly. At both very high and very low supersaturations, particle composition and precise size are not as important as accurate counting, which we attempt to demonstrate explicitly. Text has been added to Section 3.1 which now reads:

"In the low SSw region, the kappa values of these large particles have significant impacts on critical SSw 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 SSw 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 SSw spectrum (greater than the critical SSw value), which can be quite significant at lower SSW values."

Minor comment:

Title: since the wide SS range is a key point of this dataset, I would recommend adding something in the title to indicate that, such as "… Cloud Condensation Nuclei Spectra in a Wide Supersaturation Range …"

Response: Thanks for the suggestion. We have modified the title to "Long- and Short-Term Temporal Variability in Cloud Condensation Nuclei Spectra over a Wide Supersaturation Range in the Southern Great Plains Site".

Line 15: I am curious if the authors have any plan to extend this data to more recent years. If yes, it will be good to discuss in the last of the manuscript.

Response: Although initially we planned to provide this algorithm for ongoing use at the site, the suite of instruments has since changed. So no, at this time, we do not plan to extend it to additional years.

Line 67: it would be good to list out (better within a table) the instrument names and size ranges used in building the aerosol size distribution and CCN spectrum.

Response: The aerosol size distribution merging procedure, along with the instrumentation used, is described in Marinescu et al. (2019) and we used their archived, merged size distribution to initialize our procedure (Figure A1). In this work, we have added the other measurements shown in Figure A1: HTDMA, CCNC, ACSM, and nephelometer. Those data are not used in a size-specific sense, except as described in the text. We have created a table of instruments and measurements and placed it within the methods section, and included a reference to it around line 67.

Figure 1: is this example of one 45-min data sample? It would be good to present the date/time of the example.

Response: Yes, this is one of the tabulated data entries. We have added the date and time as suggested.

Line 96-97: "… which can be used to check the accuracy of the reconstructed spectra." This sentence is misleading me that I thought this paragraph is to evaluate the data accuracy using independent datasets. I suggest deleting this sentence and changing the next one as "… that were used to constrain the reconstructed CCN spectra at the low-SS (and high-SS??) end"

Response: The Reviewer raises a concern that this language may be unintentionally ambiguous. We did indeed assess the reconstructed spectra against the direct CCN observations (at the available measured SS setpoints), and these are independent observations from size distribution and hygroscopicity measurements. The extent of agreement / disagreement is then used to decide whether corrections are needed to the reconstructed CCN spectra as shown in Figure A1, not limited to the low and high SS ends.

We have modified these sentences as follows: "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% 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."

Line 106: why six instruments here? In abstract and summary it says eight instruments. Also, it would be good to list all of them in a table.

Response: Thank you for pointing out this inconsistency. Indeed, data from five instruments were merged (CPC, SMPS, APS, HTDMA, CCNc) with data from two others – the ACSM, and nephelometer – used as constraints. We have corrected this language in the Abstract, Summary and main text (shown via tracked changes). Instrumentation is listed in Figure A1 along with the steps utilizing it to generate our data product.

Line 120-121: How is k calculated from GF distribution? Appendix A says k distribution is calculated from GF distribution then averaged into one single k value. I thought the GF distribution, either single-modal or bi-modal, can be captured in k calculation. Please correct me if my understanding is wrong.

Response: The Reviewer is correct. Figure 1 and Figure A2 show the distribution of measured k (from GFs) as a function of particle size. Because of the measurement strategy used in this HTDMA, we could also define a distribution within each particle size class shown in Figure A2. From that Figure, it can been seen that the multi-modal nature of k at each size is enhanced for larger particles, where number concentrations are low. After some exploration, we decided that the overall impact on CCN distributions by using an averaged k instead of the full k distribution within a size bin was small, and applied the average size-dependent k in our subsequent calculations, as described in the text. Added the following text to Appendix A: "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."

Line 130: "…, and these particles this lowest SSw region." Is this sentence incomplete?

Response: Yes, thank you! That clause should have been deleted and has been removed in the revision.

Line 133: again, it would be great to add analysis on the data uncertainty and discuss what value can be added with this wide SS range.

Response: Agreed. We have added discussion of uncertainties, as mentioned above, as well as discussion on the value of the data product at the end of section 3.1 which now reads: "Ultimately, the regions of increased uncertainty are under SSw 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 SSw regions, this work provides best-estimate values that can be used for analysis or modelling of strong updraft (high SSw) or precipitation initiation (low SSw) scenarios."

Line 142: can you explain the advantages of using skewed log-normal distribution other than regular log-normal? Any reason why CCN number concentration (and other variables in Figure B1) fit skewed log-normal distribution better?

Response: As demonstrated in Figures 2 and B1, we found that the skewed lognormal better represented the data, and especially the tails of the distributions. As for reasons for this, we can speculate e.g. that dN/dlogDp at 100 nm is skewed toward lower values due to the rural nature of the site, whereas e.g. total particle concentration has a higher tail at high number concentrations because periods of recent new particle formation are more frequent that periods of ultraclean air masses, such as might occur after rain. We expect therefore that the distributions reflect the meteorology as well as local sources of particles at the site and might be unique for different locations around the world.

Line 151-155: This point is interesting. Can we found high correlation within each aerosol mode and low correlation between bins in different modes (such as nucleation, Aitken and accumulation modes)? will it give evidence to choose size ranges for each mode?

Response: Yes, although beyond the scope of this work, it does seem possible to use statistical techniques to define size ranges for modes when continuous size distribution data are available. Thanks for an interesting suggestion.

Figure 6: change the caption to “diurnal variability of cluster fractional likelyhood”

Response: Thank you for the suggestion, which we have adopted.

Figure 6: it would be good to change the x axis label to 3-hr or 6-hr intervals. Same as other diurnal figures.

Response: Thank you for the suggestion, which we have adopted.

Figure 8: it would be interesting to add similar plots for aerosol number size distribution.

Response: Thank you for the suggestion. Such plots can be found in Marinescu et al., 2019 (e.g. see their Figures 3 and 4).

Line 291-292: I don’t understand why “indicate short-lived events tied to diurnal cycles”. Shouldn’t the nucleation mode particles be tied to new particle formation events?

Response: We agree, clearly new particle formation has occurred relatively close to the site. From these observations, we can’t tell whether nucleation has occurred in the boundary or aloft. The diurnal nature of observed nucleation mode concentration peaks can be due to the timing of nucleation itself or the timing of meteorological conditions that bring the small particles to the site. Hence, we did not want to make too strong a statement as to their origin.

Eq. 3 in Appendix A: as it mentioned GF at a given RH is abbreviated to GF above, and GF is used in Eq.2, here GF(RH) should be GF for consistency.

Response: Thank you, we have corrected this.

Data availability: reference (Atmospheric Radiation Measurement (ARM), 1995, 2001, 2007, 2010, 2011, 2015) are not seen in the reference list.

Response: The references in the data availability section have been fixed.

Data availability: The final merged aerosol number size distribution and CCN data link (https://doi.org/10.5439/1832908) doesn't work. Please check if the DOI is correct or if the data haven't yet published in the ARM Archive.

Response: Apologies, we were aware that although the DOI has been issued, ARM had not yet completed publishing the data in the archive. This DOI is now active and an updated citation for the archived data has been included.

Line 497: not complete: "is available at XXX"

Response: We have requested that these data be included as a Supplement. Creation of the appropriate link will be handled by the journal upon acceptance.

Reviewer 2

This study investigates the temporal variability of the five-year CCN observations at the ARM SGP site. The authors first constructed the CCN spectra database covering a wide range of supersaturation (SSw) conditions by combining concurrent aerosol size distribution and hygroscopicity (kappa) obtained from multiple instruments. They then conducted three statistical analyses using this data set, namely (1) parametric fitting of the CCN PDF; (2) cluster analysis of the CCN spectra; (3) temporal autocorrelation of the CCN number concentration time series. From each analysis they concluded that (1) the skewed log-normal distribution fits the CCN PDF; (2) there are three distinct clusters of CCN spectra and they are associated with aerosol size distributions dominated by different modes of aerosol particles; each cluster has distinct seasonal and hourly fractional occurrence; (3) the time scales of CCN number is shorter (1-3 hours) in high SSw condition than in low SSw condition.

Overall, this is a clearly written manuscript. I agree that these multiple-year CCN spectra spanning the SSw condition from strong updraft in deep convection to fog are potentially useful for simulating cloud and precipitation formation as well as for the development of microphysics parameterizations. Applying cluster analysis to the CCN spectra is also a novel idea to explore the long-term dataset. The quality of the analysis and presentation meets the standard of the ACP. However, I found several questions/clarifications that need to be addressed, as listed below. If these issues can be properly responded to or revised, I would suggest the manuscript be accepted for publication.

Response: We thank the Reviewer for these remarks and for the comments that have helped improve this manuscript. Our point-by-point responses are below.

Specific Comments/Major Questions

- The scientific goal of the study and the relationship of the three statistical analyses (i.e., log-normal fitting of CCN PDF, cluster analysis of CCN spectra, and autocorrelation of CCN number concentration time series) need to be further clarified

Response: Thank you for the thorough read and detailed suggestions. We respond to each below.

When I first read the abstract, I was confused by how the statistical analyses are interrelated (e.g., the clustering was based on the fitted CCN PDF, or the autocorrelation was applied to the time series of cluster occurrence). It was until I read through the entire manuscript that I realized they are actually three separate results. Therefore I suggest to

- Revise the abstract to explicitly clarify these analyses as three different tasks and their specific objectives.

Response: We have revised the Abstract as suggested. The relevant section now reads:

"We leverage this added statistical power to examine relationships that are unclear in smaller datasets. Our analysis is performed in three main areas. First, probability distributions of many aerosol and CCN metrics are found to exhibit skewed log-normal distribution shapes. Second, clustering 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. Third, the dynamics of CCN spectral clusters and concentrations are examined with cross-correlation and autocorrelation analyses, which assist in determining the time scales of changing CCN concentrations at different supersaturations and are important for cloud modelling studies."

- Add a few sentences to state the scientific goals of this study and the organization of this manuscript to the end of Sec.1

Response: This is a good suggestion. We have added the following text to the end of section 1: "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."

- Augment the method section – At least move some of the basic descriptions of the three statistical analyses from the appendix to the method section.

Response: We understand that leaving these out of the method section entirely may be somewhat misleading or confusing and have attempted to remedy this while trying to keep the paper streamlined. We have added the following at the end of Methods (Section 2):
"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."

- Restructure section 3 – It is strange that there is no result section and the discussion section immediately follows the method section. Please separate results and discussion clearly. The median kappa test, for example, belongs to the discussion section instead of the direct results.

Response: We feel that given the range of analytical methods and results in this manuscript that results and discussion are most clearly and concisely presented together for each topic. For the sake of clarity we have renamed section 3 to "Results and Discussion".

- More systematic and quantitative discussion on the uncertainty of the dataset is needed, particularly focusing on the assumption when deriving the CCN # for extreme SSw conditions. Please also discuss how the uncertainty in CCN spectra may influence the statistical analysis results.

Response: As discussed in response to Reviewer 1, unfortunately our biggest sources of uncertainty in the CCN distributions are due to uncertainties in the underlying composite size distributions, which complicates determination of uncertainties. As such, we have settled for a qualitative discussion of the uncertainties, and where we think they become most problematic.

The reviewer makes a good point that these uncertainties may affect the results of the statistical analyses, and that this should be stated in those sections. Text has been added at the end of Section 3.1.1 reading:

“Finally, it is important to emphasize that the uncertainties in the CCN spectra discussed in section 3.1 are not necessarily reduced by this statistical fitting approach, due to their potentially systematic rather than random nature.”

And at the beginning of Section 3.1.2 reading:

“The systematic uncertainties in our CCN distributions discussed in Section 3.1 are expected to be inherited into the characteristic cluster spectra shown here.”

Near the beginning of Section 3.1.3:

“Due to the differential nature of comparisons between clusters, the effect of uncertainties discussed in Section 3.1 is likely minimized.”

And after Figure 9 (near the start of discussion of autocorrelation analysis):

“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.”

- I also think the title needs to be revised to highlight either the value of the CCN data set or the statistical analysis results. “Short- and long-term temporal variability” is a vague description

Response: We have attempted to highlight the wide supersaturation range of the dataset, while also ensuring the title is descriptive of the analyses performed, by changing the title to: “Long- and Short-Term Temporal Variability in Cloud Condensation Nuclei Spectra over a Wide Supersaturation Range in the Southern Great Plains Site”

- Please provide more details on the cluster analysis.

  For example, what was the actual input (x) to calculate the distance metrics? The original CCN spectra (i.e. CCN # as a function of SSw) or some normalization was applied? When saying the clustering was based on distribution shape (L458), do you mean the clustered results exhibit distinct functional distributions of CCN # along with SSw? Or the CCN spectral shape was used as the input?

Response: We have attempted to clarify this section to address your questions.It now reads:
“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 distributions or to CCN spectra. Distance metric 3, however, produced well-resolved clusters based on distribution shape (how the CCN spectrum changes with SSw), 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.”

As the variability of the CCN spectra is orders of magnitude higher at the higher end of SSw (and similarly the variability in the aerosol size distribution is much higher at the smaller sizes), the clustered results would mainly be more sensitive to the high SSw conditions (nucleation mode aerosols). Have you compared the clustering based on the normalized spectra, the normalized aerosol size distribution, or even the fitted CCN PDF as a function of SSw?

Response: As shown in Figure B1, it does not appear to be the case that there is relatively more variability at the high end of the CCN spectrum compared to the low end. Observed concentrations span about two orders of magnitude at 0.01% SSw, while at 5.3% SSw they span about one order of magnitude. Because of this, and the cluster distance metric used for clustering, we don't think there is reason to believe that clustering is dominated by high SSw conditions or nucleation mode aerosols (although, as discussed in the manuscript, nucleation mode aerosol behavior is well resolved between the clusters). At one point early in the project we did explore clustering with normalized spectra, but they generally produced less unique/well resolved clusters than the described method.

The authors mentioned that Kappa is not the main controlling factor of CCN spectra type. Maybe the aerosol composition in SGP is similar? Please add some discussion on the aerosol composition variation in this dataset.

Response: This is a good point, and we attempt to discuss this in detail in Figure 4 and the text preceding it. Some additional text has been added to this section, along with an additional reference to the study analyzing the kappa data used in this work, which reads:
"…with the results shown in Figure 4. Estimates are generally least reliable for lower supersaturations, with estimates below 0.2% SSw having a 95% confidence interval broader than ±50% of the estimated value. Therefore, caution should be taken interpreting estimated CCN spectra in this low SSw region. This highlights the region of the CCN spectrum that is most sensitive to observed variations in κ (Mahish and Collins, 2017), and the uncertainty in our data product below about 0.03% SSw."

- It seemed all the "size distribution" in the manuscript refer to "aerosol size distribution"? Or do some of them actually refer to the "CCN size distribution"? Please make them more explicit.

Response: This is a good point, thank you. For clarity we have added after the first reference in the introduction: "aerosol number size distributions (shortened to size distributions hereafter)" and used simply "size distribution" for consistency in the remainder of the text.

Technical corrections/Minor Comments

L23 Abstract: More specific/qualitative conclusion on the time scales

Response: Altered the text to now read: "Third, the dynamics of CCN spectral clusters and concentrations are examined with cross-correlation and autocorrelation analyses. We find that 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."

L67: Adding a table to list the instruments used in constructing the data set

Response: We have created a table of instruments and measurements and placed it within the methods section, and included a reference to it around line 67.

L95, Fig.1 Caption: blue black

Response: Corrected, thank you.

L218 reconciling seasonal particle data and cluster: please elaborate the sentence

Response: Added an additional reference to make it clear we are discussing reconciling data from previous analyses with the current one. This sentence now reads: "An important consideration for reconciling 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."

Data availability and supplement link

Response: See below. We supplied these to the journal upon submitting the manuscript for review, and the link will be provided if the article is published.

L491 "(ARM), 1995, 2001, 2007, 2010, 2011, 2015)": it's confusing that the years mentioned here are inconsistent with the years mentioned in the main text (2009-2013)

Response: These are references to additional citations for instrumentation used, and have now been cited correctly in the "Data Availability" section. The dates indeed do not match the periods of data collection, but we are following the reference information and formats suggested by DOE for this data, see https://www.osti.gov/dataexplorer/biblio/dataset/1025259 or others using provided DOIs. We agree that this is especially confusing for the data product associated with this paper, which was just activated but shows a publication date of 2009 https://www.osti.gov/dataexplorer/biblio/dataset/1832908, but we still believe it best to match the official DOE records.

L494 "merged aerosol number size distribution and CCN data (https://doi.org/10.5439/1832908)": This DOI cannot be found in the DOI System

Response: Apologies, we were aware that although the DOI has been issued, ARM had not yet completed publishing the data in the archive. This DOI is now active.

L497, " skewed log-normal fit coefficients for all CCN data, named CCN_fit_coeffs.txt, is available at XXX". missing information

Response: We have requested that these data be included as a Supplement. Creation of the appropriate link will be handled by the journal upon acceptance.