Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1297-RC1, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Classification of aerosol population type and cloud condensation nuclei properties in a coastal California littoral environment using an unsupervised cluster model" *by* Samuel A. Atwood et al.

Anonymous Referee #1

Received and published: 3 February 2019

This is a review of "Classification of aerosol population type and cloud condensation nuclei properties in a coastal California littoral environment using an unsupervised cluster model" by Atwood et al. This manuscript reports on particle size distributions and cloud condensation nuclei measurements made during a field campaign run during CalWater-2015. After cleaning the data the authors identify eight distinct clusters of particle types which they believe are statistically distinct.

The manuscript is clearly deserving of publication and I see no major issues with either the experimental technique or the data reduction and processing techniques. However,

I do see the need to further explain several of their techniques and choices and to clarify some areas.

POINTS OF CLARIFICATION (MAJOR AND MINOR):

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

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

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

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

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

Pp6, section starting 2.4.1 requires more attention than the rest of the paper. These sections require significant clarification. -It appears that the authors used a hierarchical clustering scheme to identify the ideal number of clusters. How was this done? Was it bootstrapped? What portion of the data was used? If all the data why then go back to a k-means scheme? -There is inconstancy in the reported number of variables being used in the classification. In line 14 it is reported as 24, in line 24 it is suggested that it is 20. These should be made consistent AND exactly what each of these 20 or 24 variable is should be clearly identified. -The choice of a Pearson Euclidian distance is confusing in this case and should be further justified. Is this the same thing as a Karl Pearson distance in which weights are usually standardized by standard deviation? If normalizing the weights why not just use a Euclidian distance? If a true Karl Pearson Euclidian distance function is being used why aren't the weights a reciprocal of variance rather than evenly distributed by the number of variables? This section requires significant justification and explanation. -In line 20, missing data are referenced. The fraction of missing data and which variables are most often missing should be specified.

In line 8 (Pp7), the "physical interpretability" should be clarified. Is this just empirical judgment? Pp7 line 21 and Pp11 lines 13 and 14: The reference manager again seems to be troubled. These references need to be fixed.

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

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

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

Despite these areas that require clarification this paper should clearly still be published after these minor modifications are made and I congratulate the authors on their work.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1297, 2019.