

General Comments:

The manuscript by Wang et al. focused on supervised algorithms to tackle the problem of automatic classification of large amounts of aerosol particle data. They created a labeled benchmark dataset containing ~24,000 particles with eight different coarse categories. They used the dataset from a week-long summer campaign to train five models popular for machine learning applications and compared their performance. The prediction accuracy of several supervised algorithms exceeded 97 % for the classification of the eight classes. The topic fits well within the scope of AMT. This manuscript is generally well written. Before its publication, the following comments need to be addressed.

Specific Comments:

Is there any specific reason for adopting June 29 as the demarcation between labeling and blind data? Please specify. There might be a significant uncertainty when choosing different demarcations. The authors need to address such uncertainties in the revised manuscript.

In Section 3, a brief description of the sampling lines, flow rate, and residence time for all measurements would be good. Further, it would be beneficial if the authors provide more details about the calibrations for the SPMS.

More evidence needs to be listed to support the the eight coarse particle classes used in this study. Why didn't the authors choose 7 or 9 coarse particle classes instead of 8? Some secondary particles are classified as primary emissions, e.g., K-rich particles and OC-EC particles. I also notice that V-rich particles contain 54/56Fe+ signal. Does it mean that the resolved V-rich particles were a mixed factor? How about combining the two factors together?

It would be better to add detailed comparisons (including time series and mass spectra) between the predictions and the the results obtained in a previous study using ART-2a. Otherwise, we don't know how significant the similarity variations are.