In the revised manuscript, the authors have made substantial effort to improve the approach on assessing the model performance. I only have a few minor comments for the authors to consider.

I wonder how the Kendall Tau rank correlation metric is defined/calculated. It would be better to move lines 317-323 to line 210 to give the readers some ideas first. It is also not very clear to me how Z-scores are defined. $(X-\mu)/\sigma$?

The authors use "bin" in many places which refers to grouped/averaged range of different variables, like vertical levels, altitude, longitude/latitude. I would suggest using different word so that readers can know better what exact "bin" the authors are talking about.

Lines 208-210, it is not clear to me for the definition of the mean CALIPSO aerosol layer fraction. For one column during an altitude range, is it the number of vertical levels having the subtype of interest divided by the total number of vertical levels in this altitude range? It is the "overlapping with the vertical altitude bin of interest" making me confused.

Lines 219-222, I'm lost while reading it. Please try to rephrase it or cut it into several sentences.

Figure 3, I would suggest using the same x-axis range (-0.2 to 1.0?) for profiles both including "polluted dust" and excluding "polluted dust". It would be clearer to show that removing polluted dust has a small effect on relationships with combustion aerosols.

Figure 4, it would be better that some of the figure caption goes to the main text, which is more related to the discussion/interpretation of the results.

Lines 613-614, Thorsen et al. (2011) does not have the journal name.