Dear Hailong Wang, editor,

Thank you for accepting our paper for publication, pending technical corrections! Our corrections are listed below.

We also wanted to request the addition of a few new citations of relevant recent 2022 work that we have become aware of since the last submission. The citations and their textual context is provided below and does not change the conclusions or methodology of the paper. We believe adding these citations makes this a more cutting-edge contribution, and so we hope it will not be a problem to add them in. However, we also understand if that is not possible at this time. To that end, we have added the below text into the production file uploads for now, but will happily remove again it if it is deemed necessary. Thank you again for your consideration,

Sincerely,
-Lauren Zamora, on behalf of all the authors

Technical corrections

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.

Thank you.

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.

The lines in question have been moved up, as suggested.

To clarify the software used in this analysis, which is relevant to how the Kendall-Tau and Z-scores were calculated, we added the following information to the beginning of section 2:

"All data analysis was performed using the R language and environment for statistical computing version 4.2.0 (R Core Team, 2022)."

We then added the following text to clarify how the Kendall Tau rank correlation is defined (new text in bold):

"**The robust Kendall Tau rank correlation metric (τ) was chosen to assess correlations. It is calculated by examining all possible combinations of two data points in the data set and scoring them as either concordant (positive slope) or discordant (negative slope). τ is defined as:**

$$\tau = \frac{C-D}{C+D} \quad (2)$$

where C and D are the number of concordant and discordant pairs. In the case of ties, where a pair of data points is neither concordant nor discordant, a small correction is made (τ_b, following Kendall (1945)).

τ is a nonparametric, rank-based alternative to R^2 that is robust to outliers, makes no assumptions about the data distributions, and is thus a better metric of correlation for many types of data (Shevlyakov & Oja 2016). τ ranges from -1 for a perfect negative correlation to 0 for no correlation to +1 for a perfect positive correlation. In cases where the use of R^2 would be appropriate, the magnitude of τ is a good estimator for R^2, with a maximum theoretical asymptotic difference < $\pm$0.11 (Shevlyakov and Oja 2016). For reference, Figure S3 shows corresponding R^2 values."

As requested, we also added the following information to clarify how Z-scores are defined:

"Next, we wanted to better understand where aerosol distributions between MERRA-2/FLEXPART and CALIPSO have the highest agreement. For this step, we assessed the difference in Z-scores between MERRA-2/FLEXPART concentrations and aerosol layer presence in CALIPSO. Z-scores in MERRA-2/FLEXPART for dust, BC, and OC are defined as the number of standard deviations from the mean respective dust, BC, or OC value across the study region at a given altitude level. Locations with high aerosol levels in MERRA-2/FLEXPART will have high positive Z-scores, and locations with low aerosols will have negative Z-scores. Similarly, Z-scores in CALIPSO are defined as the number of standard deviations from the mean dust or combustion aerosol layer fraction across the study region at a given altitude. Because Z-scores are unitless, they can be compared between MERRA-2/FLEXPART and CALIPSO, which has different units. This approach also enables relative comparisons between MERRA-2 and FLEXPART aerosol distribution patterns even if the concentrations are on different scales."

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 words so that readers can know better what exact "bin" the authors are talking about.

Thank you, changed as suggested.

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.

We clarified the text as follows:

"To begin, we assessed the correlation between MERRA-2/FLEXPART aerosol concentrations and mean CALIPSO aerosol layer fraction. Correlations were calculated only on cloud-free days. CALIPSO aerosol layer fraction was defined as the fraction of the CALIPSO aerosol layer of the subtype of interest - dust or combustion – within each altitude bin for an individual CALIPSO observation. Then, as CALIPSO has much finer vertical resolution than either MERRA-2 or FLEXPART, CALIPSO dust layer fraction and MERRA-2/FLEXPART dust concentrations were averaged within 20° longitude × 6° latitude bins and at model vertical resolution for FLEXPART, and at 1 km vertical resolution for MERRA-2. Similarly, CALIPSO combustion aerosol fraction was compared with MERRA-2/FLEXPART BC and OC concentrations. We excluded aerosol layers within 200 m of the surface in the lowest altitude bin and longitude-latitude-altitude bins with < 30 observations (collectively, < 1% of total data). Data were averaged either across the 8-year study period for one season (e.g., December through February) or for daytime/nighttime samples, as stated in the text.*"*

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

Rephased as follows:

"Sometimes CALIPSO had multiple overlapping layers of aerosol subtypes of interest, which is a result of CALIPSO's multiscale averaging approach (Thorsen et al., 2011). In those cases, we summed of only the portions of each CALIPSO layer that did not overlap with the other CALIPSO aerosol layer and that fell completely within the MERRA-2/FLEXPART bin. The sum of these portions was then divided by the entire height of the MERRA-2/FLEXPART bin to provide the fraction of the model bin filled by a CALIPSO aerosol layer."

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.

We have changed the x-axis to be consistent in this figure, as suggested.

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.

We have moved the following information to the text from the Figure 4 caption:

"Because MERRA-2/FLEXPART and CALIPSO indicate aerosol presence in different units, they cannot be directly compared. The Z-score differences help locate where the overall patterns agree best between the two products. For example, if at a given location, MERRA-2 aerosols were three standard deviations above the mean and CALIPSO aerosol layer presence was only one standard deviation above its mean, that would result in a Z-score difference of two. The closer the Z-score difference is to zero, the more agreement there is between the products."

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

Fixed, thanks.

Proposed addition of new references and related text

New text is shown in bold.

In section 2.1.2:

"There are several studies that have evaluated MERRA-2 Arctic aerosol distributions, mainly using MODIS aerosol optical depth and ground-based observations (e.g., Wu et al. (2020), Lee et al. (2020) and Sitnov et al. (2020)). MERRA-2 BC and OC aerosols tend to be a bit high compared to observed aerosol concentrations, although they tend to follow the qualitative trends (Vinogradova et al., 2020; Zhuravleva et al., 2020; **Xian et al., 2022**). One study found that dust optical depth and dust extinction were similar or a bit elevated compared to that of CALIPSO, but with large discrepancies in absolute concentrations (both over- and underpredicting concentrations) compared to two ground sites (Wu et al., 2020)."

In section 3.2:

"**It is important to note that the impacts on clouds of the dust and combustion aerosols of focus in this study may be much smaller than those of local marine emissions, particularly at times when the surface is not covered with snow or ice. For example, one recent satellite study found evidence that mixed phase cloud formation occurs above homogeneous freezing on average but at colder wintertime temperatures over both the Siberian and Canadian archipelago regions than over other times and Arctic locations (Carlsen and David, 2022), which the authors attribute to fewer marine ice nucleating particles being emitted there.**"

In the conclusions:

"**This study focused on dust and combustion aerosols. However, other recent studies indicate that marine aerosols may be particularly important for the Arctic ice nucleating particle budget (e.g., Carlsen and David (2022)). For example, Porter et al. (2022) found that biogenic particles may dominate the ice nucleating particle budget even over the Russian coast where we found that the dust aerosol impacts are likely to be highest. Thus, aerosol sources besides dust may enhance aerosol impacts on clouds over this region even further and could have a larger impact on clouds across the Arctic region than the long-range transported aerosols of focus in this study, particularly at times when the surface is not covered with snow or ice.**"

References

Carlsen, T. and David, R. O.: Spaceborne Evidence That Ice-Nucleating Particles Influence High-Latitude Cloud Phase, Geophysical Research Letters, 49, e2022GL098041, https://doi.org/10.1029/2022GL098041, 2022.

Porter, G. C. E., Adams, M. P., Brooks, I. M., Ickes, L., Karlsson, L., Leck, C., Salter, M. E., Schmale, J., Siegel, K., Sikora, S. N. F., Tarn, M. D., Vüllers, J., Wernli, H., Zieger, P., Zinke, J., and Murray, B. J.: Highly Active Ice-Nucleating Particles at the Summer North Pole, Journal of Geophysical Research: Atmospheres, 127, e2021JD036059, https://doi.org/10.1029/2021JD036059, 2022.

Xian, P., Zhang, J., O'Neill, N. T., Toth, T. D., Sorenson, B., Colarco, P. R., Kipling, Z., Hyer, E. J., Campbell, J. R., Reid, J. S., and Ranjbar, K.: Arctic spring and summertime aerosol optical depth baseline from long-term observations and model reanalyses – Part 1: Climatology and trend, Atmospheric Chemistry and Physics, 22, 9915–9947, https://doi.org/10.5194/acp-22-9915-2022, 2022.