Responses to Anonymous Referee #1

The authors have addressed most of my previous comments. In particular, the interpretation and comparison of the regression analysis and WRF-Chem results have been improved.

The authors also performed CESM2 experiments to test further the model sensitivity of the circulation response to dust forcing. The authors find a quantitative difference in the storm track shift response between CESM2 experiments and the WRF-Chem experiments. It is discussed that the difference is due to differences in aerosol microphysical properties and processes.

While I greatly appreciate the authors for doing the extra work of using CESM2 models, the current comparison with the WRF-Chem experiment is not correct. Unlike CTRL and NoDust from WRF-Chem (which have the same February initial condition), atmospheric circulations in CESM2 will already be different in February between the runs with and without dust. Therefore, the difference between the two CESM2 runs includes both the impact from dust forcing and initial condition differences. First, means that the difference between the two CESM2 runs is not entirely due to dust forcing. Second, this means that the different response between WRF-Chem and CESM2 is also related to the different February conditions. Ultimately, this reflects the difference between a NWP model like WRF and a GCM like CESM2.

I think this is an important issue to be revised before the paper is published. Overall, I consider the manuscript sound enough already without the extra CESM2 analysis. The authors may choose to remove the related paragraph and supplementary Figure. Another option could be to revise the discussion of the CESM2-WRF comparison.

At the same time, I think the authors can make their point stronger by correctly utilizing the CESM2 simulations. If the authors run the two CESM2 runs longer (~40 years, as the reanalysis), they can compare the interannual variability of the North Pacific storm tracks with and without dust activity directly. If the interannual variability of storm track position is greater in the run with dust, it will make the authors' argument stronger. However, as I mentioned above, the current paper is sound, so I'm leaving it here simply as a suggestion for the authors.

Responses: We thank the reviewer for this careful and insightful comment. We realize

that our description of the difference between the WRF-Chem and CESM2 experiments in the previous revision was not accurate. In WRF-Chem, the CTRL and NoDust ensembles share identical initial and lateral boundary conditions, so their difference isolates the dust effect. In the free-running CESM2 simulations, by contrast, the atmospheric circulation in the runs with and without dust has already diverged by February. Their difference therefore includes both the dust forcing and the influence of different February conditions. As the reviewer points out, the different responses in the two models largely reflect the difference between a regional weather model driven by analysis fields and a free-running global climate model, rather than differences in dust forcing alone. As the reviewer notes, the main conclusions of the manuscript do not depend on the CESM2 analysis. **We have therefore followed the reviewer's first suggestion and removed the CESM2 experiments from the revised manuscript, including the description in Section 2.3, the related discussion in Section 4, and Figure S9 in the Supplement.** We greatly appreciate the suggestion to run long free-running integrations with and without dust and compare the interannual variability of the North Pacific storm track directly. This would be a valuable test of our argument, and we plan to pursue it in future work.