Response to Reviewer 1

I have reviewed "Strong Cloud-Mediated Aerosol Cooling in TaiESM1 Diagnosed Using Cloud Radiative Kernels and APRP Decomposition" by Tsai et al., who used these methods to study the response of the climate to anthropogenic aerosols in the TaiESM Earth system model.

The manuscript is engagingly written (with a few rough spots due to length or content noted below). It is nice to see application of these state-of-the-art methods to diagnosing aerosol-cloud interactions in the model climate. A large portion of the results are necessarily TaiESM-specific, but the authors do present interesting findings that may generalize to the way we diagnose ERFaci in other ESMs (or in general). These findings concern the importance of the climate base state used in the ERFaci diagnosis, the importance of ERFaer to the historical temperature evolution across models, and the relative unimportance of atmosphere-ocean coupling. On the plus side, this makes the study interesting for the wider community. On the minus side, I have some methodological concerns specifically for these findings, which I would like to see addressed (or rebutted, since it is possible I misunderstood the method) before I am comfortable recommending publication.

We thank the reviewer for the careful and constructive assessment. The main changes in the revised manuscript are:

(1) fixed-SST piClim-control and piClim-aer simulations with TaiESM1, which follow the protocol of Zelinka et al. (2023) for the APRP-based ERF components and provide a benchmark for the residual kernel method;

(2) the historical − piNTCF contrast in the kernel analysis (revised Fig. 4), with piNTCF − piControl moved to the Supplement (Fig. S1);

(3) maps of the changes in liquid water path and cloud droplet number (revised Fig. 2) and a model-specific discussion of why ERFaci is strong in TaiESM1 (Sect. 4.1);

(4) a comparison with the fixed-SST results of Zelinka et al. (2023), with ERFaer isolines and a 1:1 line replacing the regression lines of the original Figs. 6 and 7 (revised Figs. 7 and 8);

(5) a more detailed description of the methods in Sect. 2 (cloud radiative effect versus cloud radiative response, the definition of α, histSST and the experiment protocols);

(6) a shorter Sect. 4.2, refocused on the SST-mediated cloud response and the effect of air–sea coupling.

With the fixed-SST protocol, TaiESM1's ERFaci (−1.98 W m^{-2}) is almost unchanged from the original estimate (−2.00 W m^{-2}), so the main conclusion of a strong cloud-mediated aerosol cooling is robust. The smaller components, and the residual kernel estimates for the coupled contrasts, do depend on whether the SST is coupled or fixed, as the reviewer suspected. Point-by-point responses follow.

Scientific results:

- If this study is intended to guide further model development (l. 46, l. 378-385), it would be nice to know which part of the cloud scattering response is large (Nd or LWP) since very different processes are implicated depending on the answer to this question. This shortcoming could be addressed by a figure (between Figs. 1 and 2) showing Delta LWP (or even better Delta log LWP) and a figure showing Delta CDNC (or Delta log CDNC) because (a) then the reader can see all the changes entering into Delta CRE and (b) it also give the reader an idea about the partitioning between droplet activation and LWP adjustment, which APRP doesn't provide.

For the same reason (forcing and adjustment processes are not separated by either method employed), I would dispute the use of "process-oriented" in the Conclusions (l. 458).

Reply: We agree that the APRP decomposition cannot separate droplet activation from liquid water path adjustment, and that this information is needed if the results are to guide model development. Following the suggestion, a new figure (revised Fig. 2) has been inserted between the original Figs. 1 and 2, and the subsequent figures have been renumbered. It shows the annual-mean changes in LWP and CDNUMC (vertically integrated cloud droplet number) for the same three contrasts as Figs. 1 and 3 (historical − piControl, historical − piNTCF and piNTCF − piControl), with the global-mean value in each panel, so that all changes entering ΔCRE can be followed across the figures.. We also agree that neither diagnostic separates forcing from adjustment processes. The phrase "process-oriented analysis" has been removed from the Conclusions and replaced by "diagnostic analysis".

- l. 85: probably good to cite some of the extensive pre-Chen2021 literature on the feedback-forcing inverse correlation (e.g., Kiehl 2007 https://doi.org/10.1029/2007GL031383, Forster et al. 2013 https://doi.org/10.1002/jgrd.50174).

Reply: Thank you for these references. The paragraph at lines 79–85 now notes that the inverse relationship between climate sensitivity (or feedback strength) and aerosol forcing across models was identified well before CMIP6, citing Kiehl (2007) and Forster et al. (2013), before introducing the cloud-specific compensation documented by C. Wang et al. (2021).

l. 365-377: Thinking further about the previous point in response to the discussion here. It is a bit hard to evaluate the statements about forcing-feedback compensation between models when the feedbacks are not shown. But, accepting the premise, doesn't the fact that the Delta T_s varies by a factor of 6 (by eyeball, since the actual values are not given) among the models shown argue against the "compensation mechanism

proposed by Wang et al."? I would expect the Delta T_s differences between models to be suppressed if the forcing and feedbacks compensated each other. If Fig. 6 included isolines of total ERFaer, it seems to me (again by eyeball, since the values are only shown by color scale) that Delta T_s follows ERFaer quite closely, as would conventionally be expected (e.g., Rotstayn et al. 2015, https://doi.org/10.1175/JCLI-D-14-00712.1).

It would be interesting to include a time series of the diagnosed forcing and coupled-model SST to support the statements made in this paragraph. If the historical time evolutions were presented, the presence or absence of "pothole cooling" in the 1970s and "catch-up warming" in the more recent past, Zhang et al 2021 (https://doi.org/10.5194/acp-21-18609-2021) would break the degeneracy between the strong-forcing/strong-feedback and weak-forcing/weak-feedback model temperature evolution to some extent. Otherwise, the conclusion that "the realized warming amplitude also depends on greenhouse-gas forcing, cloud feedback strength, ERFari, ocean heat uptake, and circulation adjustment. Therefore, TS should be interpreted as an integrated climate response" is obviously true but not obviously connected to what is seen in TaiESM.

Reply: We agree. The large spread in ΔTs argues against the compensation mechanism, and ΔTs instead follows the total aerosol forcing, as in Rotstayn et al. (2015). The compensation statement in Sect. 4.1 has been removed, and the generic concluding sentence has been replaced by one linking the strong ERFaer of TaiESM1 to its small ΔTs.

Following the suggestion, Fig. S2 now shows the annual ΔTs and aerosol-mediated cloud radiative response of the TaiESM1 historical and AerosolOnly simulations for 1985–2014. The historical ΔTs stays within about 0.2 K of the piControl level until the mid-1990s and then rises to about 0.8 K by 2010–2014, whereas the aerosol-mediated cloud radiative response remains strongly negative without a weakening trend, and the AerosolOnly simulation stays about 1.2 K colder than piControl throughout. This delayed warming is consistent with the behavior of CMIP6 models with strong aerosol forcing described by Zhang et al. (2021). A sentence referring to Fig. S2 has been added to Sect. 4.1.

- l. 271-280: I find it hard to believe that the difference between the diagnosed aerosol cloud response in the historical vs AerosolOnly experiments is physical. Rather, I think it must be an artifact of the method. I was very confused by this paragraph, however, so correct me as needed. I think AerosolOnly uses a coupled ocean; if so, the climate cools relative to piControl. Trying to use the 1%CO2 cloud feedback parameter may not work very well here because the cloud feedbacks are different between warming and cooling whenever there is a component due to cloud

thermodynamic phase or due to liquid vs ice sea surface. I think the bright red bands in the high-latitude oceans in the AerosolOnly aerosol-mediated cloud response (Fig. 3b) are highly suggestive that something like this is going wrong in the method, since there is no large positive ERFaci in the Southern Ocean in the actual climate. But if this is the problem, it will affect most of the rest of the globe as well (anywhere the clouds aren't pure liquid in both climate states), so it could explain why the global means are so different. I am not sure what "preindustrial background state" on l. 278 means ("background" is too imprecise), but whether it means temperature or atmospheric composition, I don't think it can explain a factor of 2 in diagnosed ERFaci (assuming that -2.08 W m-2 and -1.15 W m-2 are the numbers I'm meant to compare here).

Reply: We thank the reviewer for this comment. The reviewer is right that the factor-of-two difference is not physical. It arises from two issues, which the new fixed-SST simulations allow us to assess.

The original comparison was between non-equivalent contrasts, because historical − piControl also contains the responses to the non-aerosol historical forcings. For the equivalent contrast, historical − piNTCF (revised Fig. 4c), the residual is −1.21 W m^{-2}, close to the AerosolOnly − piControl value (−1.15 W m^{-2}), and the two contrasts also have similar total responses (−1.88 and −1.77 W m^{-2}) and surface cooling (−1.27 and −1.17 K). The residual method is also less reliable in a cooling climate. In the new piClim-aer − piClim-control simulations, ΔTs is negligible, so the kernel residual directly gives the cloud contribution to the aerosol ERF: −1.45 W m^{-2} (revised Fig. 4d), stronger than both coupled estimates and without the positive bands over the high-latitude oceans. As the reviewer suspected, these bands and the weaker coupled estimates arise from the $\alpha \cdot \Delta Ts$ term, which overcorrects for the SST-mediated cloud changes in the cooling coupled climate (see the reply to the comment on l. 335–344). The imprecise “background state” interpretation has been removed, and Sect. 3.2 has been revised accordingly.

- l. 335-344: You have invested a significant amount of work into implementing both APRP and the kernel-based method. I think you let yourself down a bit in this paragraph by simply saying "these are different quantities", and in general by mostly discussing these results separately rather than connecting them with each other. If I understand correctly, the difference is that the aerosol cooling of the coupled SST induces circulation changes that modulate the cloud responses to aerosols, which by definition (it involves Delta SST) is not part of the aerosol ERF. You already have the results to tell an interesting story along the lines of Soden and Chung (2017), (I think) simply by plotting the difference between ERFaci and the Soden/Chung aerosol-mediated cloud response. I think you should!

Reply: We thank the reviewer for this suggestion, which the new fixed-SST simulations allow us to follow. Because the APRP ERFaci is shortwave only, whereas the kernel residual includes both shortwave and longwave responses, the comparison is made between kernel diagnostics. Following Soden and Chung (2017), the difference between the coupled AerosolOnly − piControl and the fixed-SST piClim-aer − piClim-control responses isolates the cloud response to aerosol-induced changes in SST and circulation, which is not part of the aerosol ERF (revised Fig. 9c). The aerosol-induced cooling (−1.17 K) makes the total cloud radiative response more negative by 0.30 W m^{-2}, mainly over the North Pacific, but this is only about half of the $\alpha \cdot \Delta Ts$ term (−0.62 W m^{-2}) removed by the residual method. The coupled aerosol-mediated estimate (−1.15 W m^{-2}) is therefore 0.30 W m^{-2} weaker than the fixed-SST estimate (−1.45 W m^{-2}). This analysis has been added to Sect. 4.2, and Sect. 4.1 now links the fixed-SST kernel estimate to the APRP ERFaci.

- Fig. 6: Where do the values for the CMIP6 models come from? Did you compute them yourself using the Zelinka et al. (2023) method? The values are quite different from the values for the same models in Zelinka2023 Table 2. Is the difference because you used different experiments in your calculation than Zelinka et al. used in theirs? But in that case, isn't the fact that the ERF components can be so different a troubling reflection on the method?

Also, note that the modeling center name alone is not always enough to identify a model (e.g., IPSL).

Reply: We thank the reviewer for this question. The CMIP6 values in the original Fig. 6 were computed by us with the APRP method of Zelinka et al. (2023), but applied to the coupled historical − piControl difference (last 30 years) instead of the fixed-SST piClim-aer − piClim-control experiments used by Zelinka et al. (2023). The coupled difference also contains the greenhouse-gas forcing and the cloud response to the resulting changes in surface temperature and circulation, so its APRP terms are not effective radiative forcings in the strict sense. The discrepancy with Table 2 of Zelinka et al. (2023) therefore reflects the use of non-equivalent experiments rather than a deficiency of the APRP method.

In the revised manuscript, the APRP decomposition for TaiESM1 is based on new piClim-aer and piClim-control simulations that follow the same protocol as Zelinka et al. (2023), and revised Fig. 7 (former Fig. 6) shows the CMIP6 values taken directly from their Table 2 together with TaiESM1. Models are now identified by their full names (e.g. IPSL-CM6A-LR and IPSL-CM6A-LR-INCA). For TaiESM1, the fixed-SST ERFaci (−1.98 W m^{-2}) is almost identical to the coupled estimate (−2.00 W m^{-2}), whereas ERFari and the cloud-amount term differ between the two protocols.

Presentation:

l. 98-103: This paragraph is a bit choppy. I don't think it is necessary nowadays to underscore not "relying solely on aerosol optical depth". The statements about the relative importance of RF and adjustments to ERF and about the activation coefficient reflect a narrow sliver of the breadth of work being done in ERFaci world, and more to the point not questions that APRP can address, since activation and LWP adjustment are not separated. Why not simply point out that decomposing the aerosol effects on climate allows us to understand different mechanisms (ERFaer components; the Soden/Chung aerocol-mediated cloud response) and that comparing ERFaer component by component to other lines of evidence makes it more likely to catch compensating errors in the ERFaer?

Reply: We agree with the reviewer. We have revised this paragraph to move away from the previous emphasis on aerosol optical depth and the activation coefficient and instead highlight the value of decomposing aerosol effects into distinct radiative pathways. The revised text emphasizes that component-based diagnostics can help identify the mechanisms underlying the total aerosol radiative response and reveal compensating effects that may otherwise remain hidden. The revised text reads:

"These complexities also make it important to distinguish the radiative pathways through which aerosols influence climate. Separating aerosol–radiation effects from cloud-mediated radiative changes can reveal whether a model's total aerosol cooling arises primarily from aerosol optical effects, aerosol-induced modifications of cloud properties, or subsequent cloud adjustments associated with the climate response. Such component-based diagnostics are particularly useful because similar net aerosol forcing can result from different combinations of underlying processes, allowing compensating errors to remain hidden when only the total radiative response is considered. Comparing individual components with independent observational constraints and other models can further help identify such compensating errors (Bellouin et al., 2020; Zelinka et al., 2023). A decomposition of the radiative response therefore provides a basis for relating model-specific aerosol–cloud behavior to broader differences in aerosol forcing and cloud feedback across climate models."

l. 140: would be good to list the ACI-relevant parameterizations (i.e., does TaiESM inherit the Park and Bretherton cloud/boundary layer scheme and MG microphysics?) and whether the new convection scheme is coupled to cloud microphysics

Reply: Thank you for comments. Section 2.1 now describes the parameterizations relevant to aerosol–cloud interactions. TaiESM1 inherits the Morrison and Gettelman (2008) two-moment stratiform microphysics and the University of Washington moist turbulence (Bretherton and Park, 2009) and shallow convection (Park and Bretherton, 2009) schemes from CAM5, while the cloud macrophysics (cloud fraction) is replaced by the GTS scheme (Shiu et al., 2021). Aerosols are represented by SNAP (Chen et al.,

2013), and droplet activation follows Abdul-Razzak and Ghan (2000). The deep convection scheme is not coupled to the cloud microphysics.

l. 169: "combined influence of ..." add "and the ERF of the other forcing agents"? Since there is no mention of, e.g., the cloud adjustment to the CO2 forcing, does that mean Wang2021 conclude this component is negligible? If so, it would be good to state that here so the reader understands why the CO2 ERF is absent from eq. (1). Is this also the reason you say "... historical – piNTCF as the aerosol-related component ..." on l. 208 although historical – piNTCF technically includes the CO2 ERF?

Reply: We thank the reviewer for this question. The cloud adjustment to CO_2 is not neglected. The parameter α is the normalized cloud radiative response from the 1pctCO_2 experiment (the cloud radiative response divided by ΔTs), not a regression slope, so it includes both the temperature-mediated cloud feedback and the rapid cloud adjustment to CO_2. Removing α·ΔTs from the historical cloud radiative response therefore also removes the CO_2-induced cloud adjustment, which is why no separate CO_2 term appears in Eq. (1). In the revised Sect. 2.2, the sentence containing "combined influence" has been replaced by a description of α_1pctCO_2 and of its difference from the abrupt-4xCO2 cloud feedback.

The reviewer is right that the historical − piControl residual still contains the cloud adjustments to other forcing agents, such as CH_4, N_2O, ozone and land use. A sentence stating this has been added to the revised Sect. 2.2. For the historical − piNTCF contrast, the CO_2 forcing is identical in both experiments and therefore cancels. Because TaiESM1 has no interactive chemistry, CH_4 and ozone are also identical, so this contrast isolates the response to anthropogenic aerosol and aerosol-precursor emissions. The corresponding sentence in Sect. 3.1 has been clarified accordingly.

l. 190-193: APRP ERFaci and ERFari definitions differ from the IPCC definitions (because cloud adjustments to ARI are part of ERFari in the IPCC definition and part of ERFaci in the APRP definition), see the discussion after eqs. (6-7) in Zelinka2023. It would be good to mention this here.

Reply: Thank you for pointing out this distinction. We have added a clarification in Section 2.3 explaining that the APRP-based definitions of ERFaci and ERFari differ from those adopted by the IPCC. Specifically, cloud adjustments to aerosol–radiation interactions are included in ERFaci under the APRP framework, whereas they are attributed to ERFari in the IPCC definition, as discussed by Zelinka et al. (2023).

l. 218: I think it is a mathematical identity that line 1 = line 2 + line 3 for all the plots in Fig. 1, correct? Maybe a few remarks about why the piNTCF - piControl Delta AOD is so large (~30% of the historical - piControl) or equivalent why the difference between historical - piControl and piNTCF - piControl Delta AOD is so large?

Reply: The reviewer is correct: the first row equals the sum of the second and third rows, because the same piControl climatology is used in all contrasts. Because anthropogenic aerosol and precursor emissions are identical in piNTCF and piControl, the piNTCF − piControl ΔAOD (0.006, about 30 % of the historical − piControl value of 0.019) comes from aerosols outside the NTCF perturbation, mainly dust and sea salt, whose emissions respond to changes in surface wind and soil moisture under the non-aerosol historical forcings. A sentence stating this has been added to Sect. 3.1.

l. 236-243: In general, I find Figs. 1 and 2 a bit confusing because the second row of plots doesn't include the cloud feedback (both historical and piNTCF use the present-day SST) whereas the first and third rows do (piControl uses the preindustrial SST). I guess the reader will just have to live with doing a lot mental switching back and forth here. But shouldn't l. 242 then say "greenhouse-gas driven cloud adjustments and cloud feedback"? (And similarly for l. 245?) And shouldn't "greenhouse gas" be replaced by CO2, since NTCF includes the other greenhouse gases (CH4 and O3)?

Reply: We thank the reviewer for these suggestions. The sentence was revised as "greenhouse-gas-driven cloud adjustments and cloud feedback". The term "greenhouse gas" has been kept, however, because the NTCF perturbation in TaiESM1 is narrower than in models with interactive chemistry. TaiESM1 has no interactive chemistry, so CH_4 and ozone follow their prescribed historical concentrations in piNTCF as in historical, and only the anthropogenic aerosol and aerosol-precursor emissions are held at 1850 levels. As a result, piNTCF − piControl contains the forcing from all greenhouse gases (CO_2, CH_4, N_2O and ozone), together with land use and natural forcings, and replacing "greenhouse gas" by CO_2 would understate this contrast.

Regarding cloud feedback, historical and piNTCF are both coupled simulations and do not share the same SST, so the middle rows of Figs. 1 and 2 also contain a temperature-mediated cloud response. A sentence stating that all three rows combine cloud adjustments and cloud feedback has been added at the beginning of Sect. 3.1.

l. 271-281: It would be easier to follow this paragraph if the "aerosol-only" experiment were described a little more clearly, specifically whether this is a coupled or fixed-SST run (maybe Table 1 could have columns specifying which forcers are PI and which are time-varying and whether the ocean is coupled or fixed, and if fixed, which SST is being used).

Reply: The Configuration column of Table 1 now specifies which forcings are held at pre-industrial levels, whether the ocean is coupled or the SST and sea ice are prescribed, and the source of the prescribed SST. The piClim-control and piClim-aer simulations have also been added. Because TaiESM1 has no interactive chemistry, only the anthropogenic aerosol and aerosol-precursor emissions are held at 1850 levels in piNTCF, whereas CO_2, CH_4, N_2O, ozone and land use follow their historical evolution.

AerosolOnly is a fully coupled simulation, and this is now also stated in Sect. 2.1 and in the first paragraph of Sect. 3.2, which introduces AerosolOnly − piControl (revised Fig. 4b) as one of three coupled contrasts.

l. 354-364: The purpose of the least-squares fits in Figs. 6 and 7 appears to be to document the correspondence between the climate responses on the x and y axes and to show that TaiESM lies somewhere close to the line. It looks like the fit is x ~ y rather than (the more usual, though of course there is no real constraint on choice of axes in this case) y ~ x. If the fit were done the other way around, the line in Fig. 6 would probably be close to horizontal and TaiESM would look like an outlier. I say this not to quibble with the paper's conclusions but rather to say that I find the straight-line fit more confusing than illuminating.

Reply: We agree that the regression lines were more confusing than helpful, and they have been removed. Revised Fig. 7 (former Fig. 6) now shows isolines of total ERFaer (ERFari + ERFaci) instead (see also the comment on Fig. 6 below). Revised Fig. 8 (former Fig. 7) now shows the 1:1 line.

l. 362: As noted above, I am having trouble seeing where this statement about IPSL comes from. In Zelinka2023, all IPSL models have ERFaci < 0.

Reply: The reviewer is correct. The positive IPSL value came from our APRP calculation with the coupled historical − piControl difference. Revised Figs. 7 and 8 now use the values in Table 2 of Zelinka et al. (2023), in which IPSL-CM6A-LR has an ERFaci of −0.19 W m^{-2}, and the corresponding statement in Sect. 4.1 has been corrected.

l. 365: TS has not been defined

Reply: Thank you for pointing this out. ΔTs is now defined in Sect. 4.1 and in the caption of revised Fig. 8 as the change in global-mean surface temperature between historical and piControl, averaged over 1985–2014.

l. 387, though I should have noticed in Sec. 2: is the prescribed SST the one from the TaiESM historical run or an observed one? I assume the former, but please state. If the latter, that would make the historical - histsst plots difficult to interpret.

Reply: The SST and sea ice prescribed in histSST are the monthly-mean, time-varying fields from the TaiESM1 historical simulation, as specified by the AerChemMIP protocol (Collins et al., 2017). This is now stated in Sect. 2.1 and Table 1.

All of Sec. 4.2: This section is 3 multipanel figures and two pages of dense text describing many fields that change hardly at all between prescribed and interactive SST. I think this could be made substantially shorter and reader-friendlier by focusing on what's actually different between historical and histsst and then focusing on why the interactive ocean is important for those differences (through fast feedbacks in surface exchange etc.).

Reply: We agree with the reviewer's assessment. Section 4.2 has been shortened, retitled "Effect of interactive air–sea coupling", and now focuses on the differences that coupling makes. It first compares the coupled and fixed-SST aerosol estimates, which differ through the aerosol-induced SST response (revised Fig. 9c), and then shows that coupling for a given SST evolution (historical − histSST) has little effect in the global mean. The original Figs. 9 and 10 have been moved to the Supplement (Figs. S3 and S4), and the third conclusion has been revised.

Figures, especially the many-panel ones: the font is *so* tiny

Fig. 3: would be nice to have the global mean values indicated in each panel

Reply: Thank you. Revised Figs. 1–9 have been redrawn with larger fonts for labels and color bars, and global-mean values are now given in each panel of revised Fig. 4 (former Fig. 3).

Fig. 4b: would be nice to have a narrower color scale here so the spatial patterns are actually visible

Reply: Revised Fig. 5b (former Fig. 4b) uses a narrower color scale for ERFari so that its spatial pattern is visible.

Fig. 6: an isoline of total ERFaer would be quite useful to guide the eye and make the connection to the Delta T_s argument (l. 365), given how different the scales on the axes are; probably more useful than the fit line

Reply: Revised Fig. 7 (former Fig. 6) now shows dashed isolines of ERFaer = ERFari + ERFaci at 0.5 W m^{-2} intervals in place of the regression line. Combined with the ΔTs values in revised Fig. 8, these isolines show that ΔTs largely follows the total ERFaer (Sect. 4.1).