We thank all reviewers for their time in reviewing the manuscript, and appreciate their detailed and insightful comments and suggestions. Here we summarize the comments from reviewers as blue, and include our respective responses as below in black.

Reviewer 3

Review of "Meteorological Drivers of the Low-Cloud Radiative Feedback Pattern Effect and its Uncertainty" by Rachel Yuen Sum Tam and co authors for consideration in Atmospheric Chemistry and Physics.

In this study the authors study the pattern effect of low level marine clouds using a cloud controlling factor decomposition. Experiments AMIP, historical and 4xCO2 are used to estimate how the feedback depend on the time since applied forcing. While this evolution is fairly consistent and independent of radiative kernel or model, it is found that there is substantial inter model spread in how models simulate the response of clouds to atmospheric stability.

Overall, I would say the results are unsurprising, but also I haven't seen this done before so the study makes a valuable addition to the literature. My main concern with the study is that the sea surface temperatures prescribed in the atmosphere-only AMIP experiment is taken to be the true forced pattern, when in fact studies have shown that other datasets more resemble the magnitude of pattern effect simulated in coupled historical and that observations are broadly within the range of model internal variability. Hence, several of the remarks and conclusions regarding possible model biases must be adjusted or removed. When reading I was also mildly concerned with the somewhat limited/narrow selection of cited literature, so I have made some effort to include reading suggestions in the detailed comments below.

Overall, I would say my main concern is one of major weight, but is easily addressed provided the authors agree to do so. Other than this, I think the study is important and should be published.

We thank the reviewer for their constructive comments. We agree with the reviewer that AMIP simulations do not necessarily reflect the forced response, and have modified the manuscript to reflect this. We did not mean to imply that the AMIP simulations reflect the forced response. However (up until issues of errors in SST reconstructions) they do reflect the observed SST pattern. Any estimate of the historical feedback lambda_hist from observations will also contain both forced and unforced components. The utility of AMIP feedback lies in the fact it provides the most apples-to-apples comparison with observational estimates of \lambda_hist.

When placing observational constraints on ECS, we need to correct for the difference between the measured, historical, feedback and the long-term feedback. While most studies usually consider both coupled and AMIP simulations, the best estimates of the pattern effect typically rely on AMIP simulations.

We have made several modifications to the manuscript to address this point. We have included a more detailed description in the introduction of how the pattern effect is quantified and why it is necessary in estimates of climate sensitivity. Specifically we have added a new paragraph to the introduction (paragraph 2) that reads:

Estimates of Equilibrium Climate Sensitivity need to account for the pattern effect when translating the transient net feedback calculated from present day observations into an expected equilibrium feedback (Sherwood et al 2020). ECS can be estimated given knowledge of the forcing and feedback as:

$ECS = F_{2\times}/ \lambda_{eq} = F_{2\times}/(\lambda_{hist} + \Delta\lambda)$, (1)

where $F_{2\times}$ is the radiative forcing associated with a doubling of CO_2, λ_{eq} is the net radiative feedback at equilibrium, λ_{hist} is the radiative feedback over the historical period, and $\Delta\lambda$ is the pattern effect-driven difference between equilibrium and historical feedbacks. The magnitude of the pattern effect, $\Delta\lambda$ is thus usually quantified as the difference between an estimate of the equilibrium feedback drawn from an abrupt4xCO2 simulation, and a historical feedback estimated from either coupled historical simulations or historical simulations with prescribed sea surface temperatures, i.e., AMIP simulations (Andrews et al., 2022). Within models, $\Delta\lambda$ is also often quantified as the difference between the early and late part of an abrupt4xCO2 simulation (Andrews et al., 2015). Observational estimates of ECS then rely on adding a model-derived estimate of $\Delta\lambda$ on top of an observationally-derived estimate of λ_{hist} (Sherwood et al., 2020)

When describing the model simulations, we have added two paragraphs discussing the differences between *AMIP* and *historical*:

The AMIP and historical experiments are both analyzed over the 1982-2008 interval. While they both have forcing constituents consistent with the historical record, they differ in their boundary conditions and active components: AMIP is an atmosphere-only simulation with prescribed sea surface temperature and sea ice concentration variations, and the historical experiment has both ocean and atmosphere components active. While SST trends in coupled models are broadly consistent with observations, they struggle to precisely reproduce trends in historical SST patterns (Dong et al., 2021; Wills et al., 2022). Whether or not the observed patterns are consistent with the magnitude and patterns of natural variability in coupled models depends on the precise metric, interval, and region of focus. Some studies find the observed pattern consistent with modeled variability (Olonscheck et al., 2020; Watanabe et al., 2021), while others find it inconsistent (Wills et al., 2022; Rugenstein et al., 2023a). It is also possible that the discrepancy between AMIP and coupled historical simulations arises from the failure of coupled models to adequately simulate the forced response. A number of mechanisms have been proposed as explaining the observed cooling in the pacific and the failure of models to do so, such as the dynamical thermostat (Clement et al., 1996), cold tongue biases (Seager et al., 2019), aerosols (Heede and Fedorov, 2021), or teleconnections from the southern ocean (Dong et al., 2022; Kang et al., 2023). Yet another possibility is that of errors in the SST reconstruction used in AMIP simulations (Modak and Mauritsen, 2023). The reason for this discrepancy is still an active area of research (e.g. reviews by Rugenstein et al., 2023b; Watanabe et al., 2024).

In this study we focus on the differences in atmospheric response between different simulations, and are therefore agnostic to the root causes of the SST patterns and their discrepancy. Observational estimate of the historical radiative feedback $\lambda hist$ will also contain both forced and unforced components, and thus differences between AMIP simulations and long term feedbacks λeq provide the best model analog for the expected magnitude of the pattern effect (Sherwood et al., 2020; Andrews et al., 2022).

Abstract+Conclusions, the authors describe clearly what was done, but I miss a bit what was learned? For me, a striking result is that the multi model mean kernel yielded feedback changes between simulations which was a bit smaller, but not very different from observational kernels. I think this is important to convey. One might even try to quantify this.

We thank the reviewer for this comment. We have added a paragraph pointing out that while there are large differences in the net feedback between the different models and observational kernels, they yield consistent pattern effects.

Our results suggest that while model estimates are broadly consistent with observations, model-based kernels tend to underestimate the strength of the pattern effect relative to satellite-derived kernels. Considerable spread remains across estimates derived with observational meteorological kernels (Fig. S02), due to the differences in instrument capabilities, cloud detection algorithms and selection of cloud retrievals in each observational dataset (Stubenrauch et al. 2013, Minnis et al. 2011). …Regardless, the inter-model spread in the magnitude of the pattern effect is much less than the spread in the net feedbacks (see comparison between Fig. 3.1a and 3.1b, Supp. Fig. 1a, b, and c).

In addition, there are three other primary results that are now better highlighted in the summary and conclusion section.

First is the fact that the pattern effect is almost entirely dominated by the dR_EIS component. The degree to which regional changes in dR_SST cancel each other out is surprising. This points to better quantification of \partial R/\partial _EIS as *the* key main area of future work for quantifying the atmospheric response to changing SST patterns.

Second, we document regional contributions of EIS to the pattern effect. While the importance of inversion to the pattern effect was documented in the global-mean (e.g. Ceppi and Gregory 2019), we provide the first regional decomposition showing that the Southeast Pacific and Southern Ocean dominate.

Third, we document the relative spread from uncertainty in radiative kernels vs uncertainty in the meteorology (i.e. CCFs), and find that uncertainty in kernels dominate. This is not unexpected - in fact, this assumption underlies much past work using CCFs. However, it is, to our knowledge, the first time this assumption is evaluated.

The introduction is very short. I am particularly missing a paragraph on the mechanisms underlying the patterns, and some discussion of forced vs. internal variability SST patterns. This is important for the study, since the AMIP runs are later assumed to represent a true forced fast response. But some studies indicate that the AMIP SSTs are an outlier. Furthermore the East Pacific has caught up a lot since 2014, when the AMIP experiment ends, suggesting it was partly internal variability, perhaps partially associated with the hiatus period. Some papers to include could be Clement et al. (1996), Liu (1998), Pierrehumbert (2000), England et al. (2014), Kosaka and Xie (2016), Seager et al. (2019), Watanabe et al. (2020), Heede et al. (2020), Hayashi et al. (2020), Heede and Fedorov (2021), but there are probably several others that I have missed.

We thank Reviewer 3 for the suggestion and have expanded the introduction to include a brief discussion of the sources of the delayed warming in the southern ocean and east Pacific (paragraph quoted above). We also added a longer discussion on the differences between AMIP and coupled simulations and possible mechanisms for the observed patterns in Section 2.1 when we introduce the two different historical simulations (see paragraphs quoted above).

13, Here you might also add Rugenstein et al. (2016, 2019, 2020), Li et al. (2012), Knutti and Rugenstein (2015).
17, Some more papers of relevance here are Olonscheck and Rugenstein (2024) and Modak and Mauritsen (2023).
25, I suggest looking at Ceppi and Gregory (2017, 2019) and Hedemann et al. (2022).

We thank the reviewer for providing relevant literature references, and have added them in the respective statements.

44-46, Worth mentioning that Olonscheck et al. (2020) found the observed SST patterns to be broadly consistent with model internal variability. Also observational uncertainty exists in the underlying reconstructed past SSTs, and the dataset used in the standard AMIP experiment is an extreme outlier (Modak and Mauritsen 2023). Not necessarily wrong, but also not necessarily right.

We thank the reviewer for the literature suggestions. We have included these in the extended discussion on the discrepancy between AMIP and coupled simulations and their interpretation. (See paragraph quoted above).

55, Nevertheless, they are no longer the main source of community assessed uncertainty (Sherwood et al. 2020, Forster et al. 2021).
We thank the reviewer for this comment. While we agree with the reviewer that advances in observational data and modeling have reduced cloud feedback uncertainties relative to other mechanisms, the contribution to the total feedback inter-model spread by marine low clouds remains significant (Hill et al. 2025, Ceppi et al. 2024). Therefore, we modified the sentence to "inter-model spread in the total feedback estimate can be largely explained by the spread in marine low cloud feedbacks (Bony and Dufresne, 2005; Zelinka et al., 2020)" in Section 2.2.

83, I would say hours to days.

We agree with the reviewer for the clarification, and have modified the phrase from “days” to “hours to days”.

131, Unclear statement, try to rewrite or consider deleting.

We thank the reviewer for their suggestion and have rephrased the sentence to “We find the model ensemble mean results (red markers in Figure \ref{fig:1}a) have a negative (stabilizing) marine low-cloud feedback in the \textit{AMIP} experiments; \textit{historical} have a near-zero feedback, \textit{4xCO2-fast} have a weakly positive feedback, while \textit{4xCO2-slow} have a slightly more positive feedback.”.

141, Note that the pattern effect estimated from AMIP runs is an outlier, and the average SST reconstruction yields a pattern effect centered on that simulated by historical (Modak and Mauritsen 2023).

We now reference the Modak and Mauritsen paper and cite errors in the AMIP reconstruction as a possible source of discrepancy (see paragraph 2 in section 2.1, quoted above).

161-163, Good question whether it is a bias in models, or in observations, or simply an expression of natural variability, see several references mentioned earlier.
233-234, I am unsure if they should replicate them if it is due to errors in AMIP and/or internal variability.

We agree with the reviewer’s comment and have tightened the language to clarify that, on average, *coupled* simulations (historical) are biased when compared to simulations forced with the actual observed SSTS (*amip).* The paragraph now read:
“Due to these biases in dREIS , the transient low-cloud feedback in the coupled historical simulations is therefore biased towards more positive values compared to the low-feedback obtained when prescribing observed SST patterns (AMIP) simulations. Holding the assumption that the 4xCO2-slow response is representative of the future low-cloud response, using the coupled historical simulation would under-estimate the magnitude of the pattern effect, $\Delta\lambda$.”

Section 3.1.4 One has to do quite a bit of thinking to read out the results presented in this section from Fig. 1 and A1. Would it not be feasible to condense this into an estimate of how much spread comes from kernel and CCF in each case?

We thank the reviewer for their constructive comments. In Table F1, we included inter-model mean and standard deviation of feedback estimates calculated with the different combinations of model/observation/model-mean kernels and model/model-mean meteorology. We can learn the spread source by comparing the standard deviations of global low cloud feedback calculated with $mod*model_{avg}$ (source of inter-model spread is from kernels), and $model_{avg}*mod$ (source of inter-model spread is from meteorology) with *mod*mod* (source of spread comes from both meteorology and kernels). While the ensemble mean estimates are similar across model-only calculation methods across experiments (e.g. *Hist:* -0.01 Wm^{-2} *(mod*model*$_{avg}$*)/* 0.01 Wm^{-2} *(model*$_{avg}$**mod)* / -0.04 Wm^{-2} (*mod*mod)*), spread attributable to model kernels alone (*mod*model*$_{avg}$ SD = 0.46) mirror closely to the spread contributed by both meteorology and kernels (*mod*mod* SD = 0.44); and spread attributable to model meteorology only is much smaller than the other two (*model*$_{avg}$**mod* SD = 0.21).

Meteorology is the only source of spread for feedback estimates calculated with observation kernels, which all have similar SD with mod*model$_{avg}$.

209, Is there any reason this is likely? It is "possible", yes, but likely is usually meant to indicate more than 66 percent probability.

We thank the reviewer for raising this question and agreed that the choice of word is not the most accurate. We therefore edited the term from “likely” to “possible”.

Figure 2, 3 and 4, These maps are too small and distorted, squeezed horizontally.
We agree with the reviewer on their comment about the figures, which is also suggested on a similar note by Reviewer 2, and have made adjustments to Figure 2 and 3 for transposing the figure dimensions, and rearranged Figure 4’s projection and figure aspect ratios.

236-238, I do not think this can be concluded based on the presented evidence.
We agree with the reviewer’s comment and therefore reworded the statement to clarify that, on average, *coupled* simulations (historical) lead are biased when compared to simulations forced with the actual observed SSTS (*amip).*

270-271, Same issue, this cannot be concluded.
We have removed the sentence on the large bias of coupled simulations.