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 all reviewers as blue, and include our respective responses as below in black.

We have modified the introduction and included additional prior literature in response to the reviewers' comments, including adding an additional paragraph in addressing the pattern effects' definition and another paragraph on forced vs. internal variability SST patterns.

We thank the reviewers again for their suggestions and feedback.

Reviewer 1

The marine low-cloud pattern effect is an important and widely studied topic, but one that is far from being well understood. This manuscript analyzes this effect in detail in an ensemble of models using the cloud controlling factors (CCFs) method, and obtains results that are very clear, very interesting, and, in my opinion, represent a very significant advance. While the essential role of EIS has already been well established, the use of meteorological cloud radiative kernels, on the one hand, and the sensitivity of meteorological variables to average temperature, on the other, makes it possible to clearly separate what is related to the SST pattern from what is related to the response to this SST pattern. In my opinion, this is a very good manuscript that fully deserves to be published in ACP. I have only a few minor comments to make, which are presented below.

The difference between estimates using different data sets is mentioned in the manuscript without being really discussed. The manuscript highlights the importance of a good estimate of (\partial R)/(\partial EIS) for the pattern effect. But if we compare Figure 1b with Figures A1, it seems to me that the estimate of (\partial R)/(\partial EIS) differs significantly depending on the data set used. This is a point that could be further emphasized, as well as the importance of having a better estimate of this term based on observations, with possibly a discussion of the strengths and weaknesses of the different datasets and possible avenues for improvement.

We thank Reviewer 1 for the constructive feedback on the sensitivity of feedback on choice of observational datasets. We agree there is significant uncertainty in the different datasets, and have added an additional supplementary figure S2 to visualize the spread due to the differences in estimates of (\partial R) /\partial CCF). The spread of observational kernels is smaller than the spread of model kernels, yet still significant.

Past work with observational kernels (e.g. Myers et al 2021) has generally treated the spread as an indication of uncertainty, with unweighted averages of different kernels taken as a best estimate. While assessing the relative strengths and weaknesses of each dataset in detail is outside the scope of this paper, we have added a summary of key differences between datasets in the summary and conclusion section:

“Considerable spread remains across estimates derived with observational meteorological kernels (Fig. S2), due to the differences in instrument capabilities, cloud detection algorithms and selection of cloud retrievals in each observational dataset (e.g. Stubenrauch et al., 2013; Minnis et al., 2011). For example, ISCCP uses IR-VIS methods to provide cloud properties that correspond to the radiative mean from both high and low clouds and tend to misidentify high clouds that overlay low clouds and return biased-high mid-level cloud amounts. MODIS, CERES-FBCT and PATMOS-x products retrieve high cloud properties using IR methods, but distinct biases remain. Using all four observational kernels therefore provides a comprehensive range of low cloud amounts and the resulting feed-back and pattern effect estimates. However, additional observations will remain pivotal in narrowing the range of radiative sensitivities to meteorology.”

pages 3-4, Eqs 1-3 and corresponding text: adding a subscript i to R_low would make the equations and text clearer.
l 62: dR/dTg => dR_low/dT_g ; \partial R / \partial CCF_i => \partial R_low,i / \partial CCF_i
l 69: d CCF/dT_g => d CCF_i/dT_g

We thank the reviewer for their detailed comments, and have modified the in-text notations as suggested in equation 1-3 for identifying Line 62 and 69.

Reviewer 2

Tam et al. analyse a set of 16 CMIP models to study differences in marine low-cloud radiative feedback as well as the pattern effect contribution. To this end they analyse three different experiments, AMIP, historical and 4xCO2. From their analysis they find the pattern effect to be mostly driven by Southern Hemispheric EIS changes and changes between the models to be due to differences in cloud sensitivity rather than differences in cloud controlling factor changes.

I think this is a very well done analysis tackling an important issue. However, at times I found some arguments and conclusions hard to follow. Extending some of the discussions of the results would help explain the authors reasoning. Also, a deeper discussion of previous results seems necessary to place the current results into context. While these points are important to address, I have no fundamental objections to the analysis and can therefore recommend publication after minor revisions.

We thank the reviewer for the constructive comments. We agree that a better context of prior work and clarification on the novel aspects of the findings, as also suggested by Reviewer 3. We have added more context for the work, as detailed in the response to individual comments below, and in the response to Reviewer 3.

Introduction

Could do with a gentler start (see below my comments under “Pattern effect”). Also the jump to the last paragraph is very abrupt. You didn’t really map out why what you are trying to do is important and novel. Readers that are less familiar with the field will appreciate this.

We thank the reviewer for the detailed suggestion, and agree adding additional transitions and explanations will be helpful for a fuller understanding of the background of this work. We have slightly expanded the discussion of the pattern effect in the first two paragraphs, and added the following two paragraphs to the introduction.

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 λhist (Sherwood et al., 2020)

Climate models exhibit large uncertainty in the magnitude of the pattern effect - an uncertainty so large that it precludes observational estimates on the upper bound of future warming (Sherwood et al., 2020; Armour et al., 2024). Instead, this upper bound is largely constrained by paleoclimate information. However, translating past warming into future warming requires the use of models and is itself sensitive to model-estimates of the pattern effect (Cooper et al., 2024). Thus, understanding the sources of model spread in the pattern effect, and ultimately reducing that spread, is a major roadblock in improving projections of future warming.

I think the introduction would also benefit from expanding the discussion on what exactly this manuscript sets out to do. This is currently only covered in one sentence in the last paragraph, but seems to be covered in more detail in L. 114-117.

We agree with the reviewer on providing a clearer outline in the introduction and therefore have modified L114-117 and moved it to the end of the introduction.

Pattern effect

I think the manuscript would benefit from a deeper introduction to what the pattern effect is (possibly in or before the current first paragraph), to help readers that are slightly less familiar with the topic.

This extends to the discussion about quantifying the pattern effect. You mention in the abstract and in the conclusions (L.252) that “The pattern effect is defined as the difference in feedbacks between transient and long-term simulations.”. However nowhere in the introduction is this definition mentioned or discussed. A more in depth discussion on this definition and explanation of the use should be given in the introduction or method section.

We thank the reviewer for their suggestion, and have expanded the discussion of the pattern effect in the first two paragraphs, and included the significance of quantifying the pattern effect and the method of doing so.

Building on the previous paragraph, in 3.1.1/ 3.1.2 you argue that the change from AMIP->historical->4xCO2_fast->4xCO2 in the model mean as well as in individual models is a measure of the pattern effect. I think this needs more discussion/justification as it is at the core of your analysis of the pattern effect. I agree with your analysis, given that the sensitivities used for the calculations are the same throughout the experiments, the changes have to come from changes in the CCFs pattern. However, this is quite some leap for readers to make, especially for those less familiar with the field. I think your reasoning should be discussed explicitly, why the differences in the experiments are an indication of the pattern effect, maybe even refer to eq. 1 for quantitative explanation.
L. 147 In line with all the above comments, this sentence is very vague.

We agree with Reviewer 2 and have expanded the introduction. We incorporated to paragraph 2 the discussion of how the pattern effect impacts estimates of climate sensitivity and why it is estimated as the difference in net radiative feedback between transient simulations (AMIP/historical/4xCO2_fast) and equilibrium simulations (4XCO2_slow).

The distinction between AMIP and historical is discussed in more detail in section 2.1

The authors should incorporate more discussion on previous literature. For example, one of the main points, that the inter-model spread in feedback is mostly caused by differences in cloud sensitivity rather than differences in CCF changes is a somewhat know result (e.g. Klein et al. 2017). The same is true to an extent for the EIS changes. The authors analysis is very thorough and has (to my knowledge) not been done in this way before, but there are several related studies and it should be clarified what findings are agreeing (or disagreeing) with previous results and what findings are novel.

We thank the reviewer for the detailed comment. We agree that while several aspects of the results are not unexpected, many have not been quantified before. We have incorporated additional discussion of past literature, and tried to more clearly delineat our contributions.

In Section 3.1.2 we have modified the second paragraph to read:

"Our kernel-derived estimates are thus consistent with past literature on the pattern effect, which suggests the low-cloud feedback evolves to be less negative over time (Andrews et al., 2015, 2018, 2022; Myers et al., 2023), and changes in EIS are an important component to this evolution (Ceppi and Gregory, 2019). The kernel approach allows us to quantify which cloud controlling factor drive the pattern effect. We find that the pattern effect is driven almost entirely by changes in the EIS component feedback."

The second paragraph of Section 3.1.4 has been modified to read:

"Overall, models have less disagreement on CCF responses to warming than the radiative flux sensitivities to CCF changes. The CCF decomposition shows that the vast majority of the uncertainty in both the net marine low cloud feedback and the pattern effect comes from uncertainty in how marine low clouds respond to their local environment (i.e. model spread in kernels, Fig. 1c). By comparison, the model uncertainty in how meteorology changes with warming is much smaller (i.e model spread in CCF changes, Fig. 1b), with the notable exception of the historical experiment. These results hold if the CERES-FBCT observational kernels are replaced with either the ensemble mean kernels or other observational kernels (Fig. A1). In terms of specific CCFs, the largest sources of uncertainty are the sensitivities of clouds to SST and EIS, with smaller contributions from RH and WS, and negligible contributions from Tadv and ω. The fact that models have less uncertainty in the response of CCF to warming has been an underlying assumption of the approach since its inception (Klein et al., 2017; Brient and Schneider, 2016; Qu et al., 2015; Myers and Norris, 2016). However, this is, to our knowledge, the first time the relative uncertainties have been quantified."

As a side note, given that the authors have done all this work, an extension of the analysis to provide insight into what drives the differences across experiments in the models seems an interesting, novel and important question. I think this would be a not too time consuming additional analysis, but I leave it up to the authors if they want to pursue this idea.

We have incorporated a discussion of what drives differences in SSTs across experiments, both in the introduction and in section 2.1.
"GCM experiments forced by abruptly quadrupling CO2 show that warming is initially delayed in certain regions, most notably the eastern tropical Pacific and the Southern Ocean (e.g. Andrews et al., 2015; Heede and Fedorov, 2021)."

"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)."

Minor

L. 12-14: Less negative over time under which conditions? 4xCO2, historical, 1%CO2 increase. All of those?

We thank the reviewer for raising this question. Out of the literature we cited, Senior and Mitchell (2000) and Andrews and Ringer 2014 shows results in agreement with the statement using 1% per year CO2 increase experiments, while Andrews et al. 2012, Li et al. 2013, Andrews et al. 2015, Proistosescu and Huybers 2017 and Rugenstein 2016 used abrupt-4xCO2 experiments; and Armour 2017 has showed feedback changing over time using both 1ptCO2 increase and abrupt-4xCO2 experiments. Knutti and Rugenstein 2015, Rugenstein et al. 2019 and 2020 used experiments that have varying forcing levels additional to the abrupt-4xCO2 experiments; and Held et al. 2010 has consistent results using an instantaneous doubling of

CO_2 experiment. We therefore modified the sentence in the manuscript as "where the feedback becomes less negative over time after a forcing, such as a quadrupling of carbon dioxide concentration, is imposed, leading to an increase in climate sensitivity."

L. 21: This needs a source in my opinion, also as far as I am aware not all models show this behaviour.

We thank the reviewer for raising this issue and have added Andrews et al. 2015 and Heede and Fedorov (2021) in support of the statement.

L.28: Again a source seems needed here

We thank the reviewer for the suggestion. The literature cited in L. 25 discusses both mechanisms driven in the deep convection region and the region of descents, and we realize the confusion that readers have when placing the citation in between the two sentences. We therefore have relocated the citation to the end of the paragraph.

L. 67 In my mind, sensitivities are partial derivatives, since it tells me how x changes with y, all else being equal. Personally I wouldn't use the word here. However, I don't think there is a fixed definition, so this is just a suggestion.

We thank the reviewer for raising this issue, and have modified the phrase "*sensitivities* of CCFs to global surface temperature change " to "*magnitude* of CCFs to global surface temperature change".

L. 77 Are all these studies on these six specifically?

We thank the reviewer for raising this question. Not all of the cited literature are on all six cloud controlling factors, but for clarity, we have modified the sentence to "Prior work has documented in-depth how CCFs impact marine boundary layer cloudiness, covering all 6 CCFs using theory, models, and observations (Myers and Norris, 2015; Scott et al., 2020; Cesana and Del Genio, 2021; Bretherton, 2015), or focusing on specific CCFs (Lilly, 1968; Cesana and Del Genio, 2021)." for the separation of those work that studied a couple CCFs', and those that investigated all 6 CCFs as in this manuscript.

L. 81 While I think I know what the authors mean by "local large scale environment", I find this expression a bit self-contradictory

We thank the author for pointing out this confusion, and have therefore modified the phrase from "local large scale environment" to "local perturbations from the large-scale environment".

L. 111 I wouldn't call epsilon an error term. It doesn't come from measurement errors or similar things, but from inherent covariances in the system, as the authors point out themselves.

We thank the reviewer for their suggestion, and have reworded the term as “covariance term”.

L. 114-117 This information seems to belong (at least in parts) into the introduction rather than the method section
L. 117 Why is there a citation here? Aren’t those the goals you are stating?

We agree with the reviewer, and have therefore moved these sentences to the introduction. The citation in L117 was to point readers to the literature that studied the model biases in SSTs as referred to in the sentence. We therefore rephrased our statement as “(3) how coupled model biases in the evolution of historical SSTs lead to biases in the feedbacks as discussed in Andrews et al. (2022).”.

L. 168 I assume ‘pattern’ refers here to the relative importance of the terms of Fig. 1b compared to the relative importance in the subfigures of Fig. A1? In that case I would use a different word, as it might be misleading.
L. 173 as in L.168

We agree with the reviewer and have now changed the word “pattern” to “behavior” to avoid confusion with the same word as in the pattern effect.

L. 179 or a covariance term (epsilon). Was this checked for in strength in this analysis, similar to what was done for the pattern (Fig. 4).
L. 178-184 I would have expected you to do the decomposition according to eq.3, but you replaced the term using model average sensitivity with observed sensitivity in figure b. I understand that you probably wanted to “recycle” figure 1b and not create an almost identical figure again, but I still found it initially confusing. Also, it makes the 1b and 1c not apples to apples comparable, since the spread will be larger/smaller if the sensitivities are larger/smaller. I fully acknowledge that the result seems pretty robust in terms of where the spread is coming from, but it still initially confused me.
L. 191 Following up from the above comment (on L. 178-184), I see you address this here by pointing to A1d (btw, the d is missing). Why not cite this in the paragraph above and adjust the discussion accordingly to avoid confusion?

We thank the reviewer for pointing out these issues. We agree the phrase “Following Equation 3…” was confusing readers that a residual decomposition on the inter-model variance will be discussed in Figure 1. In the three panels in Figure 1, we aim to provide a qualitative overview of 1) the total inter-model spread contributed by both model kernels and model meteorology (Figure 1a); 2) how the spread contributed by observational kernels compares with those by model kernels (comparison between Fig. 1a and Fig. 1b; choosing CERES kernels as the representative kernel while MODIS/ISCCP/PATMOS-X are shown in Fig. A1), and 3) how the spread contributed by model meteorology only compares with those in the inter-model spread contributed by both model kernels and meteorology (comparison between Fig. 1a and Fig. 1c).

The proximate similarity between Fig. 1a and Fig. 1c illustrates that most of the feedback spread is attributable to the sensitivity of radiative fluxes to meteorological changes. To illustrate that models agree well on model meteorology, while also providing low-cloud feedback estimates using observational kernels in the main figure (Figure 1b), we replace multi-model mean kernels with an observational kernel, showing the intermodel-spread in feedback only contributed by CCFs. Table F1 also provides the inter-model mean and standard deviation across the 16 GCMs calculated using methods shown in each subplot in Figures 1 and A1.

We have not calculated the covariance in the global feedback estimate analysis as we think it is clearer to show the covariance in a spatial map rather than a single global value.

We agree that Fig. 1b and 1c are not comparable, and understand Figure 1 may seem confusing. We structured Section 3.1 to follow the panels in Figure 1, with Section 3.1.1 and 3.1.2 on Figure 1a, Section 3.1.3 on Figure 1b, and Section 3.1.4 on the comparison of Figure 1a with Figure 1b and Figure 1a with 1c to discuss the sources of inter-model spread coming from kernels or meteorology. Therefore we have added a statement in the manuscript that guides readers to read from panel to panel in Figure 1, and refer to Figure A1 for the results using multi-model mean kernel and individual model meteorology and the other observational kernels, which all shows the inter-model spread of meteorology.

L. 221 I would put ‘in historical and 4xCO2’ instead of ‘across experiments’. I stumbled upon this part because I stopped reading and looked at the maps and found there to be some significant difference in AMIP to the other experiments (Fig. 3a compared to 3b and 3c). That threw me off until I continued reading to see your caveat in the next sentence.
We thank the reviewer for pointing out this issue. The sentence is now rephrased as “across coupled experiments” to point out specifically historical and 4xCO2-fast experiments as the relevant experiments.

L. 236.-238. I don’t follow how you arrive at this conclusion from the before presented results. Against what is the historical estimate biased high? Against the AMIP? And why is it biased high? Also, why is the pattern effect biased low? I think this part needs significant expansion to explain your reasoning.
We have tightened the language to clarify that the *coupled* simulations (historical) are biased when compared to simulations forced with the actual observed SSTs (*amip).*
We have modified the paragraph to 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, $\Delta\lambda$.”
While this SST pattern driven bias is known, we show here that it is due almost entirely due to dREIS, rather than the direct impact of SST patterns (dRSST). We make this new result explicit in the conclusions:

“Surprisingly, we find that large regional changes in the direct impact of SSTs on low clouds, dR_{SST} cancel each other out in the global mean. Thus, the time-evolving SST patterns impact the radiative feedback indirectly, by altering atmospheric circulation and EIS. ”

L. 239 Maybe add a transition sentence here to help the reader understand what you now intend to have a look at.
We thank the reviewer for the suggestion, and have now modified the sentence to “Following Equation 3, Figure 4 maps inter-model spread in regional feedback estimates.” to point readers that we will now look into the inter-model spread of feedback spatially.

Technical
L. 51 refer to
L. 62 R-> R_low (twice)
L. 63 being a bit picky, but i is the index, the individual CCF would be CCF_i L. 69 CCF_i
L. 89 calculating -> calculated
L. 120 would cut sub-component
L. 207 to->do
We thank the reviewer for the detailed suggestions, and have accepted all above suggestions in line 51, 62 and 63 (as also suggested by Reviewer 1), 89, 120, and 207.

Figures
Fig.1: exp-> Exp.
Fig.1: Maybe put a dot between the two terms in the titles and remove the parenthesis. Also, you could consider shortening model->m and CERES-> C to make things more readable
Fig.1: Personally, I would remove the bold font in the caption.
Fig.2/3: The maps are relatively small and the label for the zonal mean plots almost not readable in the printed version. For the digital version this is of course not a problem. The authors could consider rearranging the plots to make them bigger.
Fig. 4: I would suggest to call the last column epsilon, to stay consistent with eq. 3. Fig. 4: Again, personally I would remove the bold font in the caption.
We thank the reviewer for their constructive suggestions. We have modified the figures for higher clarity as suggested in Figures 1 and 4 and rearranged Figures 2 and 3 as also suggested by Reviewer 3. However, we find it easier to identify which model/observational we are referring to in each panel, and therefore keep the model/observational dataset subscripts in these figures as is.

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 λ_{hist} will also contain both forced and unforced components, and thus differences between AMIP simulations and long term feedbacks

λ_{eq} 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.