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 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:

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

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, Δλ 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, Δλ 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 Δλ 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 CO2 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.