Author's response to "Tipping points in ocean and atmosphere circulations"

We thank Axel Kleidon for handling our manuscript and the reviewers for their feedback to improve our manuscript, which we have revised accordingly. Below are the original comments from reviewers in black and our corresponding responses in blue.

Sina Loriani for the author team

Reviewer 1

To recap, my main suggestions in the previous review stage were:

• 1. go deeper into mechanisms of circulation "tipping" and thus go beyond previous reviews; include East Asian monsoon

• 2. clearer and more objective method and criteria of selection of cases, and labeling of uncertainties

In general, the authors have implemented some changes in the direction of these suggestions. I however also see the need to go a bit further, and have some open questions that I will address below. I do not think that the article has to become substantially longer, but it should be rephrased accordingly; and/or some parts (especially impacts, which are not part of the confidence assessment) may also be shortened.

We'd like to thank you for reviewing this – rather extensive – article again, and for the constructive feedback! Especially the reflections regarding more transparency on definitions and criteria were helpful to make sure that the article provides a solid reference for a wide readership. Following your suggestions, we have amended the introduction to provide a clearer definition of "tipping", and to describe the methodology that the confidence statements are based on more transparently. Furthermore, we have added a dedicated section on East Asian monsoon. Point-by-point replies to your comments can be found below.

I also urge the authors to make more transparent what has changed between versions, and what not. I also notice that the "original" (black text) in the difference file does not agree with the previous version I reviewed. This makes it very hard to see what has changed. Often, the information in the replies was misleading (changes the authors claim to have made are not visible in the pdf I got, and pieces of text were hard to find because neither line numbers or quotes were provided).

Thank you for pointing this out! Indeed there seems to have been an unfortunate mixup on our side (some changes that indeed were not colored/implemented in the version of the submitted manuscript), apologies for that. And, at the same time, apparently there was an issue on the ESD-server side (?), since some of the changes you remarked as not being implemented *are* in fact addressed in the track-changes-file we uploaded egusphere-2023-2589-ATC1.pdf. In any case, apologies for the arisen inconveniences – we double checked that the now resubmitted file highlights all changes. All the highlighted differences are with respect to the *first* submitted manuscript.

General

1. I still suggest to discuss the EASM in its own (sub)section (see below).

We have now added a dedicated section on East Asian monsoon.

2. The confidence levels are now more explicitly explained, but still a bit vague. Are conceptual hypotheses "evidence", and how do the authors weigh results from conceptual models versus ESMs? How many models, and models of what sort, are required to show "tipping", for which forcing, in order to go from a "+" to "++"? I understand that this will always remain subjective to some extent, but I believe that some more precision is possible here. The resulting confidence assessment should be the clear result of pre-defined criteria; otherwise the criteria can appear to be "back engineered" to the subjective opinion of the authors of the (not peer reviewed) "Tipping Point Report".

We have added a dedicated subsection in the introduction providing more context to our criteria. In short, we anchor our assessment on the question "*is there a plausible positive feedback to drive self-perpetuating change beyond a forcing threshold*" and review the amount and agreement of different lines of evidence to support or refute the existence of such a feedback. This links to the discussion of the tipping point definition (see your comment and our response below).

We here approach each system with the hypothesis of such thresholded feedback dynamics (often guided by physical process understanding and conceptual or simplified models). Our confidence to classify a system as a tipping system increases with the amount and agreement of high-quality evidence from the different sources (models, paleo records, observations). Due to the complexity of the problem, at least with present-day literature, it is very challenging to make a generalised link between a confidence statement and a number of models to exhibit tipping dynamics. First, one needs to know whether these models do resolve the key feedbacks (i.e. *can* the suspected tipping process be studied with the models in the first place?) and second, generalised statements need a sufficient number of such models to participate in a model intercomparison designed to answer tipping-related questions. The emerging generation of CMIP7-type models, as well as initiatives like the Tipping Points Modelling Intercomparison Project (TIPMIP) and related endeavours are expected to provide more solid grounds for future assessments that could attempt to relate confidence statements to some quantified metric, like fraction of models exhibiting tipping dynamics in comparable settings. In the meantime, we rely on expert judgement classifying the consistency of evidence across different types of models and different experimental settings.

3. The authors replied to my previous points and often elaborate on some research results in their replies, but often left the actual manuscript unchanged. The readers of the article may benefit more if some of these replies make it into the paper. For example, can we expect some systems to be more prone to tipping than others? Are clouds not too fast and variable to be an Earth system tipping component? How does resolution affect tipping behaviour? What needs to happen to reduce uncertainties?
For example, the authors replied: "We agree that more theoretical insights from hydrodynamic theory such as critical transitions in turbulence on multiple scales (van Kan and Alexakis, 2020; van Kan, 2024) or transitions between multiple- stable states in aquaplanets (Brunetti et al., 2019; Ragon et al., 2022) can inform a better understanding of more applied research on climate tipping points going forward. We will add a corresponding discussion and literature references to the revised manuscript." However, I do not find such a

discussion in the revised paper, and do not find citation of the papers they mention. Brunetti et al. and Ragon et al. both appear in the reference list without being cited in the text, the others do not appear.
More paragraphs that the authors claim to have revised, show no changes. Examples of unchanged(?) text:
• My previous review: "Also apparently unchanged: Sentence is unclear. What does “these” refer to?" Reply: "Meant were small-scale processes, added." I see no change.
• "Although the AMOC does not collapse in this model, it seems unlikely that it will recover its former strength on human timescales.” Why does this seem unlikely? What’s the evidence? And what are “human timescales”? Authors: "This has been rewritten." I see no change.
• I suggest the authors go through my previous review and highlight the actual changes made.

We now transferred some of the responses to the main document where it seemed fit well without expanding the text too much; mainly in the discussion section (marked in blue). On the lack of change in the previously revised manuscript, please see our comment above.

Abstract
• nothing seems to have changed? I suggest to make clearer what the approach is in this paper, in the direction of my main suggestions above.
• specifically: "we classify the West African monsoon as a tipping system." sounds like you don’t consider any other system. Clarifying the method and criteria of the classification would help, and make clear that highlighting the West African monsoon is the result of this method?
• "modified wind patterns...disrupt established circulation patterns" sounds redundant - what are drivers of such shifts?
• The authors write in their reply: "The sentence aims to highlight the complex interplay of factors affecting monsoon stability. While the AMOC is not a direct driver of monsoons, research suggests its variability can influence large-scale atmospheric circulation patterns that, in turn, impact monsoon systems. We have now rewritten the sentence to reflect this more clearly." In the files I received, nothing has been changed in the abstract. The sentence is still there. In general, it would help if the authors quoted from the new version in their replies, referencing line numbers, to substantiate claims about what they have changed.

We have reworded the abstract to make clear that the reported classifications are based on the evidence found in the available literature and resolved the ambiguities you indicated above.

Introduction
• The authors added more text on the definition of "tipping"; but some of my detailed comments remain, e.g. "tipping" in general, or due to human activity in the future? Why focus on circulations, what are open questions there? How was literature collected? More precise assessment criteria.

Thank you, this is indeed important to clarify. We are analysing systems that have been fairly stable under preindustrial/Holocene conditions, and where it is plausible that they change under human activity. The guiding question is therefore: *Which components of the Earth system, with focus on circulation systems, feature strong feedbacks that could lead to*

self-sustained change after crossing a forcing threshold? In different words, “which systems *could* in principle tip?” We’d like to highlight (in agreement with your comment below) that this is a different question from “which systems *will* tip”, which is subject to current/future forcing trajectories (and uncertainties thereof) and the location of the forcing thresholds (and uncertainties thereof). The confidence statements in this review refer to the former, i.e. *could* the systems tip.

This work focuses on circulation systems as part of a special issue. It is therefore complementary to analogous assessments for systems in the cryosphere and biosphere, and provides a detailed deep-dive into the selected systems that goes beyond recent reviews addressing systems from all domains (Wang et al., 2023; Armstrong McKay et al., 2022). Circulations in the ocean and atmosphere critically connect different parts of the Earth system and hence play a vital role e.g. in potential tipping cascades (Wunderling et al., 2023).

For this narrative review, literature was collected via expert elicitation among the group of authors. We describe the confidence assessment framework above, and in the newly added dedicated subsection in the introduction.

• Definition of tipping. The authors adopt a definition that involves positive feedbacks, and even "self sustained" feedbacks. I believe that is OK as it is made transparent, but please resolve these potential issues:
• 1. There are other definitions around. Specifically, it would make sense to also note that the IPCC actually gives a different (less specific) definition; see AR6 glossary. The part "often abruptly and/or irreversibly" is actually a quote from IPCC that should be referenced. It seems that the authors decide to use a narrower definition than IPCC because positive feedbacks are essential here; this should be mentioned.
• 2. Do all examples really adhere to your definition? Coral bleeching: Why does it crucially involve self sustained positive feedbacks? As far as I know, corals die when temperature becomes too high, without the need of runaway feedbacks? Same goes for permafrost (the global permafrost - CO2 feedback is far too weak for self-sustained feedbacks).
• 3. What exactly do you mean with "self-sustained feedbacks"? If a forcing is faster than the system’s response (a typical situation in climate change), would not any system change involve "self-sustained feedbacks" as long as the system tracks the changing equilibrium, even if the system is perfectly linear?! Fig. 4 could be an example of such a situation. The problem of such a definition based on feedbacks is also that you need to understand well why a transition occurs in a model, rather than just observing that it does. Uncertainty about mechanisms hence leads to lower confidence, even though the phenomenon may be apparent.

The revised introduction has a subsection now on the tipping points definition we are adhering to, which is based on f*eedbacks strong enough to sustain change after crossing a forcing threshold* (i.e. irrespective of continued forcing).

We argue that our definition is compatible with the one from IPCC, where a tipping point is defined as *the critical threshold beyond which a system reorganizes, often abruptly and/or irreversibly* (4.7.2 in Lee et al., 2021 / IPCC AR6 WG 1 report Chapter 4). However in the IPCC, criteria for the system’s reorganisation are not defined, rather irreversibility and

abruptness are used as proxies for tipping, leading to sometimes inconsistent results such as ocean heat content being listed in the tipping elements table 4.10 of that chapter despite a threshold-free behaviour. These and similar arguments (e.g. on the ambiguity of defining *abruptness* in this context) have been made in Armstrong McKay et al. (2022), and we have included a corresponding remark in the introduction.

By extending the IPCC definition with the criterion of *feedbacks sustaining change beyond a threshold*, we require some understanding of why/how the system reorganises after crossing the (tipping-point-defining) critical threshold. Arguably, rather than constituting a problem, such a definition provides a generalisable yet consistent approach to tipping points, and can work across a wide range of systems as recently demonstrated in the Global Tipping Points Report (2023) co-authored by more than 200 experts in the various domains (including assessments of coral reefs and permafrost with the same approach – in both cases *local* rather than global feedbacks are assumed to drive tipping). On your last point; we hope that the extended tipping point definition in the introduction (and references therein) explain better now what we mean with feedbacks that sustain change after a critical forcing threshold.

• 4. It makes a difference in confidence whether a system can in principle show tipping behaviour, e.g. has shown such behaviour in the past, versus whether this system will show "tipping" as response to plausible human forcing in the future. The kind of evidence you cite and the confidence levels seem to imply the former ("system features tipping dynamics", e.g. paleo transitions), but the context of the article and the definition you adopt from McKay et al., Lenton et al., ... follows the latter. For example, if there is evidence for AMOC tipping during ice ages, what does a "++" assessment mean for the possibility of human-induced AMOC tipping? This should somehow affect the assessment, or the interpretation of the assessment.
• The different confidence levels all seem to repeat the same phrase; no need for such repetition!

As described above, we here focus on which systems *could* in principle tip (more formal: Is there a feedback that could lead to self-sustained change after a forcing threshold) and not on whether these systems *will* tip (more formal: Will this threshold be crossed). We would kindly disagree that adopting the definition from Armstrong McKay et al (2022) and related literature implies that we consider the latter. In the context of our article, “++” means that there is medium confidence in the potential for the AMOC to tip, with uncertainties in potential timing (related to threshold position), magnitude or feedback strength.

Sect. 2

• line 182: "Here we assume that all three states (Alley et al., 1999; Rahmstorf et al. 2002) are stable equilibria" Why? Is it necessary to make such an assumption in a review paper?

Agreed that this wording was misleading, it was rather a comment on the underlying conceptual three-mode model. We have rephrased this.

• SPG: Why medium confidence, the same as the AMOC? I get the impression that some ESMs do show AMOC "collapse" in some way, but are there such ESM results suggesting a

separate collapse of the SPG (independent of AMOC)? Is Fig. 6 showing "mixed-layer depth" supposed to show such an SPG collapse? Does the gyre (horizontal circulation) actually collapse?
• Same question about the Southern hemisphere. There seems to be less model evidence and agreement; yet all three ocean systems get a "++" in the confidence assessment. I guess the issue here lies in the criteria that are still a bit vague, see above.

On the SPG: This system is linked but independent from AMOC. E.g., Sgubin et al. (2017) show that there are CMIP5 models where the SPG convection collapses both with and without an accompanying AMOC collapse. We made small amendments to make clear that the SPG is driven by both wind shear and buoyancy forcing – the latter being subject to a potential *SPG convection collapse* that would reinforce (and vice versa) an overall *SPG weakening* (horizontal circulation slowdown), see also Sgubin et al. (2017) and Swingedouw et al. (2021). Mixed layer depth is an indicator for the convection. We added a corresponding statement.

On the Southern Ocean circulation: As described above, approaching each system with a conceptual understanding of a potential tipping mechanism, model evidence needs to be scrutinised with respect to the capacity of the model to resolve this mechanism in the first place. While this process understanding has long been limited for the Antarctic Overturning Circulation (Purich and England, 2023), recent work demonstrated that under adequate modelling conditions (e.g. representation of deep water formation over the continental shelf), models do show a collapse akin to potential AMOC tipping (Li et al., 2023, Zhou et al., 2023), supported by paleo evidence (Huang et al., 2020 and other references cited in the manuscript) thereby adhering to the ++ criterion in our assessment.

On a more fine-grained scale, these three systems could arguably be separated more subtly but here for all three of them we argue that there is ample evidence to classify them beyond + and too little to classify them as +++. We hope that our refined assessment criteria presentation in the introduction convincingly supports this.

Sect. 3

I do not understand the argument why the East Asian summer monsoon should not be addressed in its own subchapter. The authors now mention this system, but only as an afterthought in Sect. 3.4 (Summary), although it did not appear in the sections that are supposed to be summarised here. The authors replied that the EASM is not a "tropical" monsoon, but they already also consider ENSO, and mid-latitude dynamics, and the ocean circulation in their paper. Given that the focus of the review is a "wide range of circulation systems" in general, and that paleo evidence suggests abrupt shifts in the East Asian summer monsoon, I don’t see why this particular circulation system should not have its own (at least small) section.

We have added a dedicated section on East Asian monsoon.