Review of manuscript egusphere-2024-2170

'***Uncertainty quantification for overshoots of tipping thresholds***' by Lux-Gottschalk & Ritchie

Summary

The study by Lux-Gottschalk & Ritchie explores how parametric uncertainty introduces and impacts uncertainty in the response of tipping elements to overshoot trajectories, which temporarily exceed the critical tipping threshold.

First, the impact of uncertainties in the locating of the critical threshold and the linear restoring rate on the response of tipping elements to overshoot trajectories of varying overshoot magnitudes and durations is illustrated based on a simple dynamical system with a fold bifurcation (Sect. 2.1 and Sect. 2.2). In particular, the tipping element response is described in terms of probabilities of tipping, based on a modification of a previously derived inverse square law (Ritchie et al. 2019). Therein, the uncertainty in the location of the critical threshold leads to a larger uncertainty in the tipping response to overshoot trajectories than the linear restoring rate. Constraining the uncertain parameter can reduce uncertainties in this tipping response. In addition, the effect of changes in the mean vs the range of the distribution of the uncertain parameter is explored. In a next step (Sect. 2.3), this framework is extended to a box model of the Atlantic Meridional Overturning Circulation with a similar bifurcation structure and an uncertain parameter (diffusive timescale) that concurrently translates into an uncertain location of the critical freshwater flux threshold and the linear restoring rate.

General comments

The paper approaches the relevant and scientific interesting question of uncertainty in the tipping response to overshoot trajectories in the presence of uncertain parameters from a theoretical / conceptual side. The results could perhaps serve as a starting point for further (probabilistic) assessments of potential critical transitions in the Earth system.

The introduction contains relevant information on tipping and the concept of 'safe' overshoots. In some places, it seems repetitive or related information is spread across the introduction. Some restructuring of the introduction might help to avoid these repetitions and it could also support making the motivation for and main aims of this study clearer. For example, tipping is introduced and defined in L19-25, and some additional general information is added in L52-54. Impacts of tipping are mentioned in L26, and in L45-46 the potential consequences of AMOC tipping are described. Uncertainties are first referred to in L32-40, and laid out in more detail in L77-82.

While the approach of explaining the basic interplay between the overshoot trajectory and the location of the critical threshold, the linear restoring rate or the diffusive timescale (in the AMOC box model) in the beginning of each section before introducing uncertainty in these parameters is a good starting point for understanding the key figures displaying tipping probabilities, it is not always easy to follow the results presented in this study. This is for two main reasons:

1) The clarity of the language could be improved throughout the manuscript. In some places, the language distracts from the key messages of the paper. These key messages might be formulated more concisely. Please also see the specific comments below.
2) The methods are not given in the main manuscript but described in the Appendix. To allow the reader to properly follow the results, the models and the derivation of the tipping probability based on the (modified) inverse square law and numerical simulations should be described before presenting the impact of uncertainty in the location of the critical threshold and the linear restoring rate on the tipping response. In addition, the Bayesian Inference to constrain uncertain model parameters should be explained in such a section.

The manuscript illustrates how parametric uncertainties translate into uncertainties in the tipping response based on conceptual models, and I acknowledge that a lot can be learnt from these approaches, e.g. to showcase that constraining parametric uncertainty is important to constrain the tipping response to overshoot trajectories. It could be interesting to expand the discussion of if / to what extent / how this framework and the conceptual results could be applied to more complex situations.

I have included more specific comments, questions and suggestions below.

Specific comments

L01: Maybe replace ‘Earth’ by ‘Earth system’ or ‘climate system’. In addition, can these subsystems be considered to currently be in a stable (steady) state?

L03: Maybe replace ‘common mechanism of tipping’ by ‘possible mechanism of tipping’.

L05: Maybe replace ‘is possible’ by ‘could be possible’.

L07: Please give references in brackets here and elsewhere if not integrated into the sentence, i.e. (Ritchie et al., 2019).

L07: What is meant by ‘other system features’? Please specify.

L08: I would like to suggest to add ‘in response to a temporary overshoot’ (or similar) after ‘probability of tipping’.

L14-15: This statement appears as a key result / conclusion of the analysis in the abstract. I was wondering whether this is the case, e.g. when reading L166-168, and also what it implies for conclusions on mitigation pathways (as indicated in the previous sentence).

L16-17: To me, the sentence suggests that many parameters in the box model have been tested and the diffusive timescale has been identified as particularly relevant for determining uncertainty in the AMOC tipping behaviour. I would like to suggest to rephrase to: ‘In addition, we illustrate how constraining the highly uncertain diffusive timescale within this box model reduces the tipping uncertainty of the AMOC in response to overshoot scenarios’ (or similar).

L20: ‘sudden’ could suggest that tipping occurs abrupt in time, which may not necessarily be the case. Please consider to reformulate.

L22: ‘kicked off their current stable state’ could refer noise- and rate-induced tipping, but not necessarily bifurcation-induced tipping, which is associated with the loss of stability beyond a critical threshold in some control parameter, followed by a critical transition to a remaining stable state. Please consider to reformulate.

L28: Maybe replace ‘are assumed’ by ‘have been suggested’ (or similar).

L29-31: Maybe replace ‘can still be avoided’ by ‘could potentially be avoided’. I would also like to suggest to explicitly acknowledge that this statement is based on model simulations.

L55-72: I am not sure how helpful the description of the different mathematical mechanisms of tipping is, in particular given that the relation to overshoots is not directly described. Please consider to connect the different mechanisms to overshoot trajectories. This may also help to justify the focus on bifurcation-induced tipping in this paper as stated in L73.

L88-89: Please check the section numbering in the Appendix. In addition, I would also like to suggest to move (parts of) the Appendix into a section that focuses on the Methods in the main part of the manuscript (see general comments). This might be helpful for readers that are not familiar with the fold bifurcation and the AMOC box model, and would allow to remove some of rather detailed information of the AMOC box model in the introduction (L86-89). This also applies to the probabilistic extension of Ritchie et al. (2019) and the inverse square law, as its development is described as a main aim of this paper (L90).

L91: Maybe replace ‘fold tipping scenario’ by ‘tipping via a fold bifurcation’ (or similar).

L119: Why is the location of the critical threshold the ‘biggest uncertainty to consider’? Is this already a result of the analysis, based on the argument that follows this statement, or motivated by previous studies?

L126: Maybe replace ‘large’ by ‘far’.

L130: ‘early’ could be associated with a time dimension. Please consider changing ‘early’ to ‘low’ (or similar).

L131: I am not sure about the formulation ‘otherwise identical system’. The bifurcation diagrams also differ in e.g. the location of the basin boundary in addition to the location of the critical threshold.

L133: Maybe replace ‘a threshold further away’ by ‘a higher threshold’ (or similar).

L136: I would like to suggest to use a different terminology for cases without tipping. In the manuscript, formulations such as ‘tipping can be avoided’ are often used when there is no critical transition due to a high tipping threshold (or similar; see e.g. L158-159, L177, L185-186, L307-308, L323). This formulation could be misunderstood.

L137-138: I am not sure whether it is clear how a tipping probability is assigned here. Please see the related comments on including a section on the methods in the main part of the paper.

L139: While it may be clear to some readers, it might still be helpful to add a short explanation for the choice of this range for the location of the critical threshold.

L140: I would like to suggest to indicate the quantities on the axis of Figure 1(c) (that is, the time spent over the lowest possible threshold and the peak in external forcing) in Figure 1(a). In addition, maybe replace ‘based on’ by ‘depending on’.

L142-149: Please indicate the probability ranges (in terms of the shading and the black lines) in Figure 1(c) as an additional legend, as in Figure 8(b).

Figure 1(c): Is there a reason why the colouring stops at tipping probabilities of 50% and larger, in contrast to Figure 8(b) and given that for a known threshold (without uncertainties) there would only be one line separating *tipping* (100% probability of tipping) from *non-tipping* (see also e.g. L310-312), if I understand correctly.

L150-155: It could be helpful to describe more explicitly in the beginning of the description of the results how *uncertainty* in the response to overshoot trajectories and a low / high *tipping* *probability* relate. Only at the end of Sect. 2 (L334-335) a higher uncertainty is linked to a tipping probability moving closer to 50%.

L151: Please consider rephrasing to ‘peak forcing not even overshooting the threshold’ (or similar) instead of ‘profile not even overshooting the threshold’.

L152-153: I am not sure about the structure of this sentence. Please check.

L161: Please specify ‘distribution’ here and elsewhere (e.g. L209). I assume that in this case it is referred to the ‘distribution of the location of the critical threshold’.

L166-168: I was wondering whether it might be possible to align the (description of the) probabilities with the language used in the IPCC.

L171: ‘however’ could be removed. Maybe replace ‘factor’ by ‘parameter’ or ‘system characteristic’ (or similar).

L171-172: I would suggest to decide for one expression here. In addition, it might be helpful for the reader to give a short explanation of the strength of the linear restoring rate. Is this different or the same for both stable equilibria or is this a general system property (as indicated by the e.g. expressions ‘weaker system’, ‘strong system’; L182-187)? This could also be included in a dedicated section describing the methods in the main part of the manuscript.

L176-177: I am not sure about the formulation ‘identical systems’, as the systems also differ in e.g. the location of the basin boundary. How does the basin boundary relate to the restoring rate?

L179: Maybe replace ‘boundary for the basin of attraction’ by ‘boundary of the basin of attraction’.

L181: Which stable state do you refer to here (i.e. the upper stable state or the lower stable state or both)?

L183-184: What is meant by ‘curvature of the fold’? I am not sure whether this becomes clear without a brief explanation.

L184: What does ‘earlier’ refer to here? How is time involved?

L198: Please consider rephrasing to: ‘the tipping uncertainty is substantially smaller’ (or similar). Otherwise the sentence could suggest that an uncertain restoring force is effective in ‘reducing uncertainties’.

L205: I think the comma after ‘tipping’ is not needed here.

L206: Please consider rephrasing to: ‘An overshoot trajectory that sits...’ (or similar).

L208-209: This sentence is unclear to me. What is meant by ‘the separation between the 1% and 50% curves’?

L210: Please specify ‘uncertainty’ here and elsewhere. It should be clear whether it is referred to uncertainty in tipping or uncertainty in e.g. the location of the critical threshold.

L210: Maybe I have misunderstood this sentence, but to me, it looks as if the alternative distribution covers a different range of the restoring force proportionality factor (3.5-6.5) than its initial distribution in Figure 4(a).

L211: What is meant by the ‘width of the banding’? Do I understand correctly that this refers to the ‘distance’ between the 1% and the 50% probability curves? In general, it might be helpful for the reader to introduce and use a clear terminology when describing Figure 4(b) (and similar figures). So far, different notions have been used throughout the manuscript (e.g. uncertainty in tipping behavior, probabilistic critical boundaries, uncertainty in the tipping probability, uncertainty in the probability of an overshoot avoiding tipping).

L212-213: This sentence is not complete. Please rephrase.

L215-217: These sentences are not clear to me. If I understood correctly, a clear boundary that separates tipping from non-tipping exists for a known restoring rate without uncertainties, and would be given by the inverse square law. What does ‘(uncertainty of) the location of critical boundary (separating tipping from not tipping)’ refer to? How can there be ‘little change in the width’ and a decrease in the ‘uncertainty of the location of the critical boundary’ at the same time? Please also see the related comment for L211.

L231: Please consider rephrasing to: ‘The chosen ranges of the advective and diffusive timescales correspond to...’ (or similar).

L235-259: This is an extensive description of critical freshwater fluxes depending on the advective and diffusive timescales. I would suggest to shorten these paragraphs, if possible, and focus on the aspects that are directly relevant for the overshoot scenarios.

L238: I would like to suggest to add the missing bistability as the main reason (if I understood correctly), which then means that no critical freshwater flux exists.

L257-259: These sentences are unclear to me. What is meant by ‘the upper critical threshold moves later’? Does the second sentence refer to the lower critical freshwater flux threshold?

L261: Please add a short justification for a stabilization of the freshwater flux at 0.25 Sv.

L263: I am not sure whether a small diffusive timescale should be referred to as an ‘event’. I would like to suggest to rephrase to ‘In the case of a small diffusive timescale, the tipping threshold...’ (or similar).

L263: ‘late’ (and ‘early’) could be associated with time. Please check the formulation here and elsewhere.

L266: Why is the exact number of 0.2 Sv given here? If I understand Figure 6(b) correctly, the AMOC would recover for a reduction in the freshwater flux to below approx. 0.22 Sv. Please also add for which value of the diffusive timescale this threshold is given.

L272: Maybe it could be referred to the figure that shows the prior distribution of the diffusive timescale.

L273: Maybe replace ‘recovery of the AMOC on state’ by ‘recovery to the AMOC on state’.

L278-283: What is meant by ‘initialising the simulations in equilibrium’ in this context? I think some more explanation may be helpful for the reader here to understand how this acts as an additional ‘source of error’. Does ‘error’ refer to the discrepancy between numerics and theory? Please specify. In addition, how is the applicability of the framework presented here to more complex models effected by the discrepancy between numerics and theory?

L288: Maybe change to ‘substantial’ or ‘strong’ instead of ‘dramatic’.

L289-290: Does ‘previously’ refer to the prior distribution? This sentence is not clear to me. This also applies to ‘previously’ (and ‘now’) in L294.

L290: How is the ‘stabilisation level’ defined? I assume that it refers to the forcing level after the overshoot, but a clear definition of this term is missing (or maybe I have missed it).

L293: I am not sure if ‘guaranteed’ fits well here.

L294-295: It might be helpful to show the >99% probability of tipping (see previous related comment on the probability levels that are shown in the figures).

L332: The comma after ‘distribution’ may not be needed.

L332-335: If possible, it might be helpful to indicate this example in Figure 8(b).

L335-336: Can this also be inferred from a specific figure?

L338: Maybe replace uncertainty on model parameters' by 'uncertainty of/in model parameters'.

L340-354: In the summary of the role of the location of the critical threshold and the linear restoring rate for the system response to overshoot trajectories, the impact of uncertainties in these parameters on the tipping response could be addressed in more detail (as this is the focus of the paper).

L343: Maybe reformulate and replace 'further away' by 'high' (or similar).

L346: Something may be missing after 'especially the time over'. Please check.

L352: Maybe I misunderstood something, but I got confused by 'quasi-static'. Do the black lines in e.g. Figure 6(b) correspond to equilibrium states or quasi-equilibrium states?

L360-362: I would suggest to combine these two sentences into one sentence, e.g. 'Constraining parameter uncertainty, for instance by performing Bayesian inference on observational data can greatly reduce...' (or similar).

L362-363: Here, it could be helpful to discuss the applicability of the (probabilistic) inverse square law relationship in more complex models (i.e. beyond the conceptual models used here) in more detail.

L366-367: While the concept is interesting, I am not sure whether the introduction of an 'overshoot budget' is needed here. It is not used beyond this paragraph, and even in this paragraph it is not entirely clear to me how it relates to the main part of the manuscript.

L374-375: This sentence is not complete. Please also consider putting this statement (and a similar statement in L367-369) into context – it might still be favourable to avoid overshooting critical thresholds at all, if possible, instead of 'favouring large and short overshoots as opposed to small and long overshoots'.

L375-377: Is the need to constrain system uncertainties in the case of rate-induced tipping specifically related to overshoot scenarios or does it apply in general?

L381: I am not sure whether a 'planned overshoot' fits well as a formulation. Please consider rephrasing.

L383: ‘AMOC tipping is very unlikely’ is a strong statement, based on the AMOC box model. While limitations are discussed in the following sentences, I would still like to suggest to reformulate and clearly indicate that this statement refers to the box model.

L389: Maybe add ‘in response to possible overshoot trajectories’ at the end of the sentence.

L389-390: ‘elements tipping’?