Response to Reviewer 1

June 3, 2026

TOAD v1.0: A Python Framework for Detecting Abrupt Shifts and Coherent Spatial Domains in Earth-System Data

Jakob Harteg, Lukas Röhrich, Kobe De Maeyer, Julius Garbe, Boris Sakschewski, Ann Kristin Klose, Jonathan F. Donges, Ricarda Winkelmann, and Sina Loriani

Submitted to GMD (EGUsphere) on 22 Jan 2026

Introduction

We thank the Reviewer for their review of our manuscript and their comments. We will address each point in detail below. Our responses follow the order of the original comments. For clarity, reviewer comments are shown in shaded boxes, followed immediately by our response.

Comment 0	The paper by Harteg et al “A Python framework for detecting abrupt shifts and coherent spatial domains in Earth-System data” offers a software package that combines previously published techniques. I have several concerns about this manuscript and its software.
Response	We would like to respectfully disagree with the characterisation of TOAD as merely combining previously published techniques. While TOAD does integrate established methods for shift detection and clustering as interchangeable components (by design, since modularity is a core feature of the framework), the aggregation stage introduces a novel spacetime consensus clustering algorithm specifically developed for this work. We further note that in response to reviewer comments, we have substantially revised this aggregation algorithm: the updated method operates directly on three-dimensional cluster label fields without collapsing the temporal dimension, counting per-voxel member support across ensemble inputs with configurable spatial and temporal tolerances. This makes the aggregation both methodologically simpler to interpret and more powerful, as temporal information is fully preserved in the consensus output. The remaining contributions of TOAD, including the unified spatio-temporal embedding with equal-area regridding, the modular pipeline architecture, and the integrated ensemble synthesis workflow, together constitute a methodological contribution that we believe aligns well with the scope of GMD as a methods-oriented journal. The specific concerns raised by the reviewer are addressed in the comments below.

Comment 1	Since there is no research novelty in the paper, I think it is more suitable for a software magazine, such as The Journal of Open Source Software https://joss.theoj.org/. If the authors would like to target specifically [the] geophysical community, the choice of the experiments should be different. There should be extensive demonstration of performance on artificial data, with clear explanation of the difference between a generic tipping and abrupt change points. In particular, abrupt changes cannot be detected using early warning indicators, which means that the proposed TOAD package should be compared with equivalent packages of change point detection, on the same datasets. See the R package by Killick https://cran.r-project.org/web/packages/changepoint/index.html package by James et al https://cran.r-project.org/web/packages/ecp/index.html spectral change point package in Python https://github.com/Lucew/changepoynt and others. A new software is a welcome addition when it performs the same or better compared with already existing packages, and the paper should demonstrate this on the same datasets.
Response	We thank the reviewer for this detailed comment. Regarding the requested comparison with changepoint detection packages: ASDETECT, the default shift detection method in TOAD, is an established peer-reviewed method (Boulton and Lenton, 2019), and that paper already includes a direct comparison with the changepoint R package on multiple benchmark time series. We consider re-benchmarking ASDETECT against alternative detectors to be outside the scope of this work: using ASDETECT within TOAD's pipeline is analogous to using an established numerical solver within a larger analysis framework, where the validation of the component belongs to its own dedicated publication. The manuscript already directs readers to Boulton and Lenton (2019) for this purpose. We will, however, expand our discussion of ASDETECT's assumptions and characteristics to better guide users in deciding whether to substitute an alternative detector for their specific application, which TOAD's modular architecture explicitly supports. We agree that additional controlled validation of the full pipeline would be beneficial. As outlined in our response to Reviewer 2, we will expand the synthetic experiments to include more realistic scenarios such as autocorrelated noise and background trends. We note that "early warning indicators" such as increasing temporal autocorrelation reflective of critical slowing down are a distinct class of methods from changepoint detection packages, and are not part of TOAD at this stage. Furthermore, the distinction between abrupt shifts and tipping points is already explicitly discussed in the manuscript (Section 1, lines 62-69), and we are confident that both of these points are sufficiently clear to the reader.

Comment 2	The authors consider several modelled datasets. Can they add a 2D change point of an observed dataset? In particular, it would be interesting to know how the technique performs on observed short datasets that are currently available.
Response	We thank the reviewer for this suggestion. TOAD is designed to work on any gridded dataset, including observational data, and we agree that such an application would be an interesting demonstration. The present manuscript focuses on model output in the context of TIPMIP, for which TOAD was developed, and we have chosen to limit the demonstrations to this context to keep the paper focused. A dedicated observational application is a natural direction for follow-up work.

Comment 3	The blocks “how technique works” should be prepared in the format of pseudocode.
Response	We thank the reviewer for this suggestion. We note that Boxes 1, 2, and 3 already present the core algorithms as numbered, step-by-step descriptions; a format that serves the same purpose as pseudocode while remaining accessible to readers without a formal computer science background. We therefore do not plan to reformat these sections, but we will ensure that the boxes are clearly signposted in the text so readers can easily locate them.

Comment 4	Section 3.2 should be called “Antarctic ice sheet model data”. In this section, it is not clear how model years were generated. Similarly, amend the title of section 3.3 to state it clearly that it is a modelled example.
Response	We thank the reviewer for this suggestion. We will amend the section titles to more clearly indicate that these are modelled examples. We note that the Antarctic Ice Sheet section already describes in detail how the simulation was constructed (from 0 to ~14°C at a rate of 10^{-4} °C yr^{-1}) and that the independent axis represents GMST rather than calendar years.

Comment 5	In section 3.3., the number of parameters is rather large, and like this tuning detection of changes may become rather arbitrary – what tool can be introduced for optimised choice of parameter values?

Response	We thank the reviewer for this important point. The manuscript already discusses parameter choices and their effects throughout Section 2, including qualitative guidance on threshold selection (Section 2.3.2) and the role of the temporal weighting parameter γ. In the revised manuscript, we will add a structured overview table of all key parameters, including their default values, recommended ranges, and qualitative description of their effects on the results. We note that parameter sensitivity is an intentional feature of the framework rather than a limitation: exploring parameter space is part of the analysis workflow, and structures that persist across parameter choices can be identified as robust features via consensus clustering, as discussed in Section 4.

Comment 6	[In] the caption of figure 5, the authors mention that only negative shifts were considered. If positive shifts are excluded, it would be interesting to know the number of those – how the model performs in both [directions] is indicative of its accuracy.
Response	We thank the reviewer for this comment. The restriction to negative shifts in this figure was chosen for clarity of visualisation, as displaying all detected shifts simultaneously would lead to overcrowded and less interpretable figures. The detection method (ASDETECT) itself is symmetric with respect to positive and negative shifts, and both are included in the analysis. In the revised manuscript, we will clarify this point and provide additional information on the distribution of detected shifts in both directions (e.g. summary statistics), to give a more complete picture of the results without compromising readability of the figures.

Comment 7	I did not attempt to install and run the package due to the lack of time, but I had a look at the GitHub. I note that there are other TOAD packages on GitHub (and this toad package is not easy to find, as it is not indexed by Google): https://github.com/batrachianai/toad https://github.com/gianwario/TOAD https://github.com/amphibian-dev/toad I am not sure if the acronym is good, especially as it misses the keyword “event”. Reconsider the acronym?
Response	We thank the reviewer for this comment. We agree that discoverability and naming

	are important considerations for software packages. Since submission, we have improved the visibility of the repository and will ensure that it is clearly referenced in the manuscript to facilitate access. In particular, the package is published under ***tipmip-methods/toad*** on github such that we expect that interested users from the (tipping points) modelling and intercomparison project community will find it via TIPMIP. Regarding the acronym, we acknowledge that similar names exist in other contexts. However, TOAD (Tipping and Other Abrupt events Detector) has already been adopted within the relevant research community, and we therefore prefer to retain the name to ensure continuity.

Closing Remarks

We thank the reviewer for their comments and suggestions. We believe that our revisions will improve the clarity, positioning, and practical usefulness of the manuscript.

Sincerely,

Jakob Harteg
Potsdam Institute for Climate Impact Research, Germany
jakob.harteg@pik-potsdam.de
on behalf of all co-authors