

Response to Reviewer 2

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 careful review of our manuscript and their constructive 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 1	As it stands, I have concerns about whether the paper fits within the scope of GMD. I could see it fitting within the new methods category, but such contributions require demonstrating equivalent or superior performance relative to existing approaches. The examples, while illustrative, do not constitute a rigorous evaluation of TOAD's performance. Additionally, some methodological choices warrant justification: the detection method and choices of parameters are not sufficiently justified, and the sensitivity of results to these parameters remains untested. A number of more specific concerns are raised below.
Response	<p>We thank the reviewer for expressing their concerns. We would like to make a few general comments here.</p> <p>The manuscript is intended as a methods paper introducing TOAD as a modular and reproducible analysis framework, with the examples serving to demonstrate its capabilities across different Earth system components. We believe this contribution aligns well with the scope of GMD as a methods-oriented journal focused on the development and application of geoscientific analysis tools. The framework has already attracted interest from the research community following its presentation at recent conferences and its publication on GitHub, suggesting that it addresses a genuine gap in available tools for ensemble-based tipping-point analysis.</p> <p>To better align with the expectations for GMD, we will strengthen the methodological validation and practical guidance in the revised manuscript. In particular, we will (i) expand the synthetic test cases to include more realistic scenarios (see Comment 2), (ii) provide clearer guidance on parameter choices and their effects (see Comment 4), (iii) include a qualitative comparison with an existing approach (see Comment 6), and (iv) introduce a substantially revised aggregation algorithm that</p>

Response to Reviewer 2

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

	operates directly on three-dimensional cluster label fields, preserving temporal information and simplifying interpretation (see Comment 5).
--	--

Specific comments:

Comment 2	Method validation: The authors are upfront that the examples are proof-of-concept rather than comprehensive analyses, which is appropriate for a methods paper. However, beyond the synthetic case (Section 3.1), there is no quantitative evaluation of detection accuracy or false positive/negative rates under controlled conditions. The synthetic test is limited to a simple two-region sigmoid shift on a white-noise background, with a large magnitude change, which is far simpler than real Earth-system data. The authors mention a forthcoming benchmark suite (Röhrich et al., forthcoming) but this does not substitute for at least some basic performance assessment in the present paper. I recommend that the authors expand the synthetic data to include more realistic scenarios (e.g., autocorrelation, background trends, multiple shifts).
Response	We fully agree with this and will extend the synthetic tests to include more scenarios that go beyond the simple two-region sigmoid shift.

Comment 3	Choice of default detection algorithm: The choice of ASDTECT as the default shift detection method is surprising given the rich literature on changepoint detection methods rooted in time series analysis. Methods such as PELT (Killick et al., 2012), EnvCpt (Beaulieu and Killick, 2018), strucchange (Zeileis et al., 2002) or the BFAST framework (Verbesselt et al., 2010) just to name a few, offer statistically principled approaches to changepoint detection with well-characterised false positive/negative rates. While the authors justify the choice of ASDTECT on the grounds of speed, robustness to noise, and not requiring a reference time series, these advantages are not systematically demonstrated relative to existing detection methods. Given that TOAD is explicitly designed for use in high-stakes intercomparison efforts such as TIPMIP, the choice of default detection algorithm deserves stronger justification. The authors should either provide a comparative evaluation of ASDTECT against at least one theoretically grounded alternative, or more explicitly acknowledge that the choice of detection method is a consequential one that users should carefully consider for their specific application — particularly since the modular architecture already supports the substitution of alternative detection methods.
Response	We thank the reviewer for this important point. We agree that the choice of detection algorithm is consequential. Rather than providing a formal comparative

Response to Reviewer 2

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

	<p>benchmark, which we consider outside the scope of this work, as argued in our response to Reviewer 1, we will provide clearer practical guidance on when ASDTECT is and is not appropriate. Specifically, we will expand the discussion to explicitly characterise the types of shifts ASDTECT is designed to detect (abrupt changes in the mean state, such as step-like transitions) and the types it is not suited to (changes in variance, frequency, or recurrence structure, as in oscillatory or chaotic systems). This guidance, combined with the parameter table added in response to Comment 4, will help users make an informed choice about whether to use ASDTECT or substitute an alternative detector via TOAD's modular architecture.</p>
--	---

Comment 4	<p>Parameter sensitivity: TOAD's outputs are acknowledged to be sensitive to shift thresholds, consensus thresholds, temporal weighting γ, and clustering parameters. The paper provides limited quantitative guidance on how to choose these parameters in practice beyond the recommended threshold of $dts \geq 0.5$ for ASDTECT. For instance, how should γ be chosen for different systems? The demonstrations use values of γ ranging from 1.0 to 3.0 across the four cases with limited justification. A more systematic discussion — perhaps a sensitivity analysis for one of the demonstration cases — would significantly improve the practical utility of the paper and help users avoid arbitrary parameter choices.</p>
Response	<p>We thank the reviewer for this helpful suggestion. We agree that additional guidance on parameter selection would improve the practical utility of the paper. In the revised manuscript, we will add a structured overview of the key parameters, including shift thresholds, temporal weighting γ, and clustering parameters, covering their default values, recommended ranges, and qualitative effects on the results. We will also include a sensitivity analysis for one of the demonstration cases to illustrate how parameter choices influence the resulting patterns.</p>

Comment 5	<p>Temporal information loss in consensus clustering: The authors acknowledge that collapsing the temporal dimension before computing co-association means consensus clusters identify spatial agreement without requiring temporal synchronisation. It means that two grid cells could be assigned to the same consensus cluster even if their detected shifts occur decades apart, potentially merging physically unrelated transitions. The authors note this in Section 2.4 but the practical consequences are not illustrated. It would be helpful to show, for at least one demonstration case, the distribution of shift times within consensus clusters to illustrate when this is or is not a concern.</p>
------------------	--

Response to Reviewer 2

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

Response	<p>We thank the reviewer for calling out this important point. In response to this concern, we have replaced the aggregation algorithm entirely. Instead of collapsing the temporal dimension before computing co-association, the updated method implements a per-voxel member-support procedure that operates directly on the native three-dimensional (time, y, x) cluster label fields: for each detected event voxel, it counts how many ensemble inputs detect an abrupt shift within a configurable spatial and temporal tolerance neighbourhood, and retains voxels that meet a user-specified support threshold. After thresholding, consensus clusters are identified through connected components. Two grid cells can thus only contribute to the same consensus cluster if their detected shifts are co-located in both space and time. The concern raised by the reviewer is resolved by design in the updated algorithm.</p> <p>In the revised manuscript, Section 2.4 and Box 3 will be updated to describe the new method in full, including its parameters and interpretation. The North Atlantic Subpolar Gyre demonstration will be updated to use the new aggregation method, illustrating the time-resolved consensus output and shift-time statistics per cluster.</p>
-----------------	--

Comment 6	<p>Comparison with existing approaches: The paper situates TOAD relative to edge-detection approaches (Bathiany et al., 2020; Terpstra et al., 2025) but there is no direct comparison of TOAD's output with this approach on the same dataset. Even a qualitative comparison for one of the demonstrations would strengthen the paper's claims about TOAD's added value.</p>
Response	<p>We thank the reviewer for this suggestion. We will add a qualitative comparison between TOAD's output and the edge-detection approach of Terpstra et al. (2025) on a comparable example.</p>

Technical comments:

Comment 7	<ol style="list-style-type: none">1. Line 65: "the default method in TOAD v1.0 being ASDETECT" — consider rephrasing as "the default method in TOAD v1.0 is ASDETECT"2. Line 305: "the data is restricted" — should be "the data are restricted."3. Line 365: "oscillatory or quasi-periodic or chaotic systems" — the repeated "or" is awkward; consider "oscillatory, quasi-periodic, or chaotic systems."
------------------	---

Response to Reviewer 2

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

Response

We thank the reviewer for these corrections. All suggested changes will be implemented in the revised manuscript.

Closing Remarks

We believe that our revisions will substantially strengthen the manuscript and improve its clarity, methodological grounding, and practical usefulness.

Sincerely,

Jakob Harteg

Potsdam Institute for Climate Impact Research, Germany

jakob.harteg@pik-potsdam.de

on behalf of all co-authors