https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

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

Jakob Harteg[1,2], Lukas Röhrich[1], Kobe De Maeyer[3], Julius Garbe[1], Boris Sakschewski[1], Ann Kristin Klose[1,4], Jonathan F. Donges[1,4,5], Ricarda Winkelmann[1,2,4], and Sina Loriani[1]

[1]Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A 31, 14473 Potsdam, Germany
[2]Institute for Physics and Astronomy, University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Germany
[3]Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, The Netherlands
[4]Department Integrative Earth System Science, Max Planck Institute of Geoanthropology, Kahlaische Strasse 10, 07745 Jena, Germany
[5]Stockholm Resilience Centre, Stockholm University, Albanovägen 28, SE-106 91 Stockholm, Sweden

Correspondence: Jakob Harteg (jakob.harteg@pik-potsdam.de) and Sina Loriani (sina.loriani@pik-potsdam.de)

Abstract. Large-scale, nonlinear, abrupt, and potentially irreversible transitions in major Earth-system components are becoming increasingly likely under human pressures, with far-reaching consequences for ecosystems, climate stability, and human societies. Yet detecting and comparing such transitions across Earth System Model ensembles remains fragmented and inconsistent, hindering systematic assessment of tipping-point risks.

Here we present the first release of the Tipping and Other Abrupt events Detector (TOAD v1.0), an open-source, user-oriented Python framework for detecting abrupt changes in gridded Earth-system data. TOAD implements a modular three-stage pipeline consisting of 1) grid-level abrupt shift detection, 2) spatio-temporal clustering of co-occurring changes, and 3) consensus synthesis to identify statistically robust regions across ensemble members, variables, models, or methodological configurations and quantifies agreement in transition timing. The framework addresses key practical challenges of large-scale spatio-temporal clustering on geographic grids and provides diagnostic statistics and visualisation tools. Detection, clustering, and synthesis algorithms can be flexibly exchanged, supporting systematic method comparison and extensibility. TOAD functions as a data-introspection tool that reveals potentially tipping-relevant dynamics across spatial and temporal scales for subsequent, process-based analysis.

We apply TOAD to a synthetic benchmark, domain models of the Antarctic Ice Sheet and the global terrestrial biosphere, and a global Earth System Model ensemble of the North Atlantic Subpolar Gyre. Together, these demonstrations illustrate TOAD's applicability across diverse systems and establish a structured foundation for investigating where and when potentially tipping-relevant changes occur and for quantifying associated uncertainties, supporting coordinated assessment efforts such as the Tipping Points Modelling Intercomparison Project (TIPMIP).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

1 Introduction

Major Earth-system components—such as the Greenland and Antarctic ice sheets, the Atlantic Meridional Overturning Circulation, the Amazon rainforest, and boreal forests—can undergo large-scale, often abrupt, and potentially irreversible changes once critical thresholds are exceeded, often with grave consequences for nature and human societies (Lenton et al., 2008; Armstrong McKay et al., 2022). Such systems are commonly referred to as *tipping elements*, and the thresholds at which they transition as *tipping points*. These transitions are typically driven by positive feedbacks that amplify initial perturbations and lead to self-sustained dynamics (Levermann et al., 2012). Together with other destabilising drivers such as deforestation (Lovejoy and Nobre, 2018) and pollution (Bollasina et al., 2011), continued global warming increases the risk of crossing such tipping points (Lee et al., 2021). Current climate and energy policies imply a median warming of approximately 2.7 °C above pre-industrial levels by 2100 (Hausfather, 2025), placing the climate system within the risk range of multiple tipping elements (Armstrong McKay et al., 2022).

Despite broad agreement that tipping risks are increasing under anthropogenic forcing, large uncertainties remain regarding the critical thresholds, timescales, interactions, and impacts of these dynamics, largely due to observational gaps, incomplete process understanding, and structural differences among models (Lee et al., 2021). Addressing these uncertainties has been identified as a key research priority for the upcoming generation of coordinated climate model experiments, including within the Coupled Model Intercomparison Project (CMIP), and has motivated the Tipping Points Model Intercomparison Project (TIPMIP)—the first multi-model intercomparison effort explicitly focused on tipping dynamics through targeted experiments across Earth-system components and model classes (Dunne et al., 2024; Winkelmann et al., 2025). This creates a clear need for dedicated analysis tools capable of detecting and comparing abrupt changes across heterogeneous model ensembles.

Here, we present the first public release of the *Tipping and Other Abrupt events Detector* (TOAD v1.0): a modular, open-source Python framework for systematic detection and analysis of abrupt changes in gridded Earth-system data. TOAD provides a unified, reproducible pipeline for grid-level abrupt shift detection, spatio-temporal clustering, and ensemble synthesis. While developed in the context of TIPMIP, it is generally applicable to any multi-model efforts such as CMIP or on individual, gridded datasets. The framework features a fully modular, plug-and-play architecture for both shift detection and clustering, allowing users to flexibly select or substitute methods at each stage of the pipeline. It incorporates equal-area regridding to reduce polar bias, a unified (x, y, z, t) Euclidean embedding for consistent spatiotemporal distance calculations across different grid types, and consensus clustering based on sparse co-association matrices to identify robust spatial regions of agreement across models, variables, ensemble members, and parameter choices. For each consensus cluster, TOAD offers statistics such as the mean time (or forcing level) at which the shift occurs, and its spread across ensemble members or models. By tuning timescales, sensitivities, or clustering settings, users can investigate complementary perspectives on their data.

TOAD is designed to support a set of core scientific questions central to tipping-point research. It identifies *where* abrupt changes emerge by delineating spatially coherent domains of co-occurring shifts, and *when* these transitions occur along a forcing trajectory. When the independent axis represents, or can be mapped to, an external forcing variable (e.g. global mean temperature), the spread in shift timing across ensemble members provides a first-order estimate of the forcing range over

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

which such transitions manifest, thereby enabling a first quantification of *uncertainty*. TOAD can also be extended to detect *reversibility* under reversed forcing phases, thereby supporting systematic hysteresis assessments. More broadly, TOAD outputs provide a structured basis for downstream analyses: cluster masks supply spatial constraints for testing early-warning indicators (Dakos et al., 2024) and diagnosing feedback processes in other variables, help reveal tipping dynamics at both large and small scales, facilitate transparent comparison of tipping behaviour across models and scenarios, and can inform impact-focused studies by linking identified tipping regions with socio-ecological or sectoral datasets. By enabling the systematic detection of abrupt shifts across different variables and regions, TOAD outputs also provide a foundation for analysing interactions and potential cascade pathways between tipping elements (e.g. Högner et al., 2025). In this way, TOAD aims to provide not only a detection tool, but a foundation for a broader, integrated assessment of tipping dynamics and their consequences.

To avoid ambiguity, we clarify TOAD's scope. TOAD is a data-driven framework for detecting abrupt shifts, not a diagnostic tool for identifying tipping points in the dynamical-systems sense (i.e., analysis of bifurcations, loss of stability, or irreversibility). "Abrupt shifts" are operationally defined by the chosen detection algorithm—the default method in TOAD v1.0 being ASDETECT (Boulton and Lenton, 2019), which identifies anomalously large gradients in time or forcing space. Throughout this paper, "abrupt" therefore refers to changes that are rapid compared to background variability at the grid-cell scale, and abrupt shifts may occur on very different timescales depending on the system under consideration. Although tipping points motivate TOAD's development, determining whether a detected shift constitutes a genuine tipping process requires additional system-specific analysis.

Against this background, several approaches have been developed to detect abrupt or nonlinear changes in spatio-temporal climate data. Early efforts involved manual classification of abrupt changes in CMIP5 output by Drijfhout et al. (2015), which provided foundational evidence for broader synthesis studies such as Armstrong McKay et al. (2022). More recently, semi-automatic methods based on Canny edge-detection adapted to climate data have been applied to CMIP5 (Bathiany et al., 2020) and CMIP6 (Terpstra et al., 2025) ensembles. Additionally, system-specific criteria-based procedures have been introduced, for example, statistical catalogues of abrupt shifts in CMIP6 (Angevaare and Drijfhout, 2025).

TOAD complements these efforts by providing an integrated, modular pipeline that seamlessly incorporates existing methods for detecting (e.g. ASDETECT (Boulton and Lenton, 2019)) and clustering (e.g. DBSCAN (Campello et al., 2013)) abrupt changes, enabling reproducible synthesis of abrupt shifts across model ensembles. TOAD is designed from the ground up with ease of use and extensibility in mind, offering an end-user-focused framework that is both intuitive to apply and flexible to extend. In this way, TOAD aims to provide a reusable foundation for abrupt shift detection and multi-model synthesis, empowering users to tackle emerging research questions efficiently and transparently.

The remainder of the paper is organised as follows: Sect. 2 outlines the TOAD pipeline and its core algorithms; Sect. 3 demonstrates TOAD on a synthetic dataset with known abrupt shifts, and then on three different systems: the Antarctic Ice Sheet, global vegetation carbon dynamics, and ocean mixed-layer depth in a multi-member Earth system model ensemble; Sect. 4 discusses limitations and provides an outlook on future developments; and Sect. 5 summarises TOAD and its role in the context of TIPMIP and other ensemble-based assessments of abrupt and nonlinear Earth-system dynamics.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

2 Methods

2.1 Architecture

TOAD implements a three-stage analysis pipeline that (1) detects abrupt shifts at the grid-cell level, (2) performs spatio-temporal clustering to identify regions of dynamically coherent change, and (3) aggregates these across models, variables, ensemble members, or method parameters. The main parts of the pipeline are illustrated in Figure 1, while a full flowchart is provided in Supplementary Figure B2.

TOAD is built on *xarray* (Hoyer and Hamman, 2017) for seamless integration with standard Earth-system data formats and supports both regular and irregular latitude–longitude grids, as well as Cartesian grids. TOAD automatically detects grid type and adapts preprocessing and clustering routines accordingly (more details in Sect. 2.3). It accepts either `xarray.Dataset` objects or paths to NetCDF files, with input data structured as a 3D array with two spatial dimensions and one temporal dimension (future work may include support for three spatial dimensions). Importantly, the temporal dimension can also be used to represent a forcing variable or bifurcation parameter, allowing TOAD to be applied to a wide range of systems.

The framework follows an object-oriented design centred on the main TOAD class, which serves as the user interface for all operations. Each analysis stage automatically adds new variables to the dataset and preserves metadata through *xarray* attributes. Custom shift detection and clustering methods can be integrated by subclassing the corresponding module, enabling domain-specific extensions and method comparison within a consistent workflow.

These design choices make TOAD a flexible and reproducible platform for detecting and clustering abrupt changes in the Earth system. In the following, we describe each stage of the pipeline.

2.2 Shift detection

The first stage of TOAD's workflow identifies abrupt changes independently at each grid cell. For each one-dimensional time series, TOAD applies a detection method to produce a *detection time series*: a signal in $[-1, +1]$ whose sign indicates the direction of change (positive or negative) and whose magnitude reflects its relative abruptness. Grid cells with constant or all-NaN values are skipped automatically. This per-grid-cell approach means each time series is analysed in isolation, producing localised "pings" that indicate if and when an abrupt shift might occur within that individual time series, without knowledge of the other time series. Users can supply their own detection method by subclassing the `ShiftsMethod` base class (any method that produces a detection time series in the required format may be used), or use a built-in method.

TOAD v1.0 includes one built-in detection method, namely a modified version of ASDETECT (Boulton and Lenton, 2019): a general gradient-based approach for identifying anomalous rates of change without making assumptions about the underlying mechanism. We selected ASDETECT among many existing change point detection algorithms (see Reeves et al., 2007; Aminikhanghahi and Cook, 2017, for reviews) because it is fast, robust to noise, and doesn't require a reference time series for calibration. An overview of the original method is provided in Box 1. The modification in our implementation addresses a positional bias in the original centred segmentation scheme, which caused both a discontinuity near the centre of the detection

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

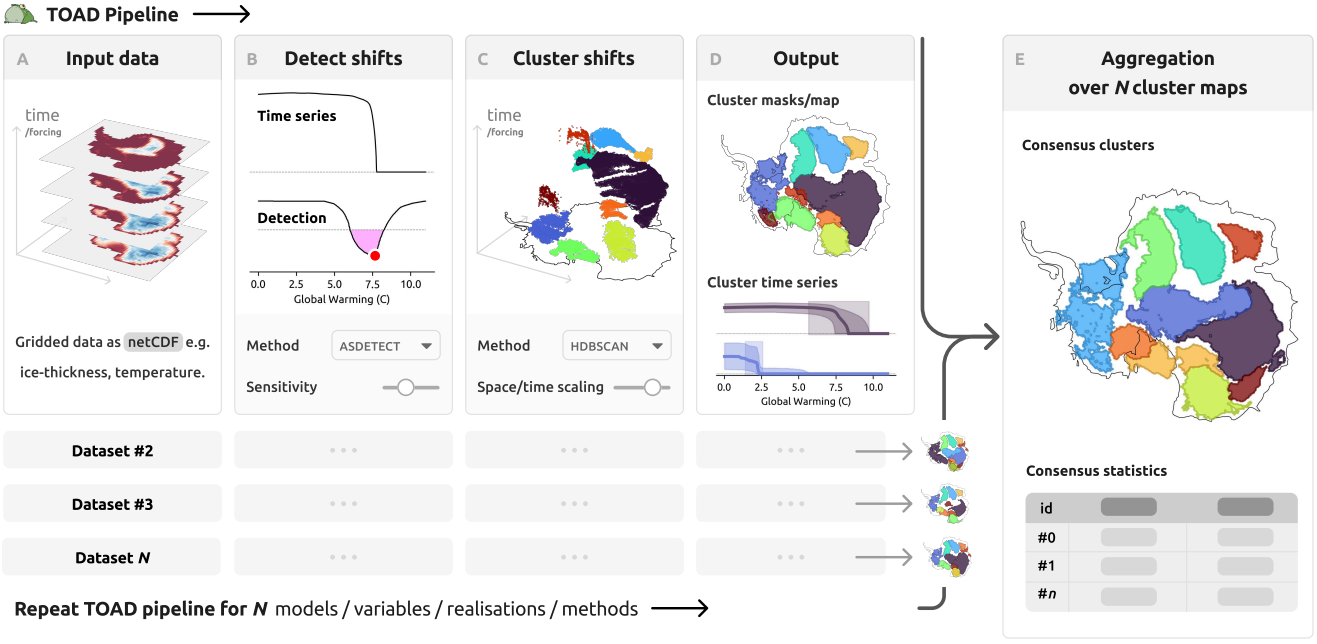

Figure 1. Overview of the TOAD analysis pipeline. (A) The input consists of gridded datasets (e.g. temperature, ice thickness) provided as NetCDF/*xarray* fields. (B) Each grid cell is analysed independently to detect abrupt shifts in its time series. Shift events are identified as local peaks in segments of the detection time series that exceed a user-defined threshold. (C) These events are clustered in a unified space–time embedding, with configurable settings that control space–time scaling, yielding (D) 3D cluster masks for each cluster, which can be collapsed to a 2D cluster map (shown here), as well as associated time series. (E) Cluster maps from multiple TOAD runs—across models, variables, ensemble members, or method configurations—are aggregated to produce ensemble-level consensus clusters and summary statistics such as mean shift time and its spread. The data used for illustration are from Antarctic Ice Sheet simulations by Garbe et al. (2020), further treated in Sect. 3.

signal and systematic differences in apparent shift magnitude depending on where in the record a shift occurs. This is described in more detail in Appendix A. We also accelerated the performance of the implementation with *numba* (Lam et al., 2015).

TOAD also introduces a `timescale` parameter for ASDETECT, allowing users to specify the minimum and maximum segmentation lengths directly in the physical units of the dataset ($\tau_{\min}$, $\tau_{\max}$). A larger $\tau_{\min}$ suppresses short-lived noise without affecting the detection of genuinely fast transitions, while a smaller $\tau_{\max}$ increases sensitivity and enables the method to resolve multiple shifts occurring in the same grid cell. In most applications, the default ASDETECT segmentation lengths provide a reasonable balance, but the timescale parameter offers finer control when needed.

This stage produces a detection time series for each grid cell. In the subsequent clustering step, abrupt shift events are extracted as peaks in the detection signal and grouped into spatio-temporal clusters.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

How ASDETECT works

ASDETECT scans across multiple temporal scales to identify points where gradients in the time series deviate significantly from the background variability. The algorithm proceeds as follows:

1. Divide the time series into non-overlapping, centred segments of length l.
2. Perform linear regression within each segment to estimate gradients.
3. Mark gradients exceeding ± 3 median absolute deviations from the median gradient as significant.
4. Repeat steps 1–3 iteratively across segment lengths $l_{\min} \ldots l_{\max}$ and normalise the cumulative detection signal by the number of segment lengths used.

The resulting detection time series yields values in the interval $[-1, +1]$, capturing both the direction and the relative strength of abrupt changes at each time step. By default, the algorithm uses $l_{\min} = 5$ time steps and $l_{\max} = N/3$, where N is the total length of the series. Setting the minimum segment length to five time steps helps filter out noise-induced artefacts, while limiting the maximum segment length to one-third of the series ensures sufficient data remains on both sides of any candidate change point for robust detection.

Box 1. Overview of the ASDETECT algorithm (Boulton and Lenton, 2019), the default shift detection method in TOAD v1.0.

2.3 Spatio-temporal clustering

The second stage of TOAD's workflow groups detected shifts into dynamically coherent regions of change by connecting individual shift events across space and time. In practice, each shift event is represented as a point in a combined space–time coordinate system: for geographic data, this is (x, y, z, t), where (x, y, z) is the Cartesian position on Earth and t is the time of the shift. Clustering then groups points that are close in *both* space and time. A resulting cluster therefore represents a set of grid cells that underwent abrupt change at similar times (or forcing levels) and in spatial proximity. The subsections below describe (i) how different grids are mapped to a common coordinate system, (ii) how shift events are selected and scaled, (iii) which clustering algorithms are supported and why HDBSCAN is the default, and (iv) how results and metadata are stored. See Figure B2 for a detailed flowchart of the clustering stage.

2.3.1 Coordinate standardisation

TOAD v1.0 supports 2D data fields defined on a variety of grid types, including Cartesian grids, regular geographic grids (lat, lon), and irregular geographic grids (lat(i, j), lon(i, j)).

For regular latitude–longitude grids, TOAD automatically applies regridding to a Hierarchical Equal Area isoLatitude Pixelation (HEALPix) grid (Gorski et al., 2005), implemented via the Python package *healpix* (Tessore, 2025), unless the user disables this step or supplies a custom regridder. HEALPix removes the polar oversampling inherent to regular geographic grids and produces nearly uniform spatial sampling, improving the consistency of distance-based clustering. Although originally developed for astronomy, it is increasingly used in Earth system applications (Chang et al., 2023; Karlbauer et al., 2024; Doblas-Reyes et al., 2025).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Since irregular geographic grids are often used specifically for polar regions, HEALPix regridding is not applied automatically for these grids but can be enabled if desired. Purely Cartesian grids are not regridded.

After regridding (when applied), all geographic coordinates are converted to three-dimensional Cartesian positions using the WGS84 ellipsoid with time appended as a fourth dimension, forming the space (x, y, z, t). Transforming coordinates into a Cartesian representation enables a unified Euclidean metric that naturally incorporates the temporal dimension and allows for fast nearest-neighbour clustering methods. Alternatives such as haversine distances cannot include time, while full spatio–temporal distance matrices scale quadratically with grid size, making them prohibitively slow or memory-intensive.

Euclidean distances closely approximate great-circle distances at the spatial scales relevant for clustering. The deviation remains below 3 percent even at 5000 km and is therefore negligible for identifying connected grid-cell neighbourhoods, where only local spatial relationships determine cluster formation. For datasets already expressed in Cartesian coordinates, no spatial transformation is performed; only the time coordinate is appended to form (x, y, t) vectors.

These coordinate transformations affect only the internal point cloud used for clustering; the original dataset stored in the TOAD object remains unchanged.

2.3.2 Shift selection and scaling

Clustering is performed on a point cloud in the space (x, y, z, t), so we must first select the data points to cluster. A threshold is applied to the detection time series to filter out only significant shifts, and then the maxima within each above-threshold interval are identified and taken as abrupt shift events. This approach captures recurring abrupt events without over-representing closely spaced detections. How should the threshold value be chosen? This depends on the detection algorithm. For ASDETECT, synthetic tests on Gaussian white-noise time series (Figure B1 in the appendix) show that small thresholds ($|dts| \lesssim 0.2$) label a large fraction of purely noisy points as shifts, whereas thresholds $|dts| \gtrsim 0.5$ reduce the false-shift rate to well below 10^{-4}. For ASDETECT, we therefore recommend a threshold of no less than 0.5. In applications where only the most pronounced events are of interest, the threshold can be increased further to remove weaker, less significant shifts.

Since spatial and temporal scales are not directly comparable, TOAD rescales the time dimension to match the mean standard deviation of the spatial coordinates. This ensures that time and space contribute equally to clustering distances by default. An additional scaling factor γ can be applied to the time dimension to control the relative weighting of time versus space, following a similar approach to Bathiany et al. (2020): larger γ values emphasise temporal proximity and produce clusters tightly grouped in time, while smaller values prioritise spatial compactness and allow broader temporal spread within clusters. This scaling factor is set to 1 by default.

2.3.3 Clustering algorithms

TOAD's clustering module supports any algorithm based on the *scikit-learn* ClusterMixin interface (Pedregosa et al., 2011), meaning that standard methods such as DBSCAN, KMeans, and AgglomerativeClustering can be used directly. Custom clustering algorithms can also be implemented following the same interface.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

The default clustering method in TOAD is HDBSCAN (Campello et al., 2013), a hierarchical density-based extension of DBSCAN (Ester et al., 1996) accessed through *scikit-learn* (Pedregosa et al., 2011). A brief explanation of HDBSCAN is given in Box 2. HDBSCAN is particularly well suited to clustering of abrupt shifts in the Earth system, because it can identify clusters of irregular shape and variable density. It handles noise robustly, does not require specifying the number of clusters in advance, and is deterministic. It requires just one intuitive hyperparameter: the *minimum cluster size*, which controls the minimum number of points required to form a cluster. DBSCAN is also useful in cases where you want to enforce a strict maximum distance between points within a cluster, which is set using the parameter `eps`. Although TOAD defaults to the *scikit-learn* implementation of HDBSCAN, users may benefit from using a faster parallelised variant (e.g. Tutte Institute for Mathematics and Computing, 2025) for large datasets.

How HDBSCAN works

HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) extends DBSCAN by identifying clusters of variable density and classifying noise points robustly. TOAD uses the *scikit-learn* implementation (Pedregosa et al., 2011). The algorithm proceeds as follows:

1. Compute mutual-reachability distances to emphasise dense regions and separate sparse noise points.
2. Build a minimum spanning tree of the mutual-reachability graph and form a hierarchical cluster tree by progressively removing edges in order of increasing distance.
3. Condense the hierarchy based on the minimum cluster size to obtain a simplified tree of candidate clusters.
4. Identify stable branches of the condensed tree (clusters with high persistence across distance scales) and select these as the final clusters; remaining points are labelled as noise.

Box 2. Overview of the HDBSCAN clustering algorithm (Campello et al., 2013), the default clustering method in TOAD v1.0.

2.3.4 Cluster masks and statistics

For each grid cell, TOAD produces a time series of integer labels indicating its cluster membership through time. We refer to this three-dimensional output (space $\times$ space $\times$ time) as a *cluster mask*. Cluster labels are sorted by decreasing cluster size (grid-cell counts) for comparability, and noise points are labelled as -1. The clustering results are added to the dataset within the TOAD object as a new variable, with incrementing suffixes if a variable of the same name already exists. Since all clustering parameters are stored as *xarray* variable attributes, clustering results can be easily reproduced and compared across different parameter settings. A cluster mask can be collapsed along the temporal dimension to form a set of two-dimensional *cluster maps*, i.e. a map for each cluster showing the spatial extent, or *footprint*. Such maps offer a starting point for physical interpretation, for example analysing local feedbacks, assessing regional stability, or examining early-warning signals in other variables.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Cluster-level statistics—such as mean timing, duration, and spatial footprint—can be computed using TOAD's statistics module. The cluster masks form the primary input for the aggregation stage (Sect. 2.4), where clustering patterns are integrated across models, variables, and ensemble members.

2.4 Aggregation

Running TOAD on different models, variables, ensemble members, or parameter settings yields a collection of three-dimensional cluster masks (space × space × time), hereafter referred to as *clusterings*. The final stage of TOAD addresses how to synthesise information across multiple clusterings and produce summary outputs. A natural starting point is to consider *where* clusters most frequently appear across the input clusterings. TOAD summarises this using the *cluster occurrence rate* diagnostic: for each grid cell, it calculates the proportion of input clusterings in which that cell belonged to a cluster. A value of 1 means the cell was always clustered; 0 means it was never clustered. This provides a simple indicator of regions where abrupt shifts are most likely to occur.

However, occurrence alone provides no information about dynamically coherent domains; in other words, it does not reveal which grid cells consistently cluster together. To address this, TOAD implements a *consensus clustering* approach inspired by Evidence Accumulation Clustering (EAC; Fred and Jain 2005). The core idea is to count how often pairs of neighbouring grid cells co-occur in the same cluster across all input clusterings, forming a sparse co-association graph. A user-specified consensus threshold $\theta_{\min}$ is then applied, and connected components in the resulting graph define the final consensus clusters. For example, setting $\theta_{\min} = 0.5$ requires that at least half of the clusterings agree on the co-association of two neighbouring cells. The temporal dimension is collapsed before computing co-association, which is necessary for computational tractability: even with spatial restriction to adjacent grid cells, including temporal connectivity would lead to an intractable number of calculations. This design choice means that consensus clusters identify regions of *spatial* agreement—grid cells that consistently cluster together across input clusterings—without requiring temporal synchronization in the clustering decision itself. Consequently, multiple abrupt shifts affecting the same or overlapping spatial regions at different times may be merged into a single consensus domain. However, temporal information is not lost: TOAD computes summary statistics for each consensus cluster, including temporal properties derived from the underlying shift detections (e.g. mean shift time and spread across input clusterings) and spatial properties (e.g. size, centroid), enabling assessment of whether spatial agreement is accompanied by temporal agreement. The algorithm is described in more detail in Box 3 and a detailed flowchart of the aggregation procedure is shown in Figure B2 in the appendix.

2.5 Diagnostic visualisation

TOAD provides core plotting tools for visualising both cluster maps and time series. Cluster maps allow spatial inspection of cluster assignments with customisable projections. For temporal analysis, TOAD supports plotting individual time series as well as aggregate statistics (mean, median, interquartile range). Combined plots enable side-by-side views of cluster maps and their corresponding time series, making it easy to interpret the spatial and temporal behaviour of detected clusters. TOAD

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

How consensus clustering works

TOAD's consensus clustering method is inspired by Evidence Accumulation Clustering (Fred and Jain, 2005), adapted for large Earth-system grids. The standard EAC approach constructs a dense co-association matrix over all point pairs and then applies hierarchical clustering (typically single-link) to obtain the final clustering. This quickly becomes intractable at typical climate model resolutions, as the number of point pairs scales quadratically with grid size. The first step toward tractability is therefore to restrict evaluation to spatially adjacent grid cells only. However, even with this spatial restriction, including the temporal dimension remains computationally prohibitive. For example, a global 1° grid contains $360 \times 180 = 64{,}800$ spatial grid cells. If we then add daily output over one year, we obtain $N \approx 2.4 \times 10^7$ space–time nodes. Even restricting connectivity to immediate neighbours (approximately eight spatial and two temporal neighbours per node) results in $E \approx 2.4 \times 10^8$ edges, which is too large to evaluate in practice. To make the method computationally tractable, TOAD therefore collapses the temporal dimension before computing co-association. The algorithm proceeds as follows:

1. Construct a sparse spatial adjacency graph connecting each grid cell to its immediate neighbours—via k-nearest neighbours on the sphere (for geographic grids) or 8-connectivity (for Cartesian grids). For regular latitude–longitude grids, HEALPix regridding is applied to obtain an equal-area neighbourhood structure. This adjacency graph is built once and reused across all clusterings.
2. For each clustering v and each cluster ID within that clustering, collapse the temporal dimension to form a two-dimensional spatial footprint: a grid cell is marked as active if it was assigned to that cluster at any point in time.
3. For each clustering v, evaluate each edge (i, j) in the adjacency graph: (i) a *vote* $V_{ij}^{(v)} = 1$ if cells i and j belong to the same cluster in that clustering (i.e. both are active in the same cluster's spatial footprint), and 0 otherwise; and (ii) an *availability count* $A_{ij}^{(v)} = 1$ if both cells are valid (non-NaN) in that clustering, and 0 otherwise. For example, if one grid cell lies outside the valid data domain (e.g. ocean for a land-only variable) and is always NaN, that edge is unavailable in that clustering.
4. Compute the consensus weight for each edge as

   $$W_{ij} = \frac{\sum_v V_{ij}^{(v)}}{\sum_v A_{ij}^{(v)}},$$

   representing the fraction of clusterings in which cells i and j are jointly clustered.
5. Retain edges with $W_{ij} \geq \theta_{\min}$, where $\theta_{\min}$ is the user-specified consensus threshold.
6. Identify connected components in the resulting sparse graph; these define the final consensus clusters. Unlike standard EAC, which uses hierarchical clustering on the full co-association matrix and may produce non-contiguous clusters, TOAD's use of connected components ensures that each consensus cluster is spatially contiguous (all cells within a cluster are connected through adjacent cells that meet the threshold). Separate clusters can border each other; if two adjacent cells belong to different clusters, the edge between them did not meet the consensus threshold.

Box 3. Overview of TOAD's consensus clustering method, used to aggregate multiple clusterings into a single spatial consensus cluster map.

also reports a range of cluster statistics, including temporal properties (start, end, duration, steepest gradient, and interquartile ranges) and spatial metrics (spatial footprint size, location).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

3 Demonstrations

The following examples serve to showcase how TOAD works, illustrate the types of results it produces, and demonstrate its versatility across diverse Earth system components. These demonstrations are not intended as comprehensive analyses of the respective systems; rather they provide proof-of-concept applications illustrating TOAD's methodological capabilities. In-depth physical interpretation and systematic evaluation of abrupt shifts in each system are left for dedicated follow-up studies.

3.1 Synthetic data

We first demonstrate that TOAD can recover known abrupt shifts in a simple synthetic dataset. We generate a dataset containing two sigmoid shift regions with different timings and magnitudes, imposed on a white-noise background (Figure 2A). TOAD successfully detects the exact grid cells on which the synthetic shifts were imposed (Figure 2B). The corresponding cluster time series (Figure 2C) clearly show both the early and late transitions, and the shading highlights the shift periods despite the background noise. While these synthetic shifts are highly idealised, the experiment demonstrates that TOAD can successfully detect and recover shift regions when the underlying structure is unambiguous; more systematic benchmarking will be possible with the forthcoming benchmark suite by Röhrich et al. (forthcoming).

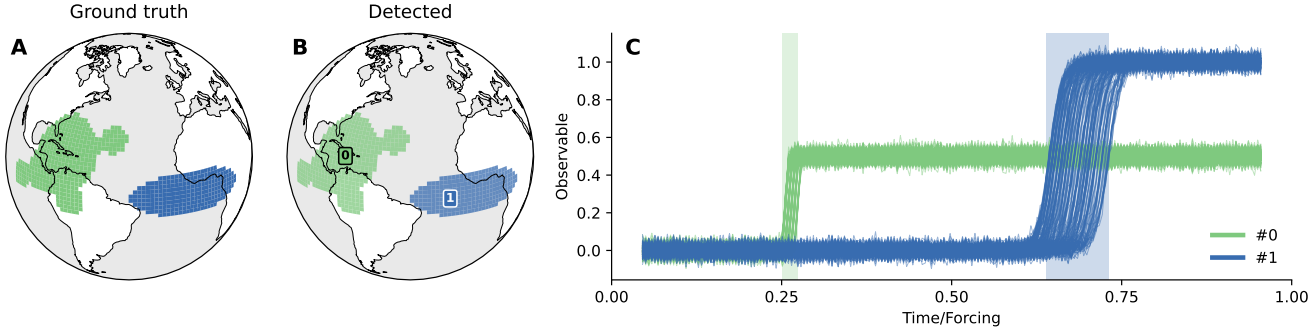

Figure 2. Synthetic shift experiment demonstrating TOAD's ability to recover known spatio-temporal transitions. (A) Two synthetic shift regions generated with white-noise background variability and spatially correlated shift timings. (B) TOAD accurately recovers both shift regions. (C) Time series from all grid cells belonging to each detected cluster. Shaded regions denote the detected transition windows.

3.2 Antarctic Ice Sheet

The West Antarctic Ice Sheet and parts of the East Antarctic Ice Sheet are widely regarded as climate tipping elements because strongly positive, nonlinear feedbacks—notably marine ice sheet instability (MISI) driven by grounding-line retreat and ocean water undercutting the ice sheet—can drive large-scale, self-reinforcing ice loss beyond critical warming thresholds (Mercer, 1978; Feldmann and Levermann, 2015; Fox-Kemper et al., 2021; Garbe et al., 2020). Here, we apply TOAD to the Antarctic Ice Sheet hysteresis simulation by Garbe et al. (2020) (Figure 3), in which the Parallel Ice Sheet Model (PISM) was forced

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

with a prescribed global mean surface temperature (GMST) starting from the pre-industrial ice-sheet state and increasing incrementally (from 0 to ∼14 °C above pre-industrial, at a rate of 10^{-4} °C yr^{-1}) until near-complete deglaciation of the Antarctic continent is reached. The warming rate is chosen to be sufficiently slow for the ice sheet to remain close to equilibrium throughout the simulation, so each model state can be interpreted as a quasi-equilibrium configuration at a given GMST. This structure makes the dataset particularly suitable for identifying dynamical instabilities as abrupt deviations from otherwise smooth trajectories.

The analysis targets the ice-thickness field (`thk`), represented on a 341×341 grid (with a model grid resolution of 16 km) across 700 GMST-indexed snapshots. To suppress numerical noise near zero ice thickness, a lower bound of 1 m is imposed before any detection. Abrupt shifts are first identified using the ASDETECT algorithm (see Sect. 2.2), applying a timescale window of $(0.5°C, 3.5°C)$—which translates to segment lengths of `lmin=24` and `lmax=174` simulation steps. Subsequently, the detected shift events are clustered with HDBSCAN, restricting analysis to negative shifts (i.e., ice-sheet thinning) and using a standard shift threshold of 0.5 together with `min_cluster_size=100`, equivalent to a spatial footprint of 25,600 km^2.

Using these parameters, 36 clusters are identified; Figure 3 displays these clusters and presents the trajectories for the eight largest. Computationally, the full detection and clustering pipeline is fast: ASDETECT (applied to all 116,281 grid cells) completes in about 1.5 minutes on an Apple M1 Pro laptop, while the HDBSCAN clustering requires ∼10 seconds (or ∼1 second using the accelerated *fast_hdbscan* implementation (Tutte Institute for Mathematics and Computing, 2025)).

The eight largest clusters (Figure 3) reveal broad, dynamically coherent regions that undergo abrupt shifts at distinct intervals along the GMST-warming trajectory. Their spatial footprints form contiguous domains across West and East Antarctica. In West Antarctica, two of the largest regions shift at low warming levels: cluster #5 spans large parts of the Ross region and collapses at warming levels between about 1 and 2 °C above pre-industrial levels, followed by cluster #2, which encompasses the Amundsen Sea and Filchner-Ronne Ice Shelf regions and collapses at warming levels of about 2 to 2.5 °C above pre-industrial levels. In East Antarctica, cluster #0—the largest cluster covering ca. 1.7 million km^2 and spanning large parts of the Aurora Subglacial Basin—collapses at warming levels between about 7 and 8°C above pre-industrial levels, while cluster #1 spans parts of the central East Antarctic plateau, including the South Pole, and collapses at even higher warming levels of about 8 to 10 °C above pre-industrial levels, broadly consistent with the findings of Garbe et al. (2020). Clusters covering multiple present-day drainage systems are expected and likely reflect the fact that as the ice sheet thins and retreats, its flow structure reorganises (Loriani et al., forthcoming). The collapse events are characterised by narrow, temporally coherent transition bands, indicating synchronised ice loss within each cluster.

Although the simulation evolves close to equilibrium, the imposed warming rate remains finite and sufficiently fast that the ice sheet cannot fully adjust to each incremental forcing level. As a result, in temperature regimes of particularly high ice-sheet sensitivity with respect to forcing, the transient simulation differs noticeably from stationary true-equilibrium runs (compare curves and triangles in Figure 2 of Garbe et al. (2020)), which is reflected in the transitions appearing gradual rather than Heaviside-like. In this setting, the detected collapse events correspond to periods of most rapid ice loss and therefore likely occur after the underlying critical threshold has been crossed, consistent with the concept of committed ice loss under transient forcing (Coulon et al., 2025).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

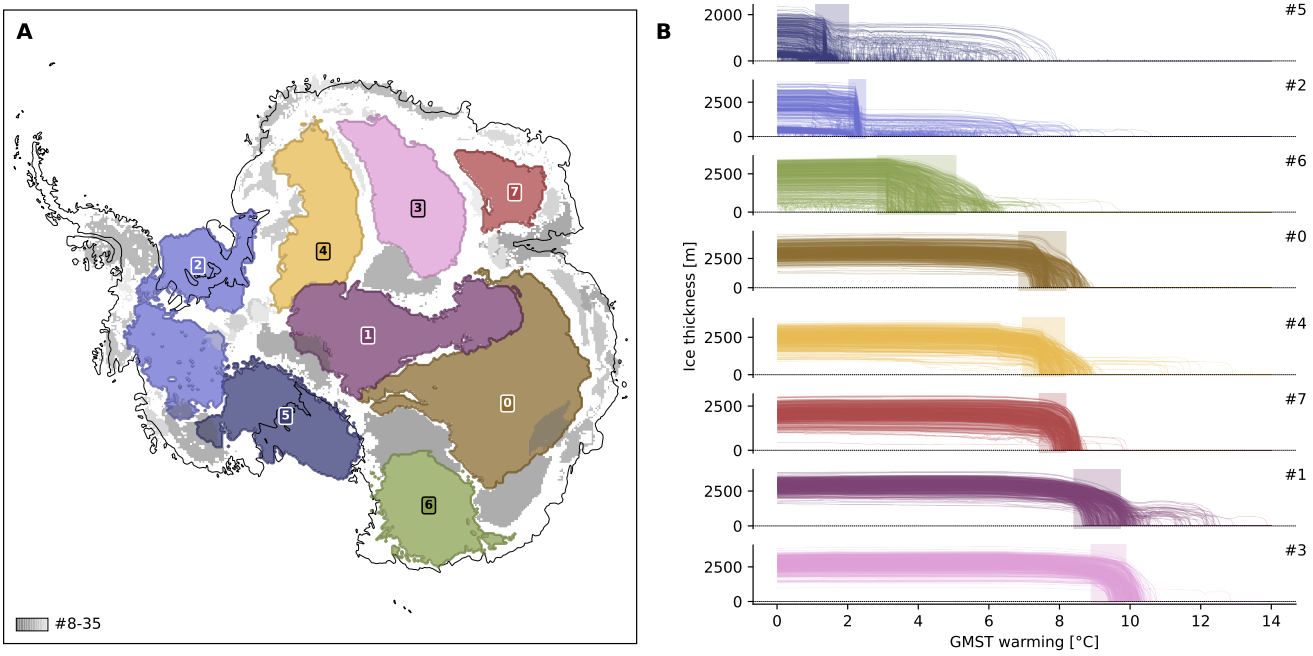

Figure 3. Clusters of abrupt ice-sheet thinning detected by TOAD in a quasi-equilibrium Antarctic Ice Sheet warming simulation. (A) Spatial distribution of the eight largest clusters (outlined and labelled) out of 36 detected clusters; remaining clusters are shown in greyscale for context. The simulation by Garbe et al. (2020) applies a prescribed GMST warming rate of 10^{-4} °C yr^{-1}, such that each GMST value represents an approximate steady state. Abrupt thinning events are detected with ASDETECT using a GMST-based timescale window of $(0.5, 3.5)$°C, a shift threshold of 0.5, and a temporal weighting $\gamma = 2.0$, and are clustered using HDBSCAN with `min_cluster_size=100`. (B) Ice-thickness trajectories for all grid cells belonging to each of the eight largest clusters. Shaded boxes indicate the GMST interval between the first and last detected shift events across all grid cells in each cluster; individual shift times are defined as the points of maximum shift rate.

3.3 Global terrestrial biosphere

As a third example, we demonstrate TOAD's application to vegetation carbon (`vegc`) output from the Lund-Potsdam-Jena dynamical global vegetation model with managed Land (LPJmL) v4.0 (Schaphoff et al., 2018). LPJmL integrates plant functional types, fire, hydrology, and biogeochemical dynamics to simulate terrestrial vegetation and carbon-cycle processes. The transient simulation starts in 1861 with historical land use (Fader et al., 2010) applied until 2005 and held constant thereafter. Climate forcing follows HadGEM2-ES under RCP8.5 (Taylor et al., 2012; Riahi et al., 2011), bias-corrected following the ISIMIP2b protocol (Frieler et al., 2017; Lange, 2018); end-of-century climate variability (2090–2099) is randomly repeated beyond 2100, and CO_2 is fixed from 2020 onward. Our analysis spans 1900–2200.

We apply standard ASDETECT to the vegetation carbon time series and perform clustering using a shift threshold of 0.75, a time weighting of 3.0, and the DBSCAN algorithm with `eps=500` and `min_samples=50`. While HDBSCAN is also

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

applicable, in this case it produced numerous low-density clusters, so instead we opted for DBSCAN. All parameters have again been chosen for illustration purposes and can be tuned to emphasise or de-emphasise different aspects of the data.

Figure 4 shows six negative-shift clusters. Major shifts are detected in North and South America and across Eurasia in the latter half of the 21st century (clusters #0, #1, #2, and #3). Apparent recoveries in some clusters (e.g., clusters #1 and #2) likely reflect transitions toward new vegetation compositions rather than a return to prior states. This also motivates future extensions of TOAD to explicitly assess hysteresis and reversibility.

3.4 North-Atlantic Subpolar Gyre

For our final example, we use TOAD to detect and aggregate abrupt changes in mixed-layer depth (`mlotst`) in the Subpolar Gyre (SPG) region, using 10 ensemble members from the fully coupled MPI-ESM1-2-LR model under ScenarioMIP SSP5-8.5 (Wieners et al., 2019).

Prior to analysis, the data is restricted to the Northern Hemisphere, converted to annual means, and temporally smoothed using a Gaussian filter with an effective width of 6 years ($\sigma = 6/\sqrt{12}$). The resulting dataset comprises 86 annual time steps on the native ocean grid. Each ensemble member is first analysed independently. ASDETECT is applied for shift detection and DBSCAN for clustering. DBSCAN was chosen after HDBSCAN was found to produce numerous low-density clusters for this dataset. Only negative shifts are considered, using DBSCAN parameters `eps = 400` and `min_samples = 25`, with the temporal dimension weighted by a factor of $\gamma = 2.0$. Panels A and B in Figure 5 show two representative examples out of the 10 ensemble members.

After running the analysis for all ensemble members, aggregation is performed as described in Sect. 2.4. Figure 5C shows the consensus cluster map with a 0% consensus threshold, corresponding to the union of all clusters detected across ensemble members and revealing numerous clustered regions across the North Atlantic. Some clustered regions at high northern latitudes likely reflect known low-resolution convection artefacts in MPI-ESM1-2-LR rather than physically realistic deep convection sites. Figure 5D shows the consensus cluster map with a 50% threshold, where only two regions remain: cluster #0 in the Irminger Sea and cluster #1 in the Labrador Sea. These regions are consistent with known deep convection sites in the SPG (Swingedouw et al., 2021). We verified that the spatial structure of detected clusters is robust to alternative smoothing strategies, including rolling means and variations in Gaussian filter width. However, detected shift timing is sensitive to the degree of smoothing: stronger smoothing can move detected shift times later by 1–2 years, consistent with the fact that ASDETECT identifies the point of maximum gradient, which smoothing displaces.

TOAD also computes summary statistics for each consensus cluster, including the ensemble-mean shift time and its inter-ensemble standard deviation. These quantities are obtained by averaging detected shift times across all grid cells and ensemble members contributing to each consensus cluster. Both clusters shift around year 2027, while the inter-ensemble standard deviation is smaller for cluster #1 (~ 4 years) than for cluster #0 (~ 6 years). Figure 5E shows the cluster occurrence rate, defined as the fraction of ensemble members in which each grid cell participates in any cluster. It indicates that parts of both the Labrador and Irminger Sea regions participate in clusters in nearly all ensemble members.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

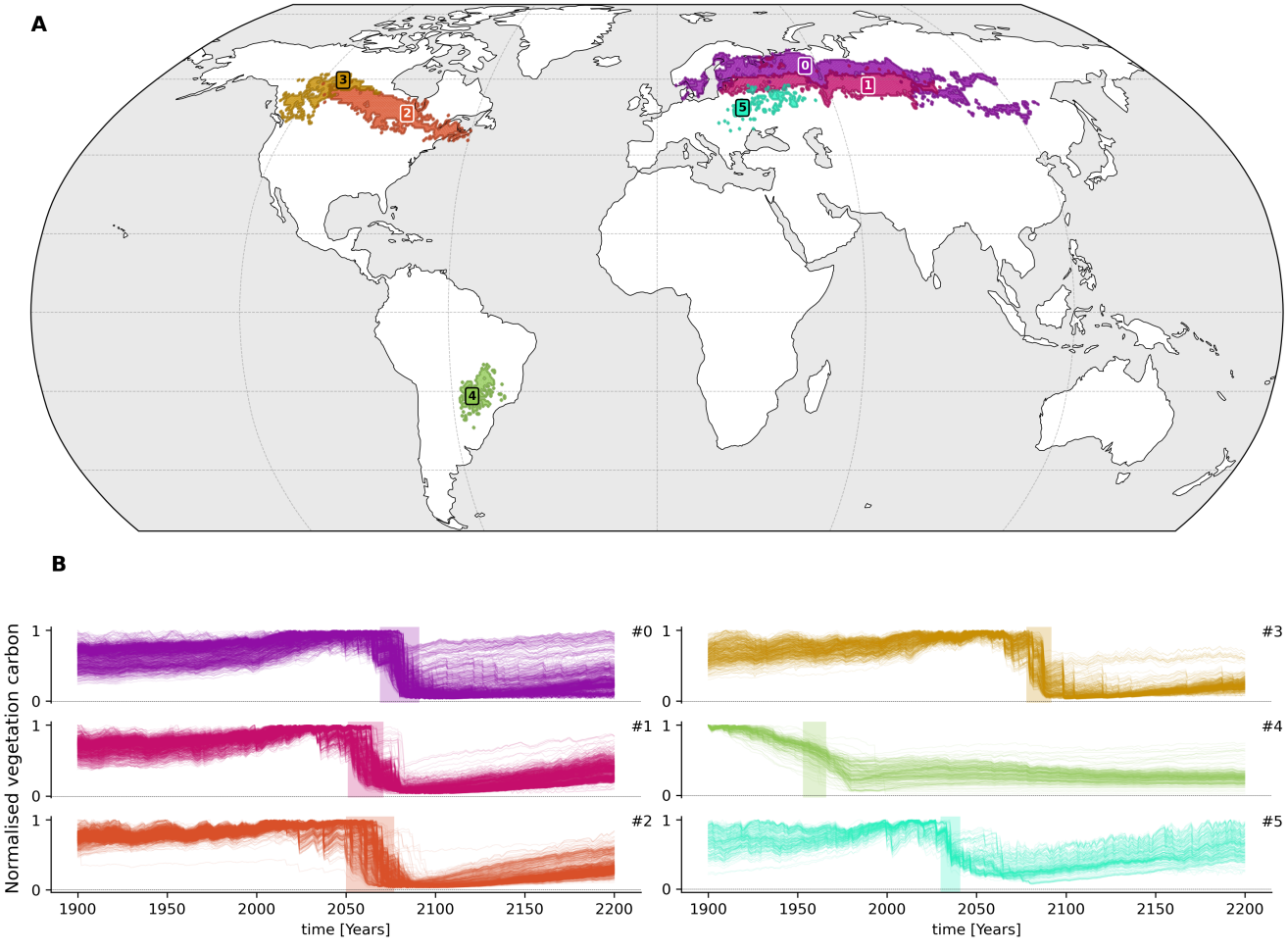

Figure 4. Abrupt vegetation carbon shifts detected by TOAD in the LPJmL4.0 global vegetation model during simulation years 1900–2200. The transient simulation starts in 1861 with historical land use applied until 2005 and held constant thereafter, forced by HadGEM2-ES climate under RCP8.5, bias-corrected following the ISIMIP2b protocol; end-of-century climate variability (2090–2099) is randomly repeated beyond 2100, and CO_2 is fixed from 2020 onward. Abrupt shifts are detected with ASDETECT using a shift threshold of 0.75 and a temporal weighting of $\gamma = 3.0$, and are clustered using DBSCAN with `eps=500` and `min_samples=50`. (A) Spatial footprint of the six largest clusters of negative shifts in vegetation carbon (`vegc`) detected between 1900 and 2200. Each cluster represents a region in which abrupt declines occur. (B) Time series for all grid cells in each cluster, normalised to their own maxima for clearer visual comparison; shaded boxes mark the temporal range over which the cluster is actively shifting.

Mixed-layer depth is used here as a local indicator of upper-ocean stratification and deep convection activity rather than as a direct measure of basin-scale circulation strength. It is important to note that Swingedouw et al. (2021) found no Subpolar Gyre collapse in MPI-ESM1-2-LR under SSP5-8.5 forcing. This does not contradict the results presented here. That study applied

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

strict, basin-scale criteria to identify an actual collapse of the SPG, whereas here we only detect abrupt, spatially coherent shifts in mixed-layer depth, which may occur without leading to a full Subpolar Gyre collapse.

4 Discussion and outlook

The examples in Sect. 3 serve as proof-of-concept demonstrations that illustrate TOAD's ability to detect and cluster abrupt shifts across diverse systems and to synthesise results from multiple ensemble members using consensus clustering. More extensive evaluation across different dynamical regimes will be needed to fully assess TOAD's performance, and proper analysis of abrupt shifts in these systems will require dedicated investigation beyond the scope of this methodological framework paper. In the following, we discuss some of the limitations of TOAD v1.0 and potential directions for future development.

TOAD is fundamentally an introspection tool designed to facilitate exploration of abrupt-shift structure across different parameter settings. Different choices of shift thresholds, temporal weightings, and clustering parameters can produce substantially different results, each revealing distinct aspects of the underlying shift dynamics. The parameter combinations shown in this paper yield spatially coherent, interpretable clusters that align with known physical processes, but alternative parameterisations could equally well emphasise different scales, timings, or spatial structures of abrupt change. This parameter sensitivity is intentional: exploring parameter space allows researchers to assess the robustness of detected patterns and investigate shifts at multiple scales, making parameter selection an integral part of the analysis workflow. The computationally efficient, object-oriented design enables users to quickly test various parameter combinations and compare results.

While manual parameter exploration is valuable, automated approaches may be useful in some contexts. TOAD v1.0 therefore provides an experimental cluster-optimisation method that allows for semi-automatic parameter selection. The optimisation backend is Optuna (Akiba et al., 2019), which samples combinations of user-selected parameters such as the shift threshold, temporal weighting (γ), and method-specific settings like the minimum cluster size. The objective functions are defined by different cluster scoring functions that reward different aspects of the clustering results, such as spatial autocorrelation, temporal consistency, nonlinearity, or a combination thereof (De Maeyer et al., forthcoming). However, this feature remains experimental and has not proven sufficiently robust for routine use across diverse applications, as the choice of appropriate objective functions and their relative weightings often requires domain-specific knowledge that cannot be easily automated.

TOAD v1.0 currently supports only 2D data fields, not 3D fields that include a vertical dimension. Extending shift detection to 3D fields is trivial since detection operates independently on individual grid cells, and the clustering step could readily accommodate a third spatial dimension by mapping it to the z component in the Euclidean embedding. However, consensus clustering faces dimensionality constraints: the same computational challenges that prevent temporal structure from being preserved in consensus clustering (see Sect. 2.4) also impede the inclusion of a third spatial dimension. Developing consensus clustering approaches that can handle higher-dimensional structures—whether temporal or spatial—is therefore an important future direction. Successfully extending consensus clustering to preserve temporal structure would not only enable direct identification of regions with both spatial and temporal agreement, but would also unlock support for 3D data fields, as the same algorithmic redevelopment would address both challenges.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

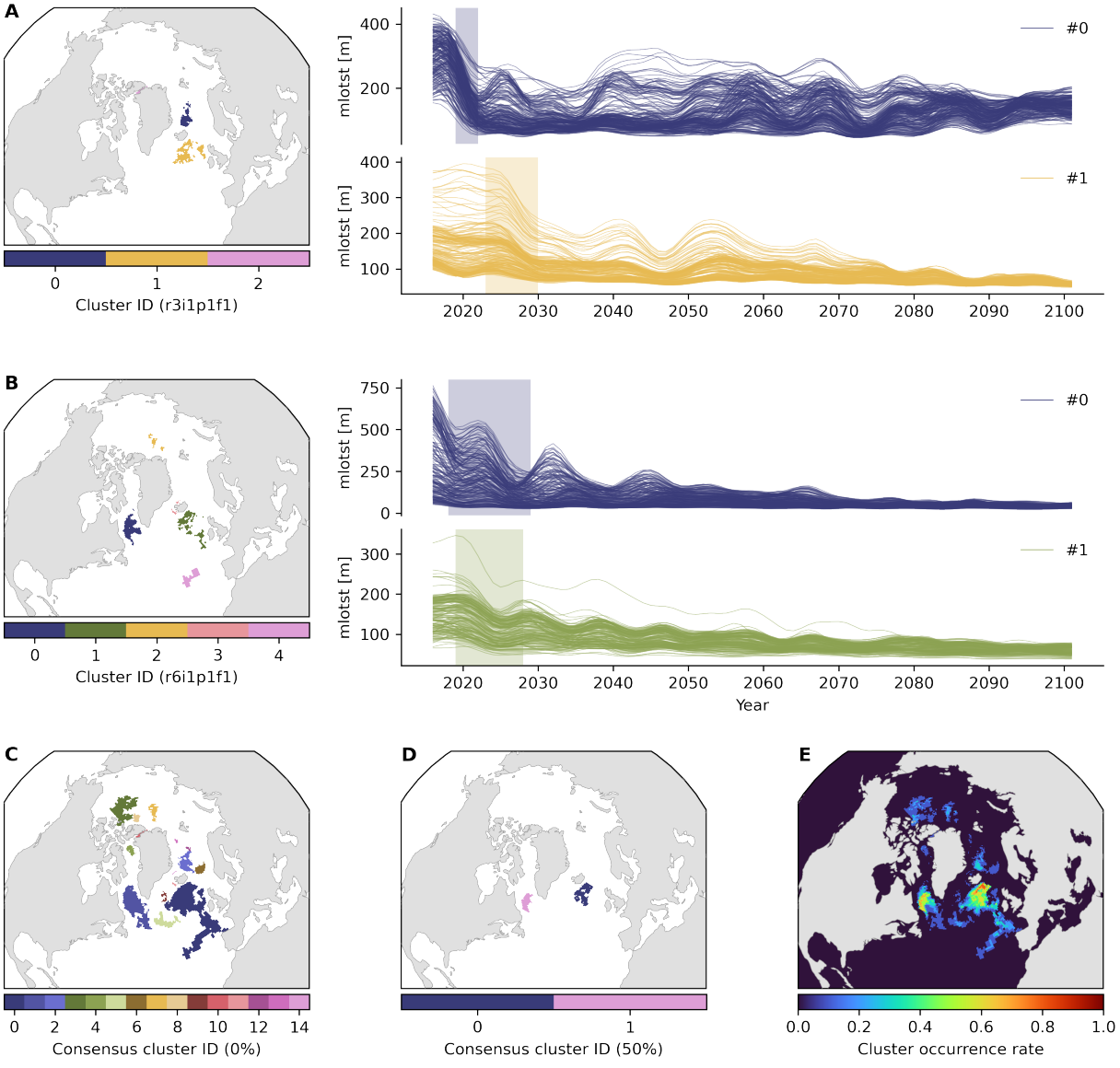

Figure 5. Mixed-layer depth shifts around the Subpolar Gyre detected by TOAD in the MPI-ESM1-2-LR fully coupled Earth system model. Abrupt shifts are detected with ASDETECT and clustered using DBSCAN with `eps=400` and `min_samples=25`, with temporal weighting $\gamma = 2.0$. Only negative shifts are considered. (A–B) Two representative ensemble members: spatial clusters (left) and time series of the two largest clusters (right; shaded boxes indicate shift periods). (C) Consensus cluster map with 0% threshold (union of all clusters across 10 ensemble members). (D) Consensus cluster map with 50% threshold, showing regions in the Labrador and Irminger Seas jointly clustered in at least half of the ensemble members. (E) Cluster occurrence rate (fraction of ensemble members in which each grid cell participates in any cluster).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Broadening TOAD's shift detection capabilities is an important development priority. While ASDETECT performs well for monotonic, step-like transitions (e.g. ice-sheet retreat or vegetation loss), it is not suited to oscillatory or quasi-periodic or chaotic systems such as monsoons or atmospheric circulation regimes, where abrupt changes may manifest as shifts in variance, frequency, or recurrence structure rather than in the mean state. These limitations reflect the properties of the chosen default detector rather than an intrinsic constraint of the TOAD framework itself. Recurrence-based approaches (Donges et al., 2011; Marwan et al., 2009) or machine-learning methods (Huang et al., 2024b) that capture nonlinear dynamical behaviour and associated transitions therefore represent promising complementary detectors, which could be integrated into TOAD's modular architecture through future community-driven extensions.

Future work will also focus on facilitating essential pre-processing steps for multi-model analyses, including spatial grid harmonisation and the merging of multiple datasets, either directly within TOAD or through tighter integration with existing tools. Another important direction is the development of explicit hysteresis diagnostics to assess reversibility under reversed forcing. In addition, alternative aggregation strategies for synthesising results across multiple models and variables, such as those proposed by De Maeyer et al. (forthcoming), represent promising avenues for future development.

TOAD occupies a distinct position within the broader ecosystem of tools for detecting abrupt or nonlinear change in climate data. Edge-detection approaches (Bathiany et al., 2020; Terpstra et al., 2025) detect sharp transitions in space and time and provide powerful large-scale multi-model assessments, but do not provide a user-extensible pipeline. System-specific catalogues of abrupt shifts (Angevaare and Drijfhout, 2025; Drijfhout et al., 2015) provide valuable inventories but rely on variable-specific criteria and do not generalise across model components. TOAD complements these efforts by offering a modular pipeline for abrupt-shift detection and synthesis along with integrated functionality for statistics, plotting, aggregation, and spatial consensus clustering. In this sense, TOAD is not an alternative detection algorithm but an integrating infrastructure capable of incorporating and comparing different methods. TOAD's *xarray*-based design also makes it technically compatible with community analysis frameworks such as *ESMValTool* (Eyring et al., 2020), where it could function as a specialised *diagnostic* for abrupt-shift analysis.

A common challenge for these approaches is the lack of systematic evaluation using standardised benchmark datasets. Establishing such benchmarks, spanning a range of dynamical regimes, is an important next step for the field. Well-designed benchmarks would enable rigorous and transparent comparison across different detection methods (Röhrich et al., forthcoming)—including future community-contributed algorithms integrated into TOAD as well as methods developed outside the TOAD framework—ultimately driving progress through shared standards and more reproducible analyses.

Viewed through the broader scientific questions that motivate TOAD, the framework provides robust answers for *where* abrupt changes occur by delineating spatially coherent regions of co-occurring shifts, and a first-order indication of *when* such changes manifest along a prescribed trajectory. The interpretation of the latter depends on both the detection method and on whether the independent axis represents physical time or an external forcing variable. Using ASDETECT, detected shift events correspond to periods of *maximum rate of change* in the analysed time or forcing series rather than to the point at which system stability is first lost. Only in the idealised limit of true equilibrium experiments with infinitesimally slow forcing would such events coincide with critical thresholds. In practical applications involving transient simulations and finite

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

forcing rates, detected shifts therefore typically lag the underlying critical threshold. A general, system-agnostic estimator of critical thresholds is thus neither realistic nor necessarily desirable. Instead, abrupt shift times identified by TOAD should be interpreted as dynamical diagnostics and may be complemented by additional indicators of transition timing—such as early-warning signals (Dakos et al., 2024) or machine-learning-based diagnostics of rate-induced tipping (Huang et al., 2024a) or nonlinear regime change (Donges et al., 2011; Marwan et al., 2009)—which could be analysed jointly with shift detectors such as ASDETECT.

Together, these considerations illustrate where TOAD already performs robustly and where further methodological development is required. Addressing these points will be essential to ensure that TOAD scales effectively to the multi-model, multi-variable analyses envisioned in forthcoming tipping-point intercomparison efforts.

5 Conclusion

We have presented the first public release of the *Tipping and Other Abrupt events Detector* (TOAD v1.0), an open and modular framework designed to address a central challenge in tipping-point research: the lack of a consistent, transparent, and scalable approach for detecting and comparing abrupt shifts in gridded Earth-system data ensembles. By unifying grid-level shift detection, space–time clustering, and ensemble synthesis within a single workflow, TOAD provides the methodological structure needed for coordinated, reproducible analyses in intercomparison efforts such as TIPMIP (Winkelmann et al., 2025).

The demonstrations across ice-sheet, vegetation carbon, and ocean mixed-layer depth simulations, as well as a synthetic benchmark dataset, illustrate TOAD's ability to identify spatially coherent domains of abrupt change. The mixed-layer depth example further demonstrates TOAD's ensemble aggregation capabilities, showing how consensus clustering can synthesise results across multiple ensemble members. In doing so, TOAD offers structured answers to several core questions in tipping-point research: it robustly identifies *where* abrupt changes occur, provides a first-order indication of *when* these transitions manifest in the data along a forcing trajectory, and quantifies *uncertainty* through ensemble spread in shift timing and spatial footprint.

A central strength of the framework lies in its flexible, modular architecture and the hyperparameters associated with each stage of the pipeline. By tuning detection timescales, clustering sensitivities, space–time scaling, or consensus settings, users can obtain complementary perspectives on their dataset and uncover patterns that would remain obscured under a single methodological choice. TOAD thus provides not only a unified workflow but also a flexible exploratory environment that helps reveal the structure and coherence of abrupt changes across models and variables.

Importantly, TOAD is not intended to yield a single definitive detection for a given dataset. Instead, it is designed as an introspective analysis tool that allows users to interrogate abrupt-change behaviour across scales. Exploring how detected patterns vary across parameter choices and scale settings is an integral part of the analysis. Structures that persist across parameter choices can be identified via consensus clustering and interpreted as robust features of the system. In this sense, TOAD functions less as a point estimator of tipping behaviour and more as a configurable lens for exploring abrupt-change

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

dynamics across spatial and temporal scales. In its bottom-up approach, it allows users to reveal "hidden" tipping behaviour that is otherwise obscured by regional averaging (Loriani et al., forthcoming).

TOAD's outputs further serve as a foundation for downstream analysis. Clusters capture physically meaningful domains for feedback analysis, early-warning signal evaluation, or multi-variable diagnostics. Ensemble synthesis enables transparent intercomparison across models and methods, together with systematic quantification of uncertainty across scenarios.

The framework remains under active development. Key priorities include extending shift detection methods to better capture nonlinear and oscillatory dynamics, developing systematic benchmark datasets, facilitating multi-model preprocessing (e.g. spatial harmonisation and dataset merging), and implementing dedicated hysteresis diagnostics. As these capabilities mature, TOAD will increasingly support comprehensive and comparable assessments of abrupt Earth-system change, such as TIPMIP and CMIP. TOAD is openly developed on GitHub, and we invite contributions from the community to help address these challenges and further expand the framework's capabilities.

Code and data availability. TOAD is available as open-source software at https://github.com/tipmip-methods/toad, including tutorials and documentation. The version used in this paper (v1.0.0) is archived on Zenodo at https://doi.org/10.5281/zenodo.18339224. The exact code, test data, and environment used to generate all figures in the main text are archived separately at https://doi.org/10.5281/zenodo.18339253.

Antarctic Ice Sheet simulation data are from Garbe et al. (2020). MPI-ESM1-2-LR mixed-layer depth data are from CMIP6 ScenarioMIP (Eyring et al., 2016; Wieners et al., 2019) and were accessed via the Earth System Grid Federation (ESGF). LPJmL vegetation carbon simulation data were provided by Boris Sakschewski and Werner von Bloh (PIK) for this study.

Author contributions. JH led manuscript writing with contributions from all authors. JH, LR, SL, and KDM developed the code. JG and BS provided data. RW and SL conceived the project. All co-authors contributed to discussion and provided feedback on the manuscript.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. We would like to thank José Licón-Saláiz, Shivani Ehrenfeucht, Max Bechthold, and Reik Donner for insightful scientific discussions and especially Nicole Cristine Laureanti for discussions on the SPG results. We thank Boris Sakschewski and Werner von Bloh for providing the LPJmL simulation data used in this study. We are also grateful to Karoline Ramin, Vincent Overbeck, Antonia Lehrmann, and Ann Sophie Messerschmidt for their contributions to earlier versions of TOAD through student projects.

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the Max Planck Institute for Meteorology (MPI-M) for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

This is ClimTip contribution #129; the ClimTip project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101137601: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

AI tools (e.g., ChatGPT, OpenAI; Claude, Anthropic) have been used to refine parts of the manuscript.

Parts of the results in this work make use of the colormaps in the CMasher package (van der Velden, 2020).

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, https://doi.org/10.1145/3292500.3330701, 2019.

Aminikhanghahi, S. and Cook, D. J.: A survey of methods for time series change point detection, Knowledge and Information Systems, 51, 339–367, https://doi.org/10.1007/s10115-016-0987-z, 2017.

Angevaare, J. R. and Drijfhout, S. S.: Catalogue of Strong Nonlinear Surprises in ocean, sea-ice, and atmospheric variables in CMIP6, EGUsphere, https://doi.org/10.5194/egusphere-2025-2039, 2025.

Armstrong McKay, D. I., Staal, A., Abrams, J. F., et al.: Exceeding 1.5 °C global warming could trigger multiple climate tipping points, Science, 377, eabn7950, https://doi.org/10.1126/science.abn7950, 2022.

Bathiany, S., Hidding, J., and Scheffer, M.: Edge Detection Reveals Abrupt and Extreme Climate Events, Journal of Climate, 33, 6379–6393, https://doi.org/10.1175/jcli-d-19-0449.1, 2020.

Bollasina, M. A., Ming, Y., and Ramaswamy, V.: Anthropogenic aerosols and the weakening of the South Asian summer monsoon, science, 334, 502–505, 2011.

Boulton, C. A. and Lenton, T. M.: A new method for detecting abrupt shifts in time series, F1000Research, https://doi.org/10.12688/f1000research.19310.1, 2019.

Campello, R. J., Moulavi, D., and Sander, J.: Density-Based Clustering Based on Hierarchical Density Estimates, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 160–172, https://doi.org/10.1007/978-3-642-37456-2_14, 2013.

Chang, A., Lee, H., Fu, R., and Tang, Q.: A seamless approach for evaluating climate models across spatial scales, Frontiers in Earth Science, 11, 1245 815, https://doi.org/10.3389/feart.2023.1245815, 2023.

Coulon, V., Klose, A. K., Edwards, T., Turner, F., Pattyn, F., and Winkelmann, R.: From short-term uncertainties to long-term certainties in the future evolution of the Antarctic Ice Sheet, Nature communications, 16, 10 385, https://doi.org/10.1038/s41467-025-66178-w, 2025.

Dakos, V., Boulton, C. A., Buxton, J. E., Abrams, J. F., Arellano-Nava, B., Armstrong McKay, D. I., Bathiany, S., Blaschke, L., Boers, N., Dylewsky, D., et al.: Tipping point detection and early warnings in climate, ecological, and human systems, Earth System Dynamics, 15, 1117–1135, https://doi.org/10.5194/esd-15-1117-2024, 2024.

De Maeyer, K., Loriani, S., Harteg, J., Donges, J. F., and Winkelmann, R.: Aggregating detected dynamics across models and variables: a framework for MIP-based Earth system analyses, in preparation, forthcoming.

Doblas-Reyes, F. J., Kontkanen, J., Sandu, I., Acosta, M., Al Turjmam, M. H., Alsina-Ferrer, I., Andrés-Martínez, M., Arriola, L., Axness, M., Batlle Martín, M., et al.: The Destination Earth digital twin for climate change adaptation, EGUsphere, 2025, 1–41, https://doi.org/10.5194/egusphere-2025-2198, 2025.

Donges, J. F., Donner, R. V., Rehfeld, K., Marwan, N., Trauth, M. H., and Kurths, J.: Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis, Nonlinear Processes in Geophysics, 18, 545–562, https://doi.org/10.5194/npg-18-545-2011, 2011.

Drijfhout, S., Gleeson, E., and Dijkstra, H. A.: Catalogue of abrupt shifts in multi-model climate simulations, Proceedings of the National Academy of Sciences, 112, E5777–E5786, https://doi.org/10.1073/pnas.1511451112, 2015.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Dunne, J. P., Hewitt, H. T., Arblaster, J., Bonou, F., Boucher, O., Cavazos, T., Durack, P. J., Hassler, B., Juckes, M., Miyakawa, T., et al.: An evolving Coupled Model Intercomparison Project phase 7 (CMIP7) and Fast Track in support of future climate assessment, EGUsphere, 2024, 1–51, 2024.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise, in: kdd, 34, pp. 226–231, https://doi.org/10.1145/23680.23683, 1996.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of CMIP6 experimental design and organisation, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., et al.: Earth System Model Evaluation Tool (ESMValTool) v2. 0–an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geoscientific Model Development, 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Fader, M., Rost, S., Müller, C., Bondeau, A., and Gerten, D.: Virtual water content of temperate cereals and maize: Present and potential future patterns, Journal of Hydrology, 384, 218–231, https://doi.org/10.1016/j.jhydrol.2009.12.011, 2010.

Feldmann, J. and Levermann, A.: Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin, Proceedings of the national academy of sciences, 112, 14 191–14 196, 2015.

Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aälgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L., Golledge, N. R., Hemer, M., Kopp, R. E., Krinner, G., Mix, A., Notz, D., Nowicki, S., Nurhati, I. S., Ruiz, L., Sallée, J.-B., Slangen, A. B. A., and Yu, Y.: Ocean, Cryosphere and Sea Level Change, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.011, 2021.

Fred, A. L. and Jain, A. K.: Combining multiple clusterings using evidence accumulation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 835–850, https://doi.org/10.1109/tpami.2005.113, 2005.

Frieler, K., Lange, S., Piontek, F., Reyer, C. P., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., et al.: Assessing the impacts of 1.5 °C global warming–simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b), Geoscientific Model Development, 10, 4321–4345, https://doi.org/10.5194/gmd-10-4321-2017, 2017.

Garbe, J., Albrecht, T., Levermann, A., Donges, J. F., and Winkelmann, R.: The hysteresis of the Antarctic Ice Sheet, Nature, 585, 538–544, https://doi.org/10.1038/s41586-020-2727-5, 2020.

Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M.: HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, The Astrophysical Journal, 622, 759, https://doi.org/10.1086/427976, 2005.

Hausfather, Z.: An assessment of current policy scenarios over the 21st century and the reduced plausibility of high-emissions pathways, Dialogues on Climate Change, 2, 26–32, 2025.

Högner, A., Di Capua, G., Donges, J. F., Donner, R. V., Feulner, G., and Wunderling, N.: Causal pathway from AMOC to Southern Amazon rainforest indicates stabilising interaction between two climate tipping elements, Environmental Research Letters, 20, 074 026, https://doi.org/10.1088/1748-9326/addb62, 2025.

Hoyer, S. and Hamman, J.: xarray: N-D labeled Arrays and Datasets in Python, Journal of Open Research Software, 5, 10, https://doi.org/10.5334/jors.148, 2017.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Huang, Y., Bathiany, S., Ashwin, P., and Boers, N.: Deep learning for predicting rate-induced tipping, Nature Machine Intelligence, 6, 1556–1565, https://doi.org/10.1038/s42256-024-00937-0, 2024a.

Huang, Y.-J., Chang, C.-W., and Hsieh, C.-h.: Detecting shifts in nonlinear dynamics using empirical dynamic modeling with nested-library analysis, PLOS Computational Biology, 20, e1011 759, https://doi.org/10.1371/journal.pcbi.1011759, 2024b.

Karlbauer, M., Cresswell-Clay, N., Durran, D. R., Moreno, R. A., Kurth, T., Bonev, B., Brenowitz, N., and Butz, M. V.: Advancing Parsimonious Deep Learning Weather Prediction Using the HEALPix Mesh, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 021, https://doi.org/10.1029/2023ms004021, 2024.

Lam, S. K., Pitrou, A., and Seibert, S.: Numba: A llvm-based python jit compiler, in: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, pp. 1–6, https://doi.org/10.1145/2833157.2833162, 2015.

Lange, S.: Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset, Earth System Dynamics, 9, 627–645, https://doi.org/10.5194/esd-9-627-2018, 2018.

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., and Zhou, T.: Future Global Climate: Scenario-Based Projections and Near-Term Information, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 553–672, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.006, 2021.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the National Academy of Sciences, 105, 1786–1793, https://doi.org/10.1073/pnas.0705414105, 2008.

Levermann, A., Bamber, J. L., Drijfhout, S., Ganopolski, A., Haeberli, W., Harris, N. R., Huss, M., Krüger, K., Lenton, T. M., Lindsay, R. W., et al.: Potential climatic transitions with profound impact on Europe: Review of the current state of six 'tipping elements of the climate system', Climatic Change, 110, 845–878, https://doi.org/10.1007/s10584-011-0126-5, 2012.

Loriani, S., Donges, J., Garbe, J., Harteg, J., Sakschewski, B., and Winkelmann, R.: Revealing hidden tipping across scales, in preparation, forthcoming.

Lovejoy, T. E. and Nobre, C.: Amazon tipping point, 2018.

Marwan, N., Donges, J. F., Zou, Y., Donner, R. V., and Kurths, J.: Complex network approach for recurrence analysis of time series, Physics Letters A, 373, 4246–4254, https://doi.org/10.1016/j.physleta.2009.09.042, 2009.

Mercer, J. H.: West Antarctic ice sheet and CO2 greenhouse effect: a threat of disaster, Nature, 271, 321–325, 1978.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, https://doi.org/10.5555/1953048.2078195, 2011.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q.: A review and comparison of changepoint detection techniques for climate data, Journal of Applied Meteorology and Climatology, 46, 900–915, https://doi.org/10.1175/JAM2493.1, 2007.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP 8.5—A scenario of comparatively high greenhouse gas emissions, Climatic Change, 109, 33–57, https://doi.org/10.1007/s10584-011-0149-y, 2011.

Röhrich, L., Harteg, J., Loriani, S., and Donges, J. F.: A Python package for generating synthetic datasets with spatio-temporal shifts, in preparation, forthcoming.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Schaphoff, S., von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., et al.: LPJmL4–a dynamic global vegetation model with managed land–Part 1: Model description, Geoscientific Model Development, 11, 1343–1375, https://doi.org/10.5194/gmd-11-1343-2018, 2018.

Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, J., and Menary, M.: On the risk of abrupt changes in the North Atlantic subpolar gyre in CMIP6 models, Annals of the New York Academy of Sciences, 1504, 187–201, https://doi.org/10.1111/nyas.14659, 2021.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Terpstra, S., Falkena, S. K., Bastiaansen, R., Bathiany, S., Dijkstra, H. A., and von der Heydt, A. S.: Assessment of abrupt shifts in CMIP6 models using edge detection, AGU Advances, 6, e2025AV001 698, https://doi.org/10.1029/2025AV001698, 2025.

Tessore, N.: healpix: Hierarchical Equal Area isoLatitude Pixelation utilities, https://github.com/ntessore/healpix, gitHub repository, version 2025.1, 2025.

Tutte Institute for Mathematics and Computing: fast_hdbscan: Fast HDBSCAN implementation for low-dimensional Euclidean space, https://github.com/TutteInstitute/fast_hdbscan, gitHub repository, version 0.2.2, 2025.

van der Velden, E.: CMasher: Scientific colormaps for making accessible, informative and 'cmashing' plots, The Journal of Open Source Software, 5, 2004, https://doi.org/10.21105/joss.02004, 2020.

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Legutke, S., Schupfner, M., Wachsmann, F., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 ScenarioMIP ssp585, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.6705, 2019.

Winkelmann, R., Dennis, D. P., Donges, J. F., Loriani, S., Klose, A. K., Abrams, J. F., Alvarez-Solas, J., Albrecht, T., Armstrong McKay, D., Bathiany, S., Blasco Navarro, J., Brovkin, V., Burke, E., Danabasoglu, G., Donner, R. V., Drüke, M., Georgievski, G., Goelzer, H., Harper, A. B., Hegerl, G., Hirota, M., Hu, A., Jackson, L. C., Jones, C., Kim, H., Koenigk, T., Lawrence, P., Lenton, T. M., Liddy, H., Licón-Saláiz, J., Menthon, M., Montoya, M., Nitzbon, J., Nowicki, S., Otto-Bliesner, B., Pausata, F., Rahmstorf, S., Ramin, K., Robinson, A., Rockström, J., Romanou, A., Sakschewski, B., Schädel, C., Sherwood, S., Smith, R. S., Steinert, N. J., Swingedouw, D., Willeit, M., Weijer, W., Wood, R., Wyser, K., and Yang, S.: The Tipping Points Modelling Intercomparison Project (TIPMIP): Assessing tipping point risks in the Earth system, EGUsphere, 2025, 1–52, https://doi.org/10.5194/egusphere-2025-1899, 2025.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

Appendix A: ASDETECT Bias Correction

The original ASDETECT algorithm, as described by Boulton and Lenton (2019), exhibits positional bias due to its use of a centred segmentation approach. As the algorithm creates segments and centres them within the record, it produces a non-uniform alignment of segment boundaries and thus oversamples particular periods of the time series—most notably the midpoint. This leads to two major issues: (i) a spurious discontinuity near the centre of the detection signal, and (ii) systematic differences in apparent shift magnitude depending on where in the record a shift occurs. These effects are illustrated in Figure A1. A fully unbiased alternative would apply a sliding window for every segment length, but this becomes computationally prohibitive for high-resolution datasets. TOAD therefore adopts a more efficient *two-sided* segmentation scheme: for each segment length, the series is segmented once from the start and once from the end, gradients from both passes are accumulated, and the detection signal is normalised by the number of segments covering each time step. We further apply an edge correction to downweight apparent abruptness near the start and end of the series, where reduced sampling yields less reliable estimates. Together, these steps remove most positional bias at only twice the computational cost of the original centred implementation, which remains available in TOAD for completeness but is not recommended for typical use.

Appendix B: Supplementary Figures

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

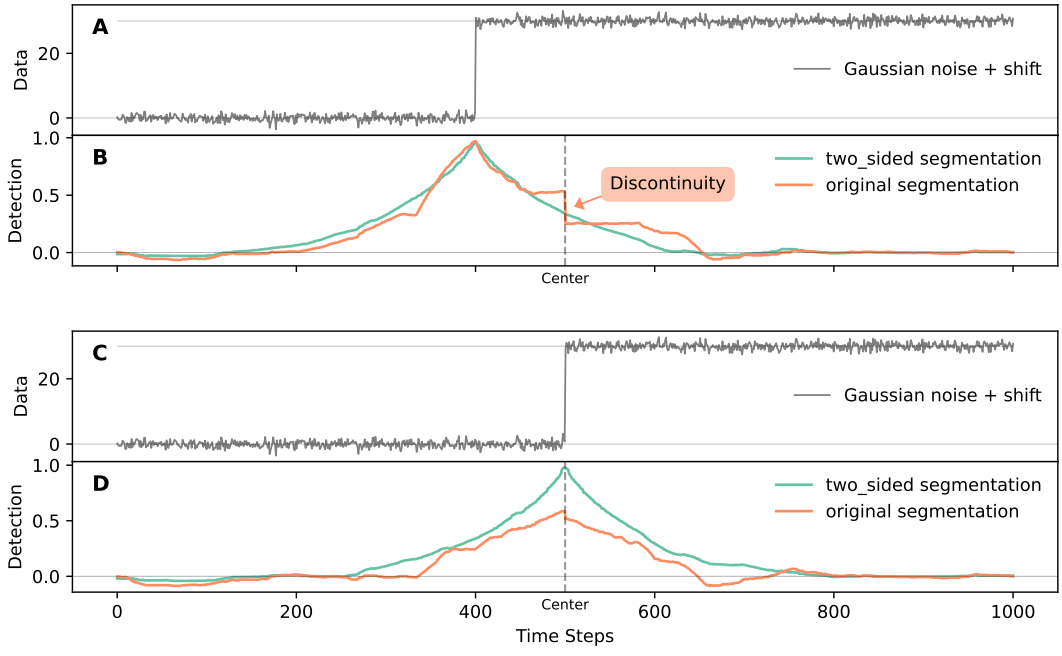

Figure A1. Comparison of centred and *two-sided* segmentation schemes applied to synthetic time series with a step-like shift and Gaussian noise. (A) Input data with a shift before the series centre (vertical dashed line). (B) Detection time series: centred segmentation (orange) introduces an artificial discontinuity at the centre, while *two-sided* segmentation (green) yields a smooth peak. (C) Input data with a shift at the series centre. (D) Detection time series: centred segmentation produces a damped signal that does not reach magnitude 1, while *two-sided* segmentation yields a single peak aligned with the true shift.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

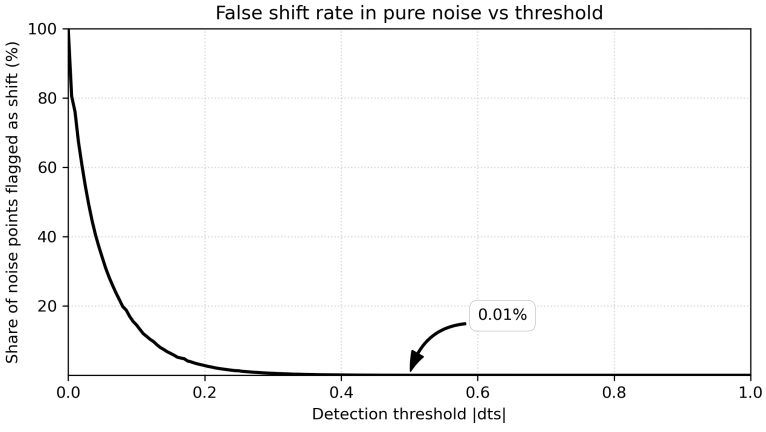

Figure B1. False-shift rate of ASDETECT as a function of the detection threshold ($|dts|$), estimated from Gaussian white-noise time series for a wide range of window lengths using *two-sided* segmentation. Even for purely noisy data, small thresholds (≤ 0.1–0.2) flag a substantial fraction of points as "shifts", whereas thresholds ≥ 0.5 reduce the false-shift rate below $\approx 0.01\%$. We therefore recommend $|dts| \geq 0.5$ as a conservative lower bound for the shift-selection threshold when using ASDETECT.

https://doi.org/10.5194/egusphere-2026-356
Preprint. Discussion started: 29 January 2026

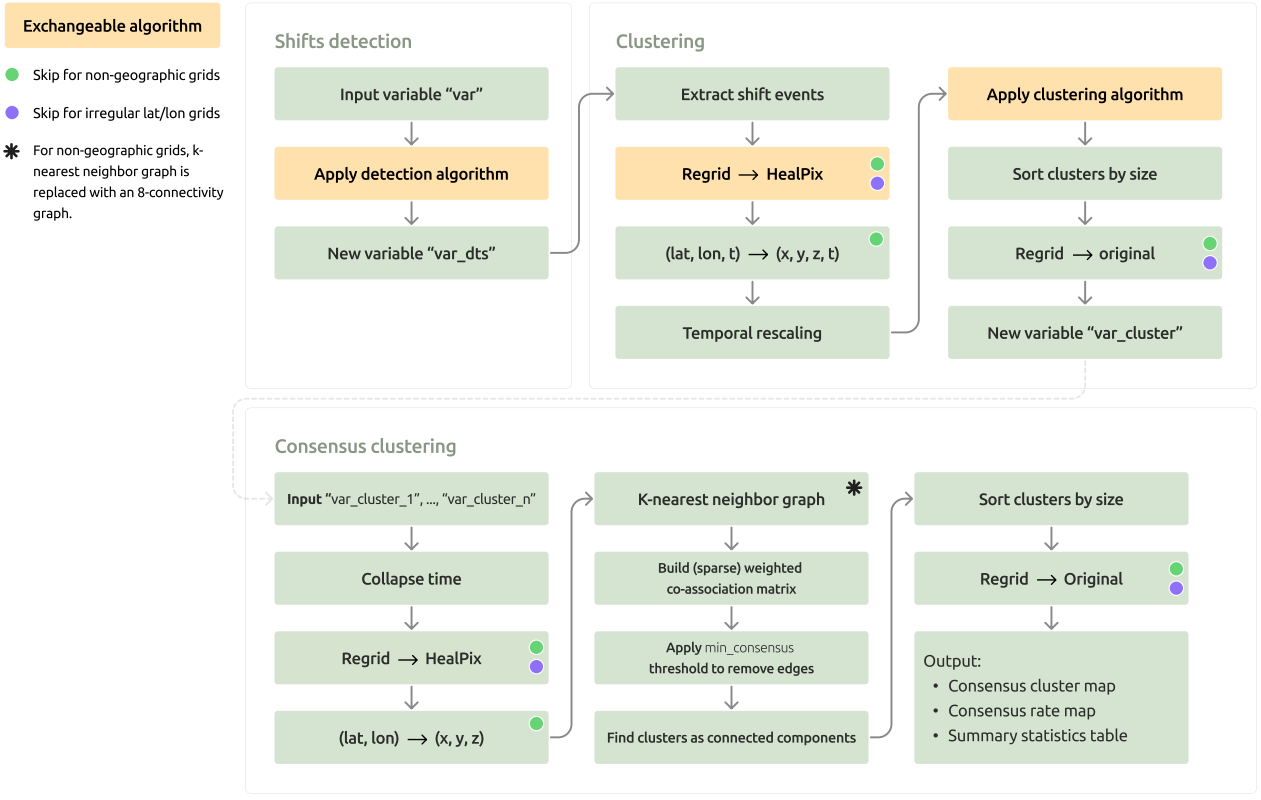

Figure B2. Detailed overview of the TOAD pipeline. In the shift detection stage, the chosen detector is applied to the input variable `var` to produce `var_dts`, a grid-cell detection time series. In the clustering stage, detected shift events are extracted, optionally regridded to an equal-area HEALPix grid for geographic datasets, embedded into Cartesian coordinates (x, y, z, t), rescaled, and clustered using the selected clustering algorithm; cluster labels are then mapped back to the original grid to form `var_cluster`. The consensus clustering shown here is an aggregation function that takes multiple cluster maps as input, collapses time, regrids and embeds coordinates into (x, y, z) space, constructs a k-nearest-neighbour (or 8-connectivity) graph, accumulates a sparse co-association matrix across inputs, applies a `min_consensus` threshold, and identifies connected components as consensus clusters. Outputs include a consensus cluster map, consensus-rate map, and a summary-statistics table. Coloured markers indicate operations that are skipped for non-geographic or irregular grids.