Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Defining Topological Weather Regimes through Persistent Homology

Marco Ezequiel Micieli

marco.micieli@cima.fcen.uba.ar

CIMA: Centro de Investigaciones del Mar y la Atmosfera https://orcid.org/0009-0001-2172-3350

Marisol Osman

CIMA: Centro de Investigaciones del Mar y la Atmosfera

Gisela Daniela Charó

CIMA: Centro de Investigaciones del Mar y la Atmosfera

Davide Faranda

LSCE: Laboratoire des sciences du climat et de l'environnement

Denisse Sciamarella

CNRS: Centre National de la Recherche Scientifique

Research Article

Keywords: topological data analysis, weather regimes, chaos topology, branched manifolds, persistent homology

Posted Date: October 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11120408/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Defining Topological Weather Regimes through Persistent Homology

Marco E. Micieli[1,2,3*], Marisol Osman[4,1,2], Gisela D. Charó[1,2,3], Davide Faranda[5], Denisse Sciamarella[2,3,6]

[1]Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Buenos Aires, Argentina.
[2]CONICET – Universidad de Buenos Aires. Centro de Investigaciones del Mar y la Atmósfera (CIMA). Buenos Aires, Argentina.
[3]Instituto Franco-Argentino de Estudios sobre el Clima y sus Impactos (IFAECI) – IRL 3351 – CNRS-CONICET-IRD-UBA, Buenos Aires, Argentina.
[4]Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Ciencias de la Atmósfera y los Océanos (DCAO). Buenos Aires, Argentina.
[5]LSCE-IPSL, CEA Saclay l'Orme des Merisiers, CNRS UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France.
[6]Centre National de la Recherche Scientifique, 75106, Paris, France.

*Corresponding author(s). E-mail(s): marco.micieli@cima.fcen.uba.ar;
Contributing authors: osman@cima.fcen.uba.ar;
gisela.charo@cima.fcen.uba.ar; denisse.sciamarella@cnrs.fr;

Abstract

The chaotic dynamics of the mid-latitude atmospheric circulation is often described using the statistical concept of Weather Regimes. Recent applications have sought to define these regimes as non-trivial topological features extracted through Persistent Homology, to make the result independent of ad hoc choices such as the number of clusters. Here, this perspective is revisited in the light of the topological description of chaotic attractors as branched manifolds. We define Topological Weather Regimes as specific parts of these structures, rather than as any non-trivial topological feature. Depending on whether they lie in the high- or the low-density subset of the points in phase space, they are computed

as zeroth- or first-order persistent homological features respectively. The statistical and topological protocols are applied independently, so that results can be compared. The approach is applied to two atmospheric datasets: the South-Atlantic and the North-Atlantic Polar-Front jets. The first one is characterized here for the first time within the weather-regime framework, while the second one has been thoroughly studied with clustering methods as well as with Persistent Homology. Each Topological Weather Regime is found to have its corresponding statistical match.

Keywords: topological data analysis, weather regimes, chaos topology, branched manifolds, persistent homology

1 Introduction

The meaning of the term *regime* varies with the discipline or the problem under analysis. In fluid mechanics, for instance, laminar and turbulent regimes correspond to low and high values of the Reynolds number taken as control parameter. In meteorology, a weather regime (WR) denotes a recurrent large-scale circulation pattern associated with preferred states of the atmospheric dynamics (Reinhold and Pierrehumbert 1982; Ghil and Lucarini 2020). The term *feature* in Topological Data Analysis (TDA) also covers objects of different nature, namely connected components and voids in a point set (Carlsson 2009; Edelsbrunner and Harer 2010). Asking topology to detect weather regimes therefore requires circumscribing what is meant by regime and by feature in each field, and bridging the gap between them. This work introduces a definition of Topological Weather Regime (TWR) that is in accordance with the particular sense that regime takes in the atmospheric sciences, and with the specific topological features that shape chaotic dynamics in phase space (Sciamarella and Charó 2024; Sciamarella 2026).

WRs account for much of the low frequency variability of the extratropical atmospheric circulation (Ghil and Robertson 2002). Owing to their quasi-barotropic structure (Blackmon et al. 1979), they typically persist longer than individual synoptic systems (Vautard 1990) and provide a useful framework for interpreting variability on subseasonal to seasonal timescales (Ghil et al. 2025). WRs have been extensively studied in both the North Atlantic and North American sectors (e.g., Vautard 1990; Robertson and Ghil 1999; Lee et al. 2023; Grams 2026), but remain comparatively unexplored over South America and the South Atlantic (Solman and Menendez 2003). Existing studies in this region are often closer to the concept of Weather Types, as they emphasize higher frequency variability within the circulation to environment framework of synoptic climatology (Yarnal and Draves 1993).

In practice, WRs are generally identified using clustering techniques applied to a reduced phase space obtained through empirical orthogonal functions (EOFs) or related dimensionality reduction methods (Mo and Ghil 1988). This approach has been highly successful and forms the basis of most operational and research applications. Nevertheless, its interpretation remains largely statistical. Clustering methods

require several ad hoc choices, including the number of clusters retained, which may influence the resulting regime structure (Christiansen 2007). Moreover, the commonly invoked recurrence hypothesis appears to hold more robustly for blocking configurations than for other circulation states (Hochman et al. 2021; Lee and Messori 2024; Legras and Ghil 1985). These limitations have motivated renewed interest in whether WRs correspond to genuine dynamical structures of the atmospheric flow rather than merely statistical aggregates of circulation patterns (Faranda et al. 2017).

The dynamical-systems perspective on regimes is long-standing (Ghil and Robertson 2002), and laboratory experiments reproducing transitions between blocked and zonal flow have provided physical evidence for their reality (Weeks et al. 1997). In this context, TDA, a family of methods that study the shape of data, was recently used to study WRs (Strommen et al. 2023). In particular, homology groups are designed to detect topological features at different orders: zeroth-order homologies (H_0) count connected components, first-order homologies (H_1) count holes, and higher-order homologies count cavities or hypercavities. To compute them from point sets, a scale-dependent construction is introduced, by connecting points within a certain distance by an edge, and filling in resulting triangles. As this distance grows, one obtains a nested sequence of constructions and one can study how a certain feature appears at multiple scales. Those features persisting, not through time, but across scales, are considered significant, while those that do not, are considered spurious features or sampling artifacts, often referred to as topological 'noise'. The computational tool is known as Persistent Homology (PH) (Edelsbrunner and Harer 2010). It was conceived for data in general, with image analysis among its most popular applications. The method does not establish which of its features matter, or for what: that decision rests with the user.

More specifically, Strommen et al. (2023) responded to the lack of an agreed definition of a regime by equating the existence of regimes with non-trivial PH features of the attractor, computed along distance and density scales (Carlsson and Zomorodian 2009). The equivalence removes the need to prescribe a number of clusters, and it applies to systems of very different kinds. Applied to the North Atlantic Polar-Front Jet (PFJ), it recovers two of the three known jet regimes. The southern one, identified by Woollings et al. (2010) and linked with the negative phase of the North Atlantic Oscillation (Madonna et al. 2017), does not come out as a non-trivial persistent feature. Replacing the density with a centrality function in the bifiltration, Anbouhi (2026) obtains a connected component consistent with the southern jet, a detection that the author qualifies as marginal, with few points in the component. These results raise the following question: can PH capture all the dynamically relevant structures underlying the large-scale atmospheric circulation? Answering it requires looking at how a chaotic dynamics is organized in phase space.

Chaotic attractors are not arbitrary point clouds in phase space. Rather, they are organized around objects called branched manifolds (Williams 1979; Birman and Williams 1983), i.e. Euclidean manifolds almost everywhere except at singular sites called junction loci where different branches merge. These junctions play a central role in organizing trajectories and are responsible for the characteristic stretching, folding, tearing and squeezing mechanisms of chaotic dynamics (Gilmore 1998; Gilmore and

Lefranc 2012). There is no chaos without branches that merge at junctions: holes, in contrast, may vary from one attractor to another and can also be present for regular dynamics, in the absence of chaos. The best-known example of chaos is provided by the Lorenz attractor (L63; Lorenz 1963). Standard homology detects a single connected component for the butterfly-shaped structure as a whole, and a void at the center of each wing, but it is not designed to detect the junctions across which trajectories repeatedly pass before switching from one wing to the other. Branched Manifold Analysis through Homologies (BraMAH; Sciamarella and Mindlin 1999, 2001) is an approach rooted in Chaos Topology that was specifically developed to characterize phase-space point clouds by assembling local manifold patches into a cell complex from which both the homologies and the junctions can be extracted, the latter from the way the boundaries of the cells intersect. This methodology has been successfully applied in several cases (Charó et al. 2020, 2021), and even extended beyond homologies (Charó et al. 2022; Ghil and Sciamarella 2023; Sciamarella and Charó 2024) to define topological modes of variability in terms of the directed algebraic topology of the branched manifold (Charó et al. 2025; Sciamarella 2026).

This work addresses the challenge of defining WRs through PH by drawing on what chaos topology has established about the branched organization of chaotic attractors (Williams 1979; Birman and Williams 1983). PH remains the computational tool and density remains a parameter of the analysis, as in Strommen et al. (2023), but here we ask a WR to correspond to an identifiable part of a branched manifold. Junctions and holes are associated with different features of a topological structure, and correspond to opposite density ranges. Retaining the densest subset of a point cloud and inspecting its connected components, one can identify the most visited parts of the manifold. Conversely, searching for holes in the least dense subset, one can identify the parts that surround voids. The two ends of the density range must therefore be analyzed separately to detect different features, rather than searched across. This leads to the definition of two classes of TWR. All the weather regimes are first computed statistically and then topologically. The two procedures are run independently, and found to return matchable results. The topological route retains the advantage that no number of regimes is fixed beforehand. The TWRs recovered do not, taken together, reconstitute the original point cloud as a whole. Describing the transitions between TWRs is left for future work.

The paper is structured as follows. Section 2 introduces the datasets, namely the Lorenz case and the South and North Atlantic PFJs. Section 3 presents the concepts underlying the topological approach, introduces a trajectory-based diagnostic to detect junctions without the prior construction of a cell complex, and states the definition of a TWR. Section 4 reports the results, with the statistical identification of the regimes coming first and the topological one second, for the Lorenz dataset, for the South Atlantic PFJ, whose wintertime regimes are characterized here for the first time, and for the North Atlantic, where all three jet regimes are recovered. Section 5 presents the conclusions.

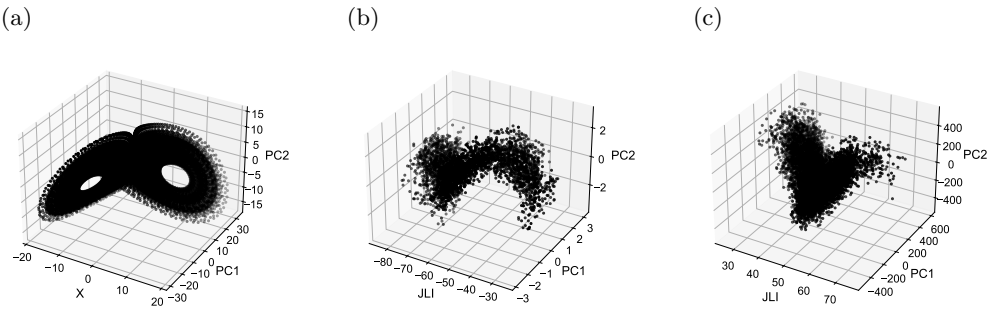


Fig. 1: Phase-space reconstructions for (a) the L63-R dataset with X and the first two principal components (PC1, PC2) of the ψ and T^* fields, (b) the JET-LAT-SA dataset with the Jet Latitude Index (JLI) and the first principal components (PC1, PC2) of $\hat{u}'_{low}$, and (c) the JET-LAT-NA dataset defined by Strommen et al. (2023).

2 Data

2.1 Lorenz-attractor

The attractor studied by Lorenz (1963) is a truncation of the Boussinesq equations for Rayleigh-Bénard convection, the buoyancy-driven motion of a fluid layer heated from below. Let us consider the mode decomposition for the stream function ψ and the temperature departure T^* from the non-convective state

$$\psi(x,z,t) \sim X(t)\sin(\pi x)\sin(\pi z)\,, \tag{1}$$

$$T^*(x,z,t) \sim Y(t)\cos(\pi x)\sin(\pi z) - \frac{1}{2}Z(t)\sin(2\pi z)\,, \tag{2}$$

A Galerkin projection retaining only these leading Fourier modes yields:

$$\dot{X} = \sigma(Y - X), \qquad \dot{Y} = X(\rho - Z) - Y, \qquad \dot{Z} = XY - \beta Z. \tag{3}$$

With the classical parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, the dynamics of this system is known to be chaotic.

We first generate a time series of 20000 points and $dt = 1.5 \times 10^{-2}$ integrating (3) using the Explicit Runge-Kutta scheme of order 5(4) and we discard the first 2000 time steps to remove the initial transient. After substituting the result into (1) and (2), one can obtain the time evolution of the ψ and T^* fields, subsequently retaining the leading principal components (PCs) associated with the joint variability of both variables.

To establish a parallel with the atmospheric cases that follow, we reconstruct the phase space using the first two PCs (PC1, PC2), alongside the state variable X, rather than the first three PCs (which yields equivalent results). This choice is illustrative: given its coupling to the circulation dynamics, X serves as an intuitive analog to the jet-latitude index used in the subsequent subsections. This reconstructed point cloud is labeled L63-R and shown in Figure 1(a).

2.2 South-Atlantic PFJ

Let us consider the subseasonal winter variability of the PFJ over southern South America and the South Atlantic Ocean. Figure 2 shows the domain defined for the analysis (85°W-25°W, 85°S-25°S), which excludes the Pacific branch of the PFJ, as the Andes Mountains significantly influence its behavior and generate strong zonal asymmetries.

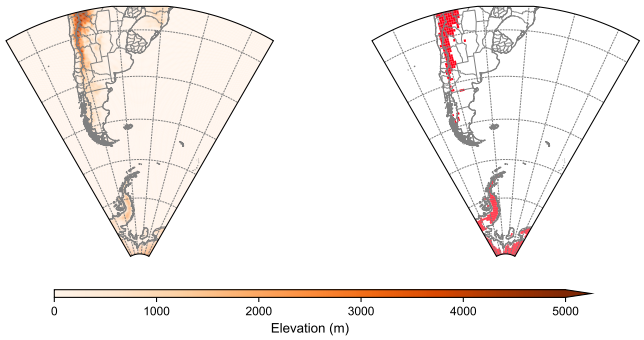


Fig. 2: For the study region: (a) ERA5 terrain elevation (shading, m) and (b) mask of points associated with areas with elevated topography (red).

The analysis employs daily-mean data of zonal wind at 800, 850 and 900 hPa levels from the fifth version of the climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ERA5; Hersbach et al. 2020), obtained by combining observational information with short-term forecasts from numerical atmospheric model simulations. The study period covers 1981-2020, focusing on the austral winter season (June-July-August). The analysis is performed at low levels due to the quasi-barotropic configuration that the PFJ acquires through the action of transient eddies accelerating the flow near the surface (Hoskins et al. 1983). This distinguishes the PFJ from the Subtropical Jet at high levels, resulting from the conservation of angular momentum (Held and Hou 1980). The distinction between these two jets over the Southern Hemisphere is clearly evident only in autumn and winter (Gallego et al. 2005). The regions with complex topography were masked by comparing the ERA5 terrain elevation and the mean daily geopotential height at 850 hPa (z_{850}). Specifically, those points where z_{850} was lower than the elevation on at least one day were masked.

To retain low-frequency variability, pre-processing was applied to the low-level zonal wind data for each grid point independently. First, the annual mean cycle was removed using a smoothed climatology based on a 61-day centered running mean for each calendar day. Subsequently, a low-pass Lanczos filter with a 61-day running window and a 10-day cutoff period was applied. The advantage of this filter over the Discrete Fourier Transform is its ability to significantly reduce Gibbs oscillations (Duchon 1979). Hereafter, subseasonal anomalies of any variable v resulting from these two steps are denoted as v'.

The long-term linear trend, calculated via the non-parametric Sen's slope estimator to ensure robustness against outliers, is then subtracted from these smoothed anomalies. These trends are statistically significant over almost the entire domain, according to the Mann-Kendall test with a 95% confidence level (not shown). Finally, the resulting anomalies are normalized by the latitude-weighted domain-averaged standard deviation, smoothed with a 61-day centered running mean for each calendar day, ensuring that all seasons contribute equally to the variance (Lee et al. 2023; Grams 2026). These normalized anomalies for a variable v are denoted as $\hat{v}'$.

Following Strommen et al. (2023), we construct a three-dimensional phase-space representing the subseasonal variability of the PFJ within the study region. The first variable is a daily index describing its latitudinal position, computed by zonally averaging the fields over all domain longitudes for each day and identifying the latitude of the maximum wind speed on the profile. This methodology, introduced by Woollings et al. (2010), is known as the Jet Latitude Index (JLI). Characterizing variability through the latitudinal position of the jet continues a long tradition that goes back to the index cycle of Namias (1950). In this study, the JLI is defined using $\hat{u}'_{850}$ at a 0.5°×0.5° resolution. To verify its robustness, this procedure was repeated at the same pressure level using the first version of the climate reanalysis from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP-NCAR; Kalnay et al. 1996) and the second climate reanalysis from the Japan Meteorological Agency (JRA55; Kobayashi et al. 2015), maintaining the original spatial resolution of the available associated datasets. The second and third variables are the principal components (PCs) associated with the first and second Empirical Orthogonal Functions (EOFs) of the vertically averaged normalized zonal wind anomalies ($\hat{u}'_{low}$), which explain approximately 50% of the total variability of this variable. The spatial resolution of $\hat{u}'_{low}$ was upscaled to 2.5°×2.5° to capture large-scale variability within the domain. These PCs also contain information on the intensity of the PFJ. This three-dimensional dataset is hereafter referred to as JET-LAT-SA and shown in Figure 1(b).

Figure 3 presents the JLI distribution estimated via Kernel Density Estimation (KDE) (Silverman 2018) from the three employed climate reanalyses for each season. A trimodal behavior is observed during winter (JJA) and particularly in spring (SON), whereas this is less evident in summer (DJF). In autumn (MAM), the JLI exhibits a distinctive bimodal pattern, with the equatorward peak almost completely disappearing. The central and poleward peaks remain well-defined throughout the year, with the central peak being most pronounced in summer and autumn. The reanalyses agree regarding the observed behavior for each season. For this reason, from now on only the results obtained for ERA5 will be shown and described.

2.3 North-Atlantic PFJ

This three-dimensional dataset (JET-LAT-NA hereafter) is similar to JET-LAT-SA and is obtained using u_{850} daily-mean data from the ERA20C reanalysis (Poli et al. 2016) over the North Atlantic-European domain (60°W-0°, 15°N-75°N) for the boreal wintertime (DJF) in the period 1901-2010. For details on this dataset the reader can refer to Strommen et al. (2023). JLI computation in this case follows the simplified

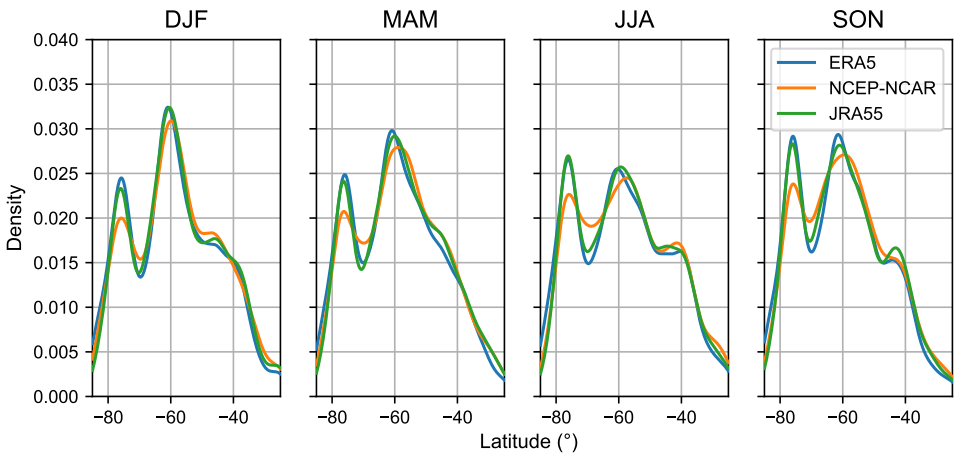


Fig. 3: South Atlantic JLI distribution estimated with KDE from ERA5 (blue), NCEP-NCAR (orange), and JRA55 (green) in summer (DJF), autumn (MAM), winter (JJA) and spring (SON) in the period 1981-2020.

approach proposed by Parker et al. (2019). The phase space we feed into the analysis (Fig. 1c) consists of the same files as provided by Strommen et al. (2023) at https://github.com/KristianJS/BifiltPH.

3 Methodology

The statistical protocol applied for the JET-LAT-SA and L63-R datasets is standard practice and is recalled here for reference; the rest of this section develops the topological protocol. For details on the statistical approach for the North-Atlantic PFJ regimes, the reader can refer to Woollings et al. (2010).

3.1 Statistical protocol

A k-means clustering (Lloyd 1982) is applied to the standardized point clouds so that all variables have the same range of variability. This partitioning method distributes the data into a predetermined number of non-overlapping clusters by minimizing the intra-cluster variance. Originally, k-means forces all cloud points to reside in a specific cluster. However, we follow a more restrictive criterion (Dorrington et al. 2022) for the case of the JET-LAT-SA dataset. Specifically, we categorize daily fields of $\hat{u}'_{low}$ exhibiting a latitude-weighted correlation of less than 0.4 with their respective cluster's average pattern as unassigned or part of a neutral state. Once the clusters are defined, an event is a period of consecutive times assigned to the same statistical regime. To link the PFJ regimes to the hemispheric annular state, we use the daily Antarctic Oscillation (AAO) index in the period 1981-2020 during austral winter provided by the National Oceanic and Atmospheric Administration (NOAA) Climate

Prediction Center (CPC) at https://psl.noaa.gov. A ±0.5 standard deviation threshold on this index defines the positive and negative AAO phases, with intermediate values considered neutral.

3.2 Topological protocol

Topological data analysis is dominated by homologies, which constitute a set of topological invariants, indexed by dimension, that allow distinguishing and classifying regular manifolds through the detection of holes of different types, such as connected components, called 0-holes, non-trivial loops, called 1-holes, cavities called 2-holes, and higher-dimensional voids. To turn homology into a computation, the manifold is first decomposed into a collection of vertices or 0-cells, segments or 1-cells, filled polygons or 2-cells and higher-dimensional cells glued along their faces. For a given cell complex, one can then introduce a boundary operator and an algebra to compute l-holes, with $l = 0, 1, \ldots$ (see Appendix A). The number of holes at each level l depends on how the cells are connected and not on their number and size (Kinsey 2012).

3.2.1 Branched manifolds

Chaotic attractors are known to lie on branched manifolds, which are regular manifolds except at junction loci, where different branches meet (see Appendix B). These 'joints' play an important role in the organization of recurrent trajectories in phase space, but are not necessarily captured with a homological analysis. For instance, the recurrent dynamics on a toroidal surface present the homologies of a torus, regardless of whether the dynamics is regular or chaotic on the torus (Mosto et al. 2024). A cell complex for which the underlying space is a branched manifold can however be designed to extract the junction(s) directly from cell intersection properties. This type of cell complex is called a BraMAH complex and is described in Appendix B.

Figure 4(a) displays a BraMAH cell complex constructed for a Lorenz attractor. Zeroth-order homology captures the single connected component of the entire structure; first-order homology captures the one-dimensional holes at the center of each wing. There are no cavities enclosed ($H_2 = 0$). If l-holes can be computed using homologies, the junction locus can be detected taking into account that three or more top-cells of a BraMAH complex share a common boundary where branches merge. In the Lorenz attractor, the 1-cells $\langle 0, 1\rangle$ and $\langle 1, 2\rangle$ form the junction locus.

3.2.2 Persistent homology

For generic point clouds, scale-dependent methods have become very popular: instead of choosing a single cell decomposition, a family of cell complexes is built based on a proximity scale parameter ε defined as the distance between points. As ε varies monotonically, it generates a nested sequence of growing complexes termed a 'filtration'. Figure 5(a)–(c) illustrates the Vietoris-Rips filtration on points sampled from a circle in a two-dimensional space. For $\varepsilon = 0$, the complex is formed by many 0-cells without 1-cells or 2-cells; at an intermediate scale $\varepsilon = 0.65$ the 0-cells are assembled into a cell complex having both 1-cells and 2-cells forming one connected component and enclosing a central void; at a large scale $\varepsilon = 1.60$, higher-dimensional cells fill

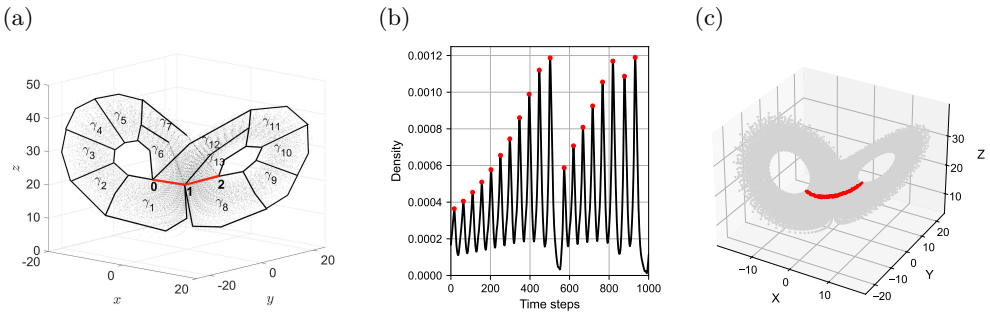


Fig. 4: (a) BraMAH complex describing the branched manifold underlying the L63 attractor. The red 1-cells $\langle 0, 1\rangle$ and $\langle 1, 2\rangle$ mark the junction locus. Trajectory-based diagnostic to detect junctions without the prior construction of a cell complex: (b) Local density peaks (red) plotted against time along the trajectory on the Lorenz attractor. (c) The red points corresponding to the local density peaks in (b) allow locating the junction locus within the gray point cloud.

the central void. Homological features consequently emerge and collapse as a function of ε. Figure 5(d) displays the persistence diagram showing the 'birth' and 'death' of the features across the filtration. In this diagram, symbols are used to mark the (birth,death) coordinates of ε values at which 0-holes (circular marker) or 1-holes (triangular marker) appear/disappear respectively. Their 'persistence', i.e. the difference between the death and birth values, denotes the 'lifetime' of the homological feature. To compensate for the fact that homologies under this framework are no longer independent of the specific cell decomposition, since they now depend strictly on the filtration parameter ε and the pairwise distances between data points, the concept of persistence is introduced as a proxy for topological invariance. Features with a large persistence survive across a wide range of scales and are considered to represent the authentic topological properties of the underlying space approximated by the point cloud, regardless of whether this space is a regular or a branched manifold. The dashed horizontal line in the plot denotes the so-called persistence horizon (persistence = 1) used for features that are still 'alive' at the largest ε scale.

To locate a persistent feature in the point cloud under analysis, one can also compute 'representatives'. Persistence records only the spatial scales at which a feature emerges and collapses: the feature itself is a homology class, an algebraic object that certifies the existence of a hole without singling out any specific points. A representative cycle is one concrete member of that class, mapped onto the data. For a 1-hole, it is a closed sequence of edges joining data points that encircles the detected void; for a 0-hole, it reduces to the set of points forming the connected component. Representatives thus translate a topological feature into a subset of phase-space states. A representative is not unique: two cycles of the same class differ by a boundary, that is, they can be deformed into one another across sampled regions (see Appendix A), and the computed cycle is a particular choice among them. This non-uniqueness, together with the sensitivity to threshold settings, motivates the tracking procedure

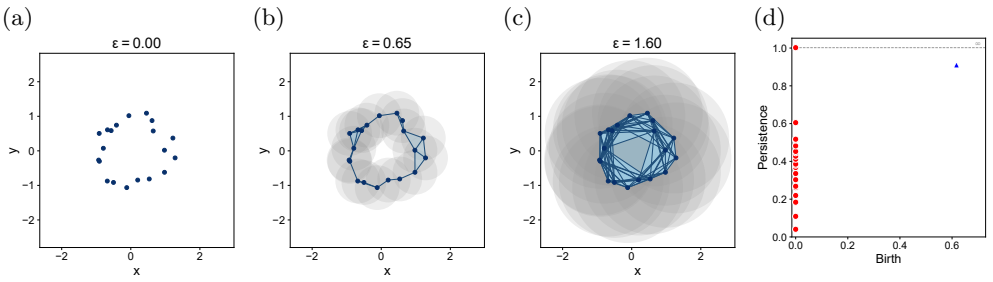


Fig. 5: Vietoris-Rips filtration of a circular point cloud at three progressive scales of the distance parameter ε. (a) $\varepsilon = 0.00$: Initial point cloud representing disconnected H_0 components. (b) $\varepsilon = 0.65$: Formation of a persistent 1-hole (H_1). (c) $\varepsilon = 1.60$: Occlusion and death of the central void due to the configuration of higher-dimensional cells. (d) Persistence diagram for the H_0 (red circles) and H_1 (light blue triangles) homological features of the example point cloud. Persistence values equal 1 for features that are still 'alive' at the largest ε scale.

of Appendix C. Notice, however, that there is no persistent feature related to the existence of a junction in a branched manifold underlying a dataset.

3.2.3 The role of density

Density constitutes a fundamental variable for WR detection through PH, and it can be handled in more than one way. Strommen et al. (2023) organized the analysis as a distance-density bifiltration: nested subsets retaining the densest $10\%, 20\%, \ldots$ of the cloud are formed, a distance filtration is computed on each, and the system is declared to have non-trivial topology if some feature, at some density threshold, outlives the level expected from Gaussian noise. Regimes are then read off from the non-trivial features. The protocol adopted here also relies on density estimation: the density distribution of the point cloud is approximated through a multidimensional KDE $\hat{f}$ with Gaussian kernels, the bandwidth being set by Silverman's rule of thumb for standardized data (Silverman 2018, Sec. 4.3.2).

Before considering any persistent homological features of the point cloud, let us focus on a metric property concerning local density for point clouds with an underlying branched-manifold structure. In the regions of the point cloud where branches merge, local density is expected to rise due to the superposition of trajectories belonging to merging dynamical paths. Local-homology approaches can detect singular strata from samples (Stolz et al. 2020), at the cost of assembling a local complex around each point, but the diagnostic introduced here to detect junctions requires only a temporal ordering of the points $\mathbf{x}_i$ in the point cloud inherited from the discrete sampling of the trajectory: taken in this order, the values $\hat{f}(\mathbf{x}_i)$ trace the local density along each contiguous segment of the trajectory in phase space. The ordering need not span a single uninterrupted trajectory: time-ordered segments suffice, as in observational records organized in seasonal blocks. Where the trajectory crosses a set on which several branches overlap, the local density displays a narrow maximum with a sharp

peak. These spikes differ from the broader maxima that may be produced by longitudinal or transverse squeezing, where the flow contracts into a narrower region; in either case the density rises and falls gradually. Instead, when branches merge, they do so on a subset of phase space of lower dimension than the branched manifold itself (see Appendix B), so the density rises and falls abruptly as the trajectory traverses it. The reference scale for sharpness is the residence time in the region: a junction concentrates the rise and fall of the density within a few consecutive samples, while metric maxima evolve on the timescale of the transit itself. The coordinates of the peaks locate the junction as a set of points in phase space. The application of this diagnostic to a Lorenz-attractor point cloud is shown in Figure 4(b) and (c). Notice the correspondence of the red points in (c) with the 1-cells $\langle 0, 1\rangle$ and $\langle 1, 2\rangle$ of the BraMAH complex shown in (a).

Let us now turn back to the detection of weather regimes from the topological analysis of a phase-space point cloud. If the point cloud is structured as a branched manifold, its densest subsets, which may partition into multiple disconnected components, are expected to contain junctions. Each such component corresponds to a highly visited part of the manifold: whether this involves a junction locus, or only longitudinal or transverse squeezing, can be examined with the trajectory-based junction diagnostic. On the other hand, the holes in the point cloud will encircle regions in phase space which remain unvisited. The role of density can hence be used to partition the full point cloud into three subsets corresponding to different density ranges: a high-density range targeting squeezing or confluence zones detected as connected components, a low-density range aimed at identifying voids, and an intermediate-density range that is excluded from the analysis to keep the two ends apart.

These considerations lead to a definition of topological weather regimes that assigns each homological order to one specific end of the density spectrum in advance. Zeroth-order homologies (H_0) will be computed exclusively for the high-density range, while first-order homologies (H_1) will be computed for the low-density range. This methodology of partitioning a point cloud to separately detect low- and high-density features via PH will hereafter be referred to as BraMAPH (Branched Manifold Analysis through Persistent Homology). Within BraMAPH, the trajectory-based diagnostic completes the analysis of the high-density end, searching for the signature of a junction.

3.2.4 Detection of topological WRs

We define a TWR as a part of the branched manifold, i.e. as a region of phase space that the point cloud samples. The arguments above allow defining topological WRs operationally.

- A zeroth-order topological WR or **0-TWR** is detected as a persistent connected component of the high-density subset, by computing the persistent homologies of order 0. The trajectory-based diagnostic then applies to the component. The onset of the regime is the set of times at which the trajectory enters the component within each continuous segment of the record. A maximum of the local point density in

the temporal neighborhood of the onset, confined to a few consecutive samples, is read as the signature of a junction.

- A first-order topological WR or **1-TWR** is detected through a persistent 1-hole of the low-density subset, by computing persistent homologies of order 1. The persistent 1-holes single out candidate voids, and their representative cycles map these candidates onto the original point cloud. The artificial void generated by the excision of the dense core is located and set aside. Representative cycles are computed, and this is done across neighboring density thresholds within the low-density range, which ensures that the feature is not an artifact of a single threshold. For further details see Appendix C.

Every quantity conditioned on a TWR in what follows (conditional distributions, onsets, circulation composites, day assignments) is computed over the member points of the regime: for a 0-TWR, the points of its connected component at a reference high-density range, specified for each dataset; for a 1-TWR, the points on its tracked representative cycles, which sample the region that surrounds the void.

All PH features are here computed with the Ripser C++ library (Bauer 2021), and persistence is deemed significant above the 0.4 threshold adopted from Strommen et al. (2023), who obtained it as the maximum persistence achieved across 20 independent realizations of a standard multivariate Gaussian distribution; since the point clouds are standardized, the threshold applies without rescaling.

Both classes of TWR qualify as WR candidates because both realize in phase space the temporal persistence and recurrence that the meteorological notion of regime demands. In the high-density case, trajectories repeatedly return to the region that a component occupies. In the low-density case, trajectories repeatedly return to the region that surrounds a void. A 1-TWR is not the void itself but the region that surrounds it. Transit through a component and encirclement of a void are two forms of recurrence, told apart by the order of the homology and the density range at which it is computed.

4 Results

4.1 Statistical results

Given the trimodal distribution of the JLI in the JET-LAT-SA dataset and the X variable in the L63-R dataset, shown in black in Figures 6 (a) and 6 (b) respectively, 3 clusters are defined in both cases. Indeed, various validation metrics also suggest that this is the optimal number (not shown). In that sense, the conditional distributions of the JLI and X by the defined clusters, illustrated with different colors in Figure 6 (a) and 6 (b), highlight a different peak from the original distribution of these variables. Based on the JLI values in each cluster, the statistical JET-LAT-SA WRs are named Southern Jet (green, J_S), Central Jet (red, J_C) and Northern Jet (blue, J_N). On the other hand, the statistical L63-R WRs describe relevant regions of the Lorenz attractor: the left wing (green, LW), the vicinity of the junction region (red, JU) and the right wing (blue, RW).

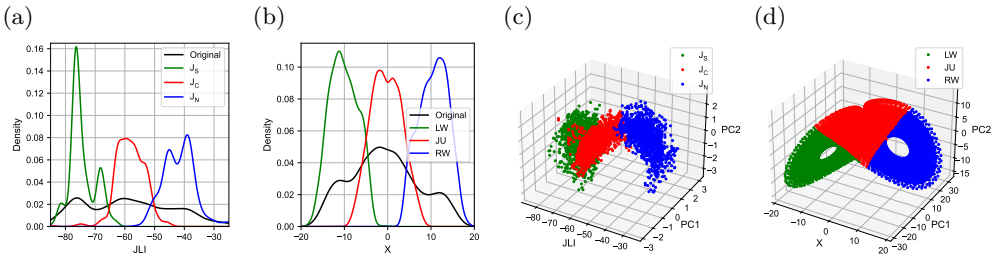


Fig. 6: (a) Climatological South-Atlantic JLI (ERA5) winter distribution (black line) and conditional distributions by the statistical JET-LAT-SA WRs (J_S, J_C, J_N) in the period 1981–2020. (b) Distribution of variable X for L63-R (black line) and conditional distributions by the statistical L63-R WRs (LW, JU, RW). Phase-space point clouds with colors corresponding to the k-means partition in 3 clusters applied to (c) JET-LAT-SA and (d) L63-R datasets.

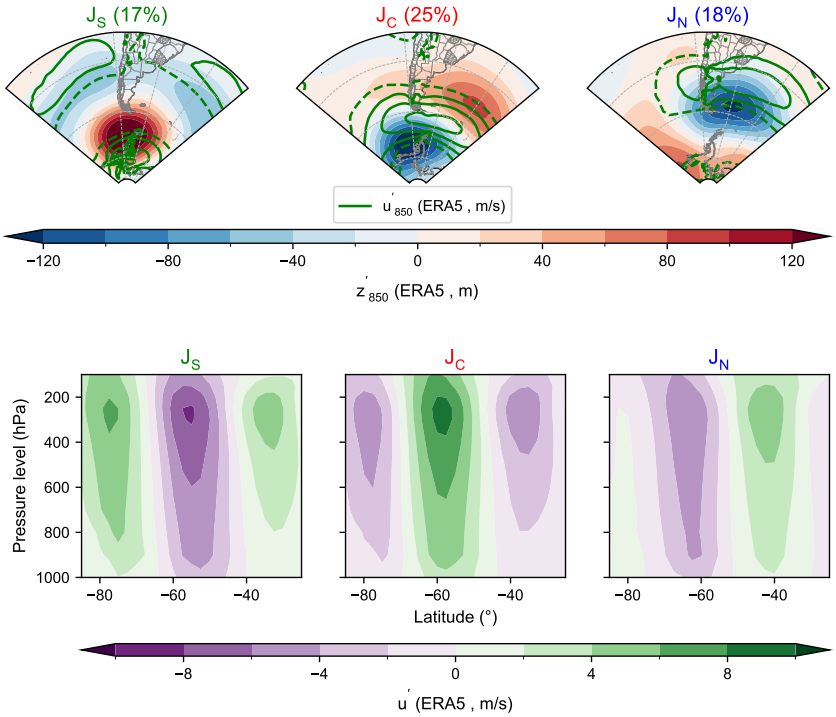


Fig. 7: Top row: z'_{850} (shading, m) and u'_{850} (positive values every $2\frac{m}{s}$ in solid contours; $0\frac{m}{s}$ isoline in dotted line) ERA5 composites (JJA 1981-2020) of all assigned days in J_S (left), J_C (center), and J_N (right). The climatological frequency of occurrence is displayed above each column. Bottom row: Vertically and zonally averaged u' (shading, $\frac{m}{s}$) ERA5 composites (JJA 1981-2020) between 85°W and 25°W for all assigned days in J_S, J_C, and J_N.

With the neutral state taken into account, Figure 7 (top row) illustrates the horizontal structure characterizing each statistical WR in the JET-LAT-SA dataset. As in the conditional JLI distributions in Figure 6 (a), the regimes differ primarily in the latitudinal position of the PFJ based on the location of the maximum values of u'_{850}. The most frequent regime is J_C (central panel) which features a zonally elongated region centered at 60°S with westerly wind anomalies up to 8 m/s. The z'_{500} field indicates this jet results from an anticyclonic circulation over the Atlantic and a cyclonic circulation over the Antarctic Peninsula. The second most common regime is J_N (right panel), characterized by a meridionally disturbed and less intense jet, with eastward anomalies up to 6 m/s displaced towards the equator, largely driven by a cyclonic circulation north of the Antarctic Peninsula. J_S (left panel) is the least frequent but most intense regime, with eastward anomalies up to 10 m/s restricted to the pole, resulting from an anticyclonic circulation over the Antarctic Peninsula. These results are robust: the spatial latitude-weighted correlation between the $\hat{u}'_{low}$ composites of these ERA5 regimes and those from NCEP-NCAR and JRA55 is close to 1 and temporal assignments across datasets are highly consistent (not shown). The zonal mean vertical distribution of positive zonal wind anomalies shown in Figure 7 (bottom row) confirms that all regimes exhibit a quasi-barotropic configuration extending practically to the surface.

The Southern Annular Mode, also referred to as the Antarctic Oscillation (AAO), describes the leading annular variability of the Southern Hemisphere extratropical circulation and its associated circumpolar westerlies. The PFJ regimes analyzed here can be interpreted as regional expressions of this broader annular variability, together with additional sectoral dynamical influences. Figure 8 (left panel) displays the climatological JLI distribution for ERA5 alongside the distributions conditioned by AAO phases. A considerable intensification (weakening) of the central (equatorial) maximum occurs during the positive AAO phase, and vice versa during the negative phase. The neutral phase distribution closely mirrors the climatology. This aligns with the occurrence frequencies of PFJ regimes per AAO phase, see Figure 8 (right panel). Specifically, J_C is significantly more (less) frequent during the positive (negative) AAO phase, and vice versa for J_N. In contrast, J_S is not strongly related to the AAO, displaying only a slight decrease in frequency during the negative phase. The southern regime is thus largely independent of the annular mode, and could instead be associated with blocking-like anticyclonic circulation at high latitudes, consistent with the anticyclonic center over the Antarctic Peninsula in Figure 7 (top row).

The boxplots of event durations in the study period for each statistical JET-LAT-SA WR are shown in Figure 9 (a). J_C is associated with events of longer duration than the other two, with a median of 6 days. J_S follows with a median of 5 days, while J_N has the lowest median of only 4 days. A similar behavior is observed in Figure 9 (b), which shows the distribution of event durations of the statistical L63-R WRs. In this case, the longest duration events occur in the vicinity of the junction region (*JU* regime), whereas the turns around the wings have shorter durations (*LW* and *RW* regimes).

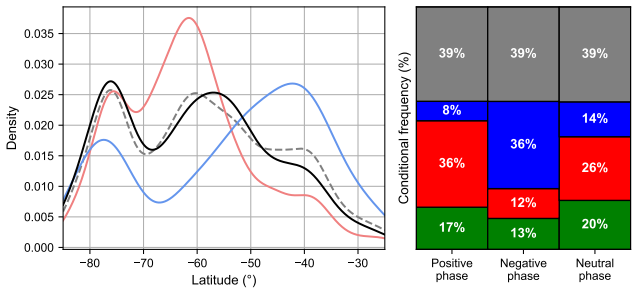


Fig. 8: (Left panel) Climatological (gray dotted line) and conditional (solid lines) South Atlantic JLI (ERA5) winter distributions (JJA 1981-2020) estimated with KDE, grouped by AAO phases; (right panel) Conditional occurrence frequency of statistical JET-LAT-SA WRs (bars) and unassigned states (gray bar) by AAO phase (JJA 1981-2020).

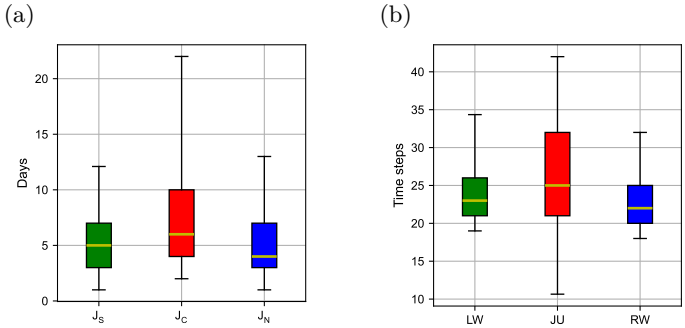


Fig. 9: Boxplots of event durations for statistical (a) JET-LAT-SA (days) and (b) L63-R (time steps) WRs. The yellow line is the median, boxes show the interquartile range and whiskers are defined using Tukey's rule.

4.2 Topological results

4.2.1 Lorenz-attractor case

Figure 10 illustrates the density-based decomposition of the L63-R point cloud. Based on this distribution, the point cloud is partitioned into a high-density subset (top 10%) and a low-density subset (bottom 60%). This decomposition separates the junction from the two wings, allowing each structure to be analyzed independently.

Figure 11 shows the ten most persistent topological features for the high and low-density ranges. In the high-density subset (Fig. 11a), a single dominant 0-hole stands out from all remaining connected components, exhibiting a persistence well above the threshold of 0.4. This feature corresponds to the junction region of the attractor. In contrast, the low-density subset (Fig. 11b) is characterized by three persistent 1-holes exceeding the same threshold. These features are mapped back onto the phase-space

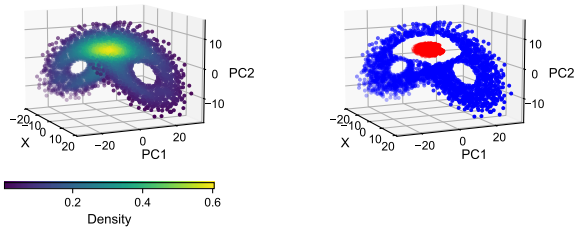


Fig. 10: For the L63-R dataset, (left panel) point-cloud density estimated with KDE and (right panel) density-based partition of phase-space into the high-density (top 10%, red) and low-density (bottom 60%, blue) subsets.

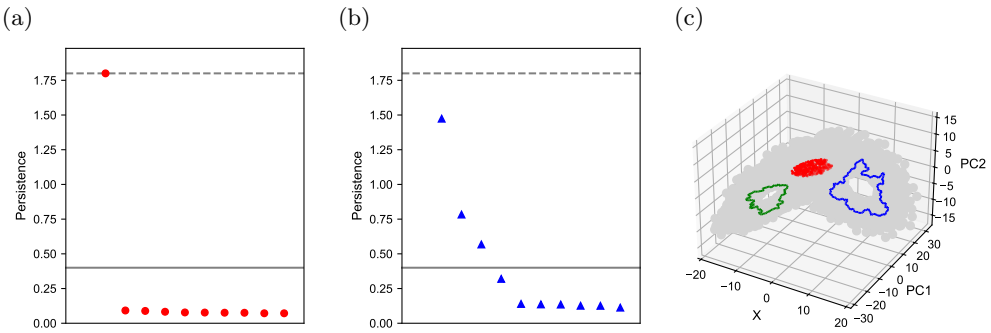


Fig. 11: For the L63-R dataset, ten most persistent topological (a) 0-holes (red dots) and (b) 1-holes (blue triangles) of the high-density (top 10%) and low-density (bottom 60%) ranges. The solid line denotes the 0.4 persistence threshold and the dashed line the infinity persistence. (c) Phase-space mapping of the topological features on the L63-R point cloud. Red points correspond to the most persistent 0-hole in the high-density range, isolating the junction region, while the blue and green curves represent the second and third most persistent 1-holes in the low-density range, capturing the loops around the two wings.

coordinates in Figure 11(c). The second and third most persistent cycles correspond to the two loops of the Lorenz attractor (green and blue representative cycles), whereas the most persistent cycle encloses the central void induced by the removal of the high-density junction.

Figure 12(a) reports the junction diagnostic: the local density grows as the trajectory approaches the junction regime and peaks around onset, before relaxing as the system leaves it. In this case the junction locus is known from the BraMAH complex and lies within the 0-TWR: the trajectory-based diagnostic yields a set of red points in the point cloud which play the role of the 1-cells, marked as red line segments in Figure 4(a).

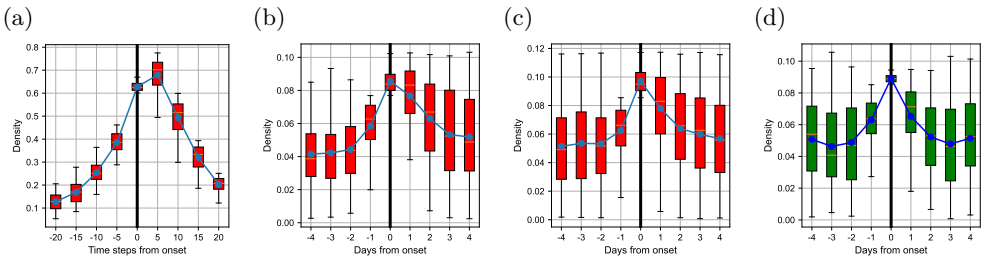


Fig. 12: Trajectory-based diagnostic for the 0-TWRs of the three datasets. (a) L63-R: average evolution of local point-cloud density (blue dotted line) as a function of time relative to the onset of the 0-TWR. (b) 0-TWR1 in the JET-LAT-SA: the signature of the junction locus can be read from the density rising to a sharp local maximum around onset. (c) 0-TWR1 and (d) 0-TWR2 in the JET-LAT-NA: both cases present the signature of a junction locus. Boxes show the interquartile range and whiskers are defined using Tukey's rule.

The topological results for L63-R return the three structural elements of its known branched manifold: one junction region, as a 0-TWR, and two wing voids, each surrounded by a 1-TWR. The two wing cycles can also be recovered with standard PH applied to the full cloud. The junction region, instead, only emerges when the densest subset is analyzed on its own, since the full cloud forms a single connected component whose zeroth-order homology is trivial. These are the same three regimes returned by the statistical partition of Section 4.1: *JU* falls on the junction region isolated by the high-density 0-hole, while *LW* and *RW* correspond to the wings enclosed by the two low-density cycles.

4.2.2 South-Atlantic case

Figure 13 illustrates multiple threshold configurations for the density-based partition of the phase space. Specifically, each panel isolates the corresponding lowest and highest percentiles of the density distribution; for example, panel (a) contrasts the 5% least dense subset (blue point cloud) against the 5% most dense subset (red point cloud). Subsequent panels increment this symmetric partition to 10%, 15%, and 20%. The tracking of 1-holes (Figure 14c) and the comparison with the statistical regimes (Figure 20) extend these ranges in steps of 5%, up to the 35% least dense and the 50% most dense subsets.

A central region of maximum density surrounded by two low-density side lobes is consistently observed across the evaluated high- and low-density ranges. These structures stand out clearly in Figure 14. The high-density range exhibits up to two significantly persistent 0-holes, as shown in panel (a). In fact, the two high-density persistent 0-holes in the case of 20% most dense subset correspond to groups of points in the phase space that are very close to each other: one being considerably larger in number of points than the second one. The second 0-hole exceeds the threshold only for the 20% subset: it is not recovered across the evaluated high-density ranges and

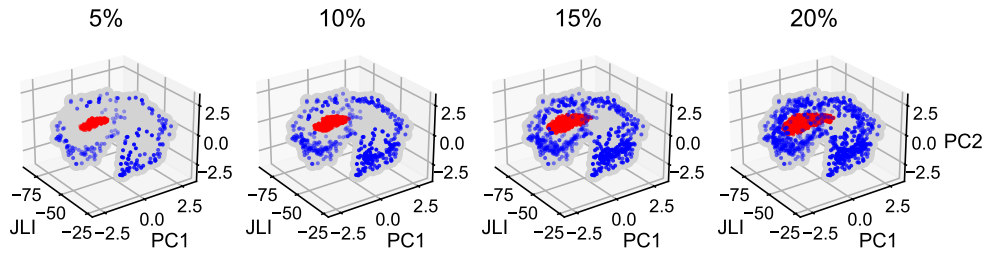


Fig. 13: For the JET-LAT-SA dataset, density-based partition into the high-density (red) and low-density (blue) subsets varying the high- and low-density ranges based on different highest and lowest percentiles of the density distribution.

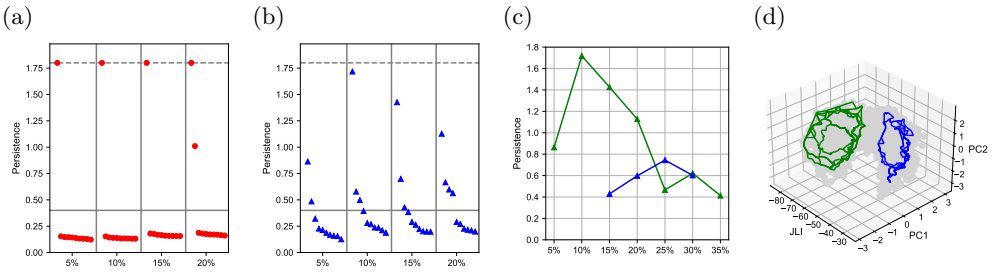


Fig. 14: For the JET-LAT-SA dataset, ten most persistent (a) 0-holes and (b) 1-holes corresponding to the high- and low-density ranges in Figure 13, respectively. The solid line denotes the 0.4 persistence threshold and the dashed line the infinity persistence. (c) Persistence of the two main sets of matched 1-holes across low density ranges and (d) phase-space mapping of these sets on the original point cloud (gray).

is discarded. The largest component, recovered across the four ranges, is hereafter 0-TWR1; its member points are those of the 10% component. For the low-density level of the analysis, there is a single H_1 feature with maximal persistence for the four threshold values, as seen in panel (b), and up to four 1-holes that significantly exceed the persistence threshold.

Figure 12(b) reports the junction diagnostic applied to 0-TWR1. As in the Lorenz case, the local density peaks around onset: 0-TWR1 displays the density signature of a junction, i.e. of a confluence of branches in a branched-manifold structure.

The tracking of representative 1-cycles across low-density ranges for the JET-LAT-SA point cloud is detailed in Figure 14(c) and (d). Two groups of cycles are present in multiple consecutive ranges (c), indicating robust topological structures; they are hereafter 1-TWR1 and 1-TWR2. Representative cycles for both groups in the phase space (d) confirm that these features capture the low-density side lobes of the point cloud, as previously shown in Figure 13.

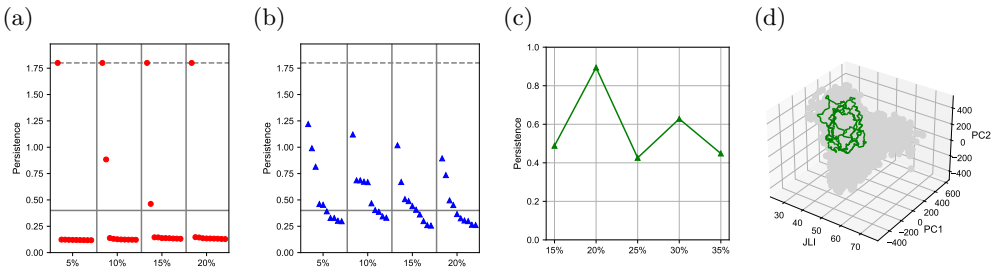


Fig. 15: For the JET-LAT-NA dataset, ten most persistent (a) 0-holes and (b) 1-holes corresponding to the high- and low-density ranges of the JET-LAT-NA point cloud, respectively. The solid line denotes the 0.4 persistence threshold and the dashed line the infinity persistence. (c) Persistence of the main set of matched 1-holes across low density ranges and (d) phase-space mapping of this set on the original point cloud (gray).

4.2.3 North-Atlantic case

Figure 15 presents persistence diagrams for the high- and low-density ranges for the JET-LAT-NA phase space. In the 10% and 15% high-density ranges, two 0-holes exceed the persistence threshold (a); they are hereafter 0-TWR1 and 0-TWR2, with member points taken from the 10% components, and correspond to the non-trivial topological features reported by Strommen et al. (2023). Mirroring the South Atlantic case, the low-density range adds significant information through the persistent 1-holes (b). Matching these 1-holes across different density thresholds for the low-density range reveals a robust set, hereafter 1-TWR1; see Figure 15(c). The JET-LAT-NA phase-space geometry shows that 1-TWR1 corresponds to a low-density lobe associated with lower North-Atlantic JLI values (d).

Figure 12 reports the junction diagnostic applied to 0-TWR1 (c) and 0-TWR2 (d). In both cases the local density grows as the trajectory approaches the regime, peaks sharply at onset, and returns to its background level within the ±4-day window of the composite. Both dense components of the JET-LAT-NA phase space thus display the density signature of a junction: as for the Lorenz attractor and the South Atlantic jet, the persistent high-density components are read as confluences of branches.

Further examining the relationship between the North Atlantic JLI and these topological JET-LAT-NA features, Figure 16(a) illustrates how its conditional distribution differs from the climatology (black line) when restricted to the member points of the two most persistent 0-holes (0-TWR1, 0-TWR2) and of the robust 1-hole (1-TWR1). The conditional distributions for 0-TWR1 and 0-TWR2 isolate the central and polar peaks, respectively, while 1-TWR1 isolates the equatorial peak. This clear separation becomes more evident when the JET-LAT-NA points that are part of these topological features are plotted separately, as shown in Figure 16(b) with different colors. These results are consistent with the circulation ERA20C composites for the topological JET-LAT-NA features shown in Figure 17. We can now reconstruct the three classic North Atlantic PFJ regimes documented by Woollings et al. (2010): while the

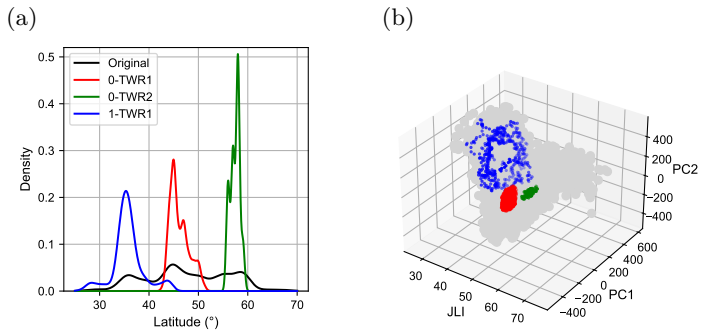


Fig. 16: (a) Climatological North-Atlantic JLI (ERA20C) winter distribution (black line) and conditional distributions defined by the JET-LAT-NA TWRs in the period 1901-2010. (b) Phase-space points associated with these TWRs above the original point cloud (gray).

central and northern jets map to the dense connected components (0-TWR1 and 0-TWR2), the southern jet is captured by the low-density cycle 1-TWR1, see Figure 17(c). This relationship is observed even when taking a database of lower spatial resolution such as NCEP-NCAR reanalysis and in a shorter time frame from the beginning of the satellite era (not shown).

Strommen et al. (2023) report the two dense components but not the southern jet, whose absence they trace to the southern mode appearing as a shoulder rather than a peak of the density distribution. The result here keeps what they detect and adds the southern regime, recovered from the low-density range. With the centrality-based bifiltration of Anbouhi (2026), the southern jet appears at the high-centrality end as connected components of ten to twenty points, when the centrality is computed over 30 to 100 nearest neighbors. Here it is a 1-TWR: a first-order feature of the low-density range, tracked across thresholds, whose member points support the conditional distribution and the composite of Figures 16(a) and 17. The northern jet is where the two basins differ, appearing as a high-density component in the North Atlantic but as a low-density feature in the South Atlantic. In both regions, the full set of regimes can be detected and classified as distinct.

4.3 South-Atlantic regimes through the statistical and the topological lenses

It is pertinent to assess whether the topological features identified through PH correspond to the circulation patterns identified by the statistical WRs. Figure 18 compares the normalized vertical-average zonal wind anomalies for the statistical PFJ regimes against the composites associated with the TWRs for the JET-LAT-SA dataset. The composites conditioned on these topological features resemble the statistical regimes: J_C exhibits a high correlation with 0-TWR1, J_S with 1-TWR1, and J_N with 1-TWR2.

Given the dominance of the positional variability of the PFJ in distinguishing the statistical regimes as shown previously in Figure 6 (a), Figure 19(a) shows how the JLI

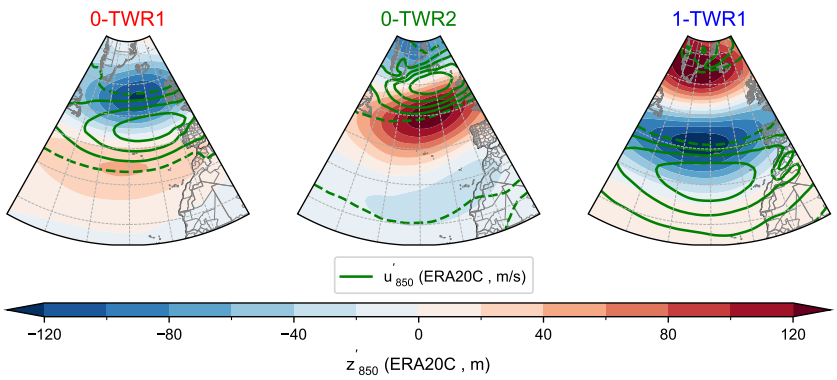


Fig. 17: z'_{850} (shading, m) and u'_{850} (green contours) ERA20C composites (DJF 1901-2010) for days associated with JET-LAT-NA TWRs: 0-TWR1 (left), 0-TWR2 (middle) and 1-TWR1 (right).

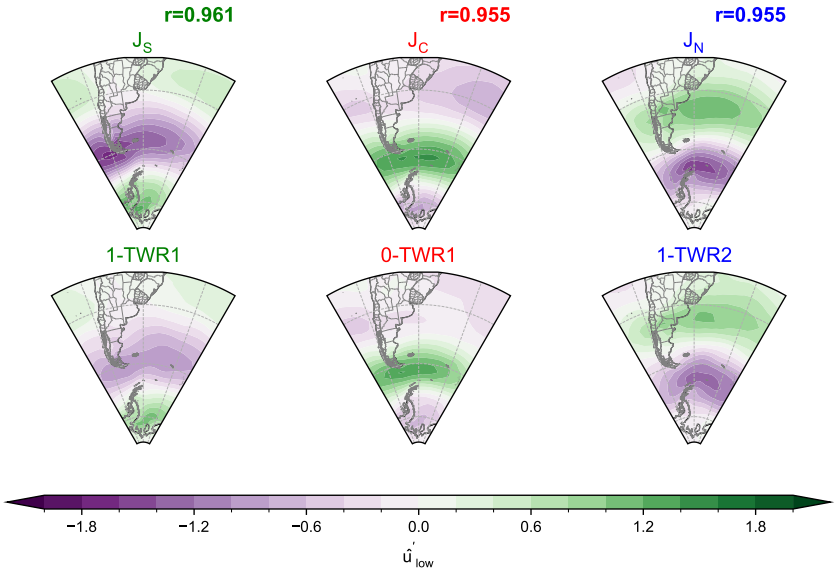


Fig. 18: Top row: $\hat{u}'_{low}$ ERA5 composites (JJA 1981-2020) of assigned days for PFJ regimes: J_S (left), J_C (middle), and J_N (right). Bottom row: corresponding composites for JET-LAT-SA TWRs: 1-TWR1 (left), 0-TWR1 (middle) and 1-TWR2 (right). Linear latitude-weighted correlations by column are displayed at the top for each case.

distribution is modified now considering days within each TWR. Again, a distinct peak from the original trimodal climatological distribution becomes much more pronounced

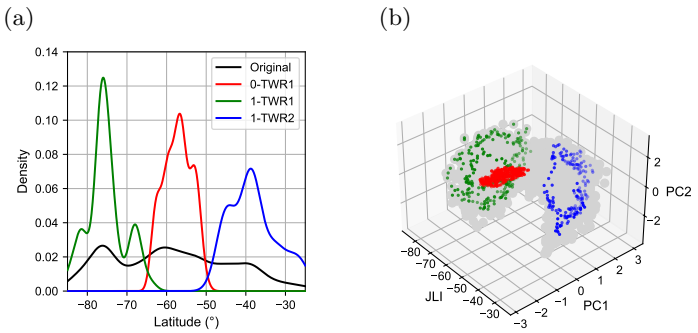


Fig. 19: (a) Climatological South-Atlantic JLI (ERA5) winter distribution (black line) and conditional distributions defined by the JET-LAT-SA TWRs in the period 1981-2020. (b) Phase space points associated with these TWRs above the original point cloud (gray).

in each case. A global view in phase space of the points that reside in each topological feature, shown in Figure 19(b), demonstrates how the JLI as coordinate allows for clear differentiation between them.

Figure 20(a, top) shows the linear-latitude-weighted correlation between the composites of the statistical JET-LAT-SA WRs (top row in Figure 18) and the corresponding composite for 0-TWR1 when it is identified in different high-density ranges. The relationship with J_C is robust when considering different high-density subsets. Another way to view this is the contribution of statistical WR states to this TWR, as shown in Figure 20(a, bottom), where the equivalence between J_C and 0-TWR1 is clearer in the most restrictive cases.

The same analysis can be performed based on what is shown in Figure 20(b) and (c), which are equivalent to Figure 20(a), but instead consider different low-density ranges to identify 1-TWR1 and 1-TWR2, respectively. Based on the correlations (top) and state contributions (bottom), 1-TWR1 is predominantly related with J_S states, and 1-TWR2 is dominated by J_N states. In the first case, the equivalence is more evident in the most restrictive low-density ranges, but for the last one the opposite behavior is observed. Topology thus confirms the statistical description: each statistical regime translates into a high- or low-density topological feature. J_C is related to the most persistent high-density component, and J_S and J_N to the two robust low-density features. Two of the three regimes are thus first-order features of the low-density range: an analysis confined to the high-density range retains the central jet alone. The correspondence holds in the circulation composites, in the conditional jet-latitude distributions, and in the point-by-point assignment across density thresholds. The regimes are therefore recovered as distinct structural elements of the dynamics.

5 Conclusions

The introduction asked whether PH can capture all the dynamically relevant structures underlying the large-scale atmospheric circulation. The answer given here has

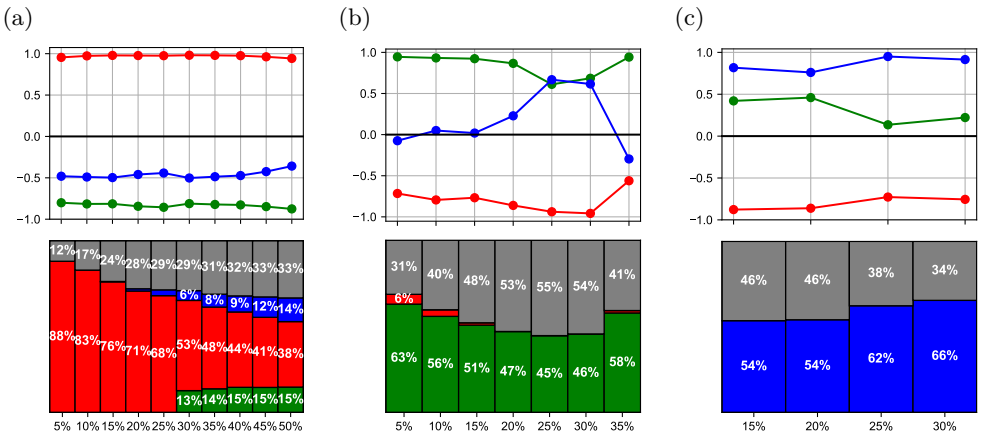


Fig. 20: For the JET-LAT-SA dataset, (top) linear latitude-weighted correlation and (bottom) percentage contribution of regime-assigned points by statistical WRs to (a) 0-TWR1 across high-density ranges, (b) 1-TWR1 and (c) 1-TWR2 across low-density ranges. Gray bars denote unassigned points.

two parts. On its own, PH cannot: junctions, the sites where the branches of a branched manifold merge, are not homological features of the attractor as a whole, since the point cloud forms a single connected piece whose zeroth-order homology is trivial. A density sweep may surface them among many other features, but nothing in the sweep says what they are. The branched manifold supplies what is missing: it prescribes, in advance, that connected components are to be sought at the dense end of the distribution, where converging branches overlay, and voids at the sparse end, which trajectories encircle without entering. With that prescription, PH recovers every regime documented in the three datasets examined, and the answer becomes affirmative, once a weather regime is required to be an identifiable part of the branched manifold rather than any non-trivial feature. This requirement closes the gap between what regime means in the atmospheric sciences and what feature means in TDA. Both elemental parts of a branched manifold realize in phase space the recurrence that the meteorological notion demands, each in its own manner: a high-density component is revisited, and so is the region that surrounds a void, while the void itself is never entered: there, recurrence takes the form of repeated encirclement of states the flow does not visit. The definition assigns each form of recurrence its homological order and its end of the density range, with no number of regimes fixed in advance. The trajectory-based diagnostic then looks for the density signature of a junction locus, encountered in all three datasets.

The Lorenz attractor, whose branched manifold is known beforehand, provides an illustrative example. Treated as an atmospheric dataset, it returns the same three regimes by the statistical and the topological routes, run independently: the cluster around the junction (JU) coincides with the persistent high-density component, and the two wing clusters (LW, RW) with the regions surrounding the voids that the

low-density cycles enclose. The junction regime is, moreover, the one with the longest event durations, on the attractor as in the South Atlantic jet.

For the South Atlantic PFJ, this work delivers both the statistical baseline and its topological counterpart. The wintertime regimes of this jet are characterized here for the first time within the weather-regime framework: three regimes separated by the latitudinal position of the jet, consistent with the trimodal winter distribution of the JLI and robust across three reanalyses. The central jet J_C, the most frequent, is favored during the positive AAO phase and the northern jet J_N during the negative one, while the southern jet J_S, the least frequent but the most intense, is largely independent of the annular mode and accompanies blocking-like anticyclonic circulation at high latitudes. The topological analysis retrieves this portrait term by term: J_C matches the persistent high-density component 0-TWR1, which the trajectory-based diagnostic reads as hosting a junction, while J_S and J_N match the robust low-density cycles 1-TWR1 and 1-TWR2, with the agreement holding from the circulation composites down to the assignment of individual days across density thresholds. Two of the three regimes of this basin are therefore low-density features, which an analysis confined to the dense end would not retain.

For the North Atlantic PFJ, the two persistent high-density components coincide with the non-trivial features of Strommen et al. (2023) and correspond to the central and northern jets of Woollings et al. (2010); both dense components display the density signature of a junction locus. The southern jet, which the density-based bifiltration left undetected because its mode appears as a shoulder of the density distribution rather than a peak, is recovered as a robust low-density cycle. Moving from non-trivial persistent homologies in a bifiltration to the search of junctions and voids structuring a branched manifold thus seems to enlarge, rather than restrict, what is detected. The comparison between basins yields a distinction that the statistical description alone does not draw: the northern jet is a high-density feature in the North Atlantic and a low-density feature in the South Atlantic. Regimes bearing the same synoptic name may play different roles on the branched manifold: the classification into 0-TWRs and 1-TWRs carries dynamical information of its own.

The trajectory-based density diagnostic is not tied to the atmospheric setting, and can be applied to detect potential junction loci as subsets of phase-space points from embedded datasets, in general. To our knowledge, this is the first report extracting branched-manifold junctions from point clouds with no prior cell complex construction whatsoever. It also brings the time ordering of the data into the topological protocol. The definitions we put forward operationally remain within PH and inherit the dependence on scale and on threshold settings; what we add is the prescription of where each class of feature is to be sought. The TWRs are interpreted as specific parts of the branched manifold: their union does not recompose the full structure. Describing how trajectories leave one regime and reach another is left for future work, involving directed features that can be obtained with a directed algebraic topology approach (Charó et al. 2022; Sciamarella and Charó 2024; Sciamarella 2026).

Acknowledgments

The authors wish to thank Michael Ghil for careful reading and insightful discussions.

Statements and Declarations

Funding

M. E. Micieli acknowledges funding from CONICET. The authors also acknowledge funding from the ANR project TeMPlex (ANR-23-CE56-0002), PI: Denisse Sciamarella.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by Marco E. Micieli and Marisol Osman. Data formal analysis was performed by Marco E. Micieli. The first draft of the manuscript was written by Marco E. Micieli and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

ERA5 reanalysis data are freely available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). Python codes used for the topological analysis are available at https://git.cima.fcen.uba.ar/marco.micieli/BraMAPH_approach.git.

Appendix A Homologies

The k-th homology group is defined as the quotient group $H_k = Z_k/B_k$, where Z_k represents the group of k-cycles (closed chains of k-dimensional cells with a null boundary) and B_k denotes the group of k-boundaries (chains that constitute the boundary of a $k+1$-dimensional chain). Under this formulation, two k-cycles z_1 and z_2 belong to the same homology class (i.e., they represent the same hole) if their difference is a boundary, $z_1 - z_2 = \partial c$, where c is a $(k+1)$-chain.

Geometrically, this implies that one cycle can be continuously deformed into the other through a sequence of higher-dimensional cells within the complex. If two closed trajectories in phase space encircle the same void, they are homologically equivalent provided the region between them has been visited by the system. In this case, the algebraic difference between the trajectories is the boundary of that intermediate sampled region.

The rank of this quotient group, known as the k-th Betti number β_k, quantifies the number of independent k-dimensional holes. To provide an intuitive interpretation for phase space analysis, β_0 counts the number of connected components, indicating

whether the dynamically visited states form a single continuous domain or isolated patches. The β_1 invariant counts independent holes, which represent the existence of states surrounding an unvisited region; on a torus, $\beta_1 = 2$ corresponds to the two fundamental cycles (poloidal and toroidal). Finally, β_2 counts enclosed volumetric cavities, such as the single empty interior ($\beta_2 = 1$) bounded by the surface of the torus.

In PH, the extraction of these invariants from finite point clouds is done using simplicial cells, i.e., cells having the minimal number of vertices per dimensional layer (vertices, edges, triangles, and so on). A k-chain within a complex is defined as a formal linear combination of k-simplices, each multiplied by a scalar coefficient. Standard software packages, such as Ripser (Bauer 2021) and GUDHI (Maria et al. 2014), compute PH with coefficients that can only take values of 0 or 1. A representative cycle, defined as a specific chain $z \in Z_k$ belonging to the detected homology class, provides a concrete geometric instantiation of the abstract invariant mapped onto the original data points.

Appendix B Regular vs. branched manifolds

While PH provides a general recipe to extract persistent homologies from datasets, none of the cell complexes of a scale-dependent filtration provides a faithful stand-alone topological representation of the manifold underlying the point cloud. A special recipe was conceived with that purpose (Sciamarella and Mindlin 1999, 2001). A BraMAH complex (Charó et al. 2020, 2021; Sciamarella and Charó 2024; Sciamarella 2026) is a construction that integrates multiple scales in a single scaffold, and that can be used to compute homologies with integer coefficients, thus providing a complete topological characterization by simultaneously capturing the generators of the homology groups as well as intrinsic orientability properties.

A BraMAH complex can be built for both regular or branched manifolds. Table B1 shows examples of regular and branched manifolds for a few paradigmatic spaces. The existence of a junction locus is the primary feature distinguishing chaotic attractors from regular manifolds. For instance, while a regular torus and Deng's toroidal chaos (Mosto et al. 2024) share the same Betti numbers, only the latter exhibits a junction locus. Because the junction locus prevents the assignment of a global orientation across the entire cell complex, the orientability for branched-manifold structures must be computed independently for each branch passing through the junction. In the Rössler attractor, for instance, the coexistence of an orientation-preserving and an orientation-reversing branch within the same support constitutes its fundamental topological signature.

A junction locus is a set of lower-dimensional cells along which the branches are glued. Each of these cells has codimension one and is shared by more than two top-cells, that is, by more than two cells of the highest dimension present in the complex. A branched manifold may have several disjoint junction loci, each involving a different set of merging branches. In the BraMAH complex of the Lorenz attractor there is only one, and it is connected: it is given by the chain of 1-cells $\langle 0, 1\rangle + \langle 1, 2\rangle$ shown in Figure 4(a), the two 1-cells sharing a common 0-cell. A BraMAH complex represents the manifold itself, and the junction locus is read off from the cells.

Table B1: BraMAH properties for regular and branched manifolds.

Case	β_0	β_1	β_2	Junction Locus	Orientability
Toroidal surface	1	2	1	No	Preserving
Cylindrical surface	1	1	0	No	Preserving
Möbius band	1	1	0	No	Reversing
Rössler attractor	1	1	0	Yes	Preserving & Reversing branches
Lorenz attractor	1	2	0	Yes	All Preserving branches
Deng's toroidal chaos	1	2	1	Yes	All Preserving branches

Appendix C Representative-cycle tracking

Because the representative 1-cycles of observational atmospheric datasets are sensitive to the selected partition thresholds, the following three-stage procedure is implemented:

1. **Minimal cycle extraction:** when two significant cycles share at least one edge, they are taken to represent the same hole, and the one of shortest geometric length is retained, yielding a minimal loop per feature.
2. **Centroid tracking:** the distance between the spatial centroids of the retained representative 1-cycles is computed iteratively between consecutive density thresholds. Pairs of cycles in adjacent partition steps are linked as provisional matches through an optimal one-to-one assignment that minimizes the inter-centroid distances.
3. **Geometric consistency constraints:** to validate a provisional link and ensure that the tracking follows the same feature, two structural criteria are enforced:

$$d(C_i, C_j) < \min(R(z_i), R(z_j)) \tag{C1}$$

$$|\Delta L| = |L_i - L_j| < \mathcal{P}_\alpha, \tag{C2}$$

where $d(C_i, C_j)$ is the distance between centroids, $R(z)$ is the spatial radius of the cycle z (the maximum distance between its centroid and any of its constituent 0-cells), and the difference in total geometric length $|\Delta L|$ must fall below a predefined distance percentile $\mathcal{P}_\alpha$ of the pairwise-distance distribution of the full point cloud; $\alpha = 95$ is used throughout this work.

References

Anbouhi S (2026) Improving topological detection of weather regimes in climate dynamical systems. Climate Dynamics 64:356. https://doi.org/10.1007/s00382-026-08296-9

Bauer U (2021) Ripser: efficient computation of Vietoris-Rips persistence barcodes. Journal of Applied and Computational Topology 5(3):391–423. https://doi.org/10.1007/s41468-021-00071-5

Birman JS, Williams RF (1983) Knotted periodic orbits in dynamical systems I: Lorenz's equations. Topology 22(1):47–82

Blackmon ML, Madden RA, Wallace JM, et al (1979) Geographical variations in the vertical structure of geopotential height fluctuations. Journal of the Atmospheric Sciences 36:2450–2466

Carlsson G (2009) Topology and data. Bulletin of The American Mathematical Society 46:255–308. https://doi.org/10.1090/S0273-0979-09-01249-X

Carlsson G, Zomorodian A (2009) The theory of multidimensional persistence. Discrete & Computational Geometry 42(1):71–93

Charó GD, Artana G, Sciamarella D (2020) Topology of dynamical reconstructions from Lagrangian data. Physica D: Nonlinear Phenomena 405:132371. https://doi.org/10.1016/j.physd.2020.132371

Charó GD, Artana G, Sciamarella D (2021) Topological colouring of fluid particles unravels finite-time coherent sets. Journal of Fluid Mechanics 923:A17. https://doi.org/10.1017/jfm.2021.561

Charó GD, Letellier C, Sciamarella D (2022) Templex: A bridge between homologies and templates for chaotic attractors. Chaos: An Interdisciplinary Journal of Nonlinear Science 32(8):083108. https://doi.org/10.1063/5.0092933

Charó GD, Sciamarella D, Ruiz J, et al (2025) Topological modes of variability of the wind-driven ocean circulation. Chaos: An Interdisciplinary Journal of Nonlinear Science 35(9):093121. https://doi.org/10.1063/5.0261968

Christiansen B (2007) Atmospheric circulation regimes: Can cluster analysis provide the number? Journal of Climate 20:2229–2250. https://doi.org/10.1175/JCLI4107.1

Dorrington J, Strommen K, Fabiano F (2022) Quantifying climate model representation of the wintertime Euro-Atlantic circulation using geopotential-jet regimes. Weather and Climate Dynamics 3(2):505–533. https://doi.org/10.5194/wcd-3-505-2022

Duchon CE (1979) Lanczos filtering in one and two dimensions. Journal of Applied Meteorology and Climatology 18(8):1016–1022

Edelsbrunner H, Harer JL (2010) Computational Topology: An Introduction. American Mathematical Society, Providence, RI

Faranda D, Messori G, Yiou P (2017) Dynamical proxies of North Atlantic predictability and extremes. Scientific Reports 7:41278. https://doi.org/10.1038/srep41278

Gallego D, Ribera P, García-Herrera R, et al (2005) A new look for the Southern Hemisphere jet stream. Climate Dynamics 24:607–621. https://doi.org/10.1007/

s00382-005-0006-7

Ghil M, Lucarini V (2020) The physics of climate variability and climate change. Reviews of Modern Physics 92(3):035002. https://doi.org/10.1103/RevModPhys.92.035002

Ghil M, Robertson AW (2002) "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting? Proceedings of the National Academy of Sciences 99(suppl 1):2493–2500. https://doi.org/10.1073/pnas.012580899

Ghil M, Sciamarella D (2023) Dynamical systems, algebraic topology and the climate sciences. Nonlinear Processes in Geophysics 30(4):399–434. https://doi.org/10.5194/npg-30-399-2023

Ghil M, Chavez E, Groth A, et al (2025) Extratropical sub-seasonal-to-seasonal oscillations and multiple regimes: The dynamical systems view. In: Robertson AW, Vitart F (eds) The Gap Between Weather and Climate Forecasting: Sub-Seasonal to Seasonal Prediction, 2nd edn. Elsevier, Amsterdam, p 137–180, https://doi.org/10.5281/zenodo.17116632

Gilmore R (1998) Topological analysis of chaotic dynamical systems. Reviews of Modern Physics 70(4):1455. https://doi.org/10.1103/RevModPhys.70.1455

Gilmore R, Lefranc M (2012) The topology of chaos: Alice in stretch and squeezeland. John Wiley & Sons

Grams CM (2026) A life cycle definition of year-round weather regimes in the North Atlantic European region. Weather and Climate Dynamics 7(3):1641–1680. https://doi.org/10.5194/wcd-7-1641-2026, URL https://wcd.copernicus.org/articles/7/1641/2026/

Held IM, Hou AY (1980) Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. Journal of the Atmospheric Sciences 37:515–533

Hersbach H, Bell B, Berrisford P, et al (2020) The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146(730):1999–2049

Hochman A, Messori G, Quinting JF, et al (2021) Do Atlantic-European weather regimes physically exist? Geophysical Research Letters 48(20):e2021GL095574. https://doi.org/10.1029/2021GL095574

Hoskins BJ, James IN, White GH (1983) The shape, propagation and mean-flow interaction of large-scale weather systems. Journal of the Atmospheric Sciences 40:1595–1612. https://doi.org/10.1175/1520-0469(1983)040⟨1595:TSPAMF⟩2.0.CO;2

Kalnay E, Kanamitsu M, Kistler R, et al (1996) The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77(3):437–472

Kinsey LC (2012) Topology of surfaces. Springer Science & Business Media

Kobayashi S, Ota Y, Harada Y, et al (2015) The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan Ser II 93(1):5–48. https://doi.org/10.2151/jmsj.2015-001

Lee SH, Messori G (2024) The dynamical footprint of year-round North American weather regimes. Geophysical Research Letters 51:e2023GL107161. https://doi.org/10.1029/2023GL107161

Lee SH, Tippett MK, Polvani LM (2023) A new year-round weather regime classification for North America. Journal of Climate 36(20):7091–7108. https://doi.org/10.1175/JCLI-D-23-0214.1

Legras B, Ghil M (1985) Persistent anomalies, blocking and variations in atmospheric predictability. Journal of the Atmospheric Sciences 42(5):433–471. https://doi.org/10.1175/1520-0469(1985)042⟨0433:PABAVI⟩2.0.CO;2

Lloyd SP (1982) Least squares quantization in PCM. IEEE Transactions on Information Theory 28:129–136. https://doi.org/10.1109/TIT.1982.1056489

Lorenz EN (1963) Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20(2):130–141. https://doi.org/10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2

Madonna E, Li C, Grams CM, et al (2017) The link between eddy-driven jet variability and weather regimes in the North Atlantic-European sector. Quarterly Journal of the Royal Meteorological Society 143(708):2960–2972. https://doi.org/10.1002/qj.3155

Maria C, Boissonnat JD, Glisse M, et al (2014) The GUDHI library: Simplicial complexes and persistent homology. In: Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4, Springer, pp 167–174, https://doi.org/10.1007/978-3-662-44199-2_28

Mo K, Ghil M (1988) Cluster analysis of multiple planetary flow regimes. Journal of Geophysical Research: Atmospheres 93(D9):10927–10952. https://doi.org/10.1029/JD093iD09p10927

Mosto C, Charó GD, Letellier C, et al (2024) Templex-based dynamical units for a taxonomy of chaos. Chaos: An Interdisciplinary Journal of Nonlinear Science 34(11):113111. https://doi.org/10.1063/5.0233160

Namias J (1950) The index cycle and its role in the general circulation. Journal of Meteorology 7(2):130–139. https://doi.org/10.1175/1520-0469(1950)007⟨0130:TICAIR⟩2.0.CO;2

Parker T, Woollings T, Weisheimer A, et al (2019) Seasonal predictability of the winter North Atlantic oscillation from a jet stream perspective. Geophysical Research Letters 46(16):10159–10167. https://doi.org/10.1029/2019GL084402

Poli P, Hersbach H, Dee D, et al (2016) ERA-20C: An atmospheric reanalysis of the 20th century. Journal of Climate 29(11):4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Reinhold BB, Pierrehumbert RT (1982) Dynamics of weather regimes: Quasi-stationary waves and blocking. Monthly Weather Review 110(9):1105–1145. https://doi.org/10.1175/1520-0493(1982)110⟨1105:DOWRQS⟩2.0.CO;2

Robertson AW, Ghil M (1999) Large-scale weather regimes and local climate over the Western United States. Journal of Climate 12(6):1796–1813. https://doi.org/10.1175/1520-0442(1999)012⟨1796:LSWRAL⟩2.0.CO;2

Sciamarella D (2026) Functorial invariants for chaos topology from data. Physical Review E 114(2):024205. https://doi.org/10.1103/zx76-5115

Sciamarella D, Charó GD (2024) New elements for a theory of chaos topology. In: Amster P, Benevieri P (eds) Topological Methods for Delay and Ordinary Differential Equations. Advances in Mechanics and Mathematics, Springer Birkhäuser, Cham, https://doi.org/10.1007/978-3-031-61337-1_9

Sciamarella D, Mindlin G (1999) Topological structure of chaotic flows from human speech data. Physical Review Letters 82(7):1450. https://doi.org/10.1103/PhysRevLett.82.1450

Sciamarella D, Mindlin G (2001) Unveiling the topological structure of chaotic flows from data. Physical Review E 64(3):036209. https://doi.org/10.1103/PhysRevE.64.036209

Silverman BW (2018) Density estimation for statistics and data analysis. Routledge

Solman SA, Menendez C (2003) Weather regimes in the South American sector and neighbouring oceans during winter. Climate Dynamics 21(1):91–104. https://doi.org/10.1007/s00382-003-0320-x

Stolz BJ, Tanner J, Harrington HA, et al (2020) Geometric anomaly detection in data. Proceedings of the National Academy of Sciences 117(33):19664–19669. https://doi.org/10.1073/pnas.2001741117

Strommen K, Chantry M, Dorrington J, et al (2023) A topological perspective on weather regimes. Climate Dynamics 60(5–6):1415–1445. https://doi.org/10.1007/s00382-022-06395-x

Vautard R (1990) Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. Monthly Weather Review 118(10):2056–2081. https://doi.org/10.1175/1520-0493(1990)118⟨2056:MWROTN⟩2.0.CO;2

Weeks ER, Tian Y, Urbach JS, et al (1997) Transitions between blocked and zonal flows in a rotating annulus with topography. Science 278(5343):1598–1601. https://doi.org/10.1126/science.278.5343.1598

Williams RF (1979) The structure of Lorenz attractors. Publications Mathématiques de l'IHÉS 50:73–99. https://doi.org/10.1007/BF02684770

Woollings T, Hannachi A, Hoskins B (2010) Variability of the North Atlantic eddy-driven jet stream. Quarterly Journal of the Royal Meteorological Society 136(649):856–868. https://doi.org/10.1002/qj.625

Yarnal B, Draves J (1993) A synoptic climatology of stream flow and acidity. Climate Research 2:193–202