8th International Workshop on Climate Informatics
September 19-21, 2018
CI 2018 Hosted by the National Center for Atmospheric Research in Boulder, CO

TOWARDS TOPOLOGICAL PATTERN DETECTION IN FLUID AND CLIMATE SIMULATION DATA

Grzegorz Muszynski[1,2], Karthik Kashinath[2], Vitaliy Kurlin[1], Michael Wehner[2], Prabhat[2]

***Abstract*—Increasingly massive amounts of high-resolution climate datasets are being generated by observations as well as complex climate models. As the unprecedented growth of data continues, a massive challenge is to design automated and efficient data analysis techniques that can extract meaningful insights from vast datasets. In particular, a key challenge is the detection and characterization of weather and climate patterns. Machine learning, including deep learning, are currently popularly used for these tasks. These techniques, however, do not incorporate geometric features of data and temporal persistence information. In this paper, we develop a novel approach to pattern detection and characterization based on dynamical systems, manifold learning and topological data analysis (i.e., persistent homology) that utilize important geometric and topological properties of underlying patterns in datasets.**

I. MOTIVATION

A large volume of data is currently produced by high resolution global climate model simulations, observational campaigns and data assimilation (i.e., climate reanalysis products combining a model with a range of observational data). Both the volume and the complex nature of climate data pose many challenges to design an automated and efficient data analysis techniques. Here we focus on one particular aspect of climate data analysis, the detection of weather patterns. Typically, pattern detection in the climate community has been done using standard image processing techniques such as image thresholding to detect weather patterns in climate data [1]. However, an open challenge in the the community, across many types of weather patterns, is converging on reliable and consistent threshold values for different variables [2]. Machine learning and deep learning have recently started to gain recognition and popularity due to their revolutionary success in commercial applications and fields such as computer vision, robotics and control systems [3], [4].

We propose an alternative approach to weather pattern detection that lies at the intersection of dynamical systems, manifold learning and topological data analysis [5]. Most existing machine or deep learning methods applied to climate research do not incorporate geometric and topological features of the underlying patterns, nor do they include any temporal persistence information, both of which are strongly linked to the physics and evolution of the dynamical system under consideration (*i.e.*, the global climate system).

We develop and test this approach using canonical fluid flow simulations with persistent repeating spatio-temporal patterns, such as vortex streets. We hypothesize that these patterns can be detected, identified and characterized by the combination of the time-delay coordinate embedding [6], manifold learning (i.e., diffusion maps) [7], and recent advances in applied topology (i.e., persistent homology) [8]. This fluid dynamical system will serve as a test bench for the development of a detection approach for other similar coherent spatio-temporal patterns, such as atmospheric blocking and Rossby waves in climate data. Atmospheric blocking is a persistent weather pattern (high-pressure system) that significantly influences the climatological westerly flow at mid-latitudes, especially the North Atlantic, Europe, and the North Pacific [9]. The impacts of blocking can persist for several weeks over certain geographical region and can have dramatic adverse consequences on populations and economies [10].

In this paper, we present a proof-of-concept of this approach. To the best of our knowledge, this approach is novel and potentially highly promising for pattern detection of wave-like phenomena.

II. APPROACH

This approach incorporates temporal, geometric and topological information to detect persistent repeating patterns in fluid and climate simulations.

Corresponding author: G Muszynski, muszyng@liverpool.ac.uk
[1]Department of Computer Science, University of Liverpool, United Kingdom [2]Lawrence Berkeley National Laboratory, CA, United States

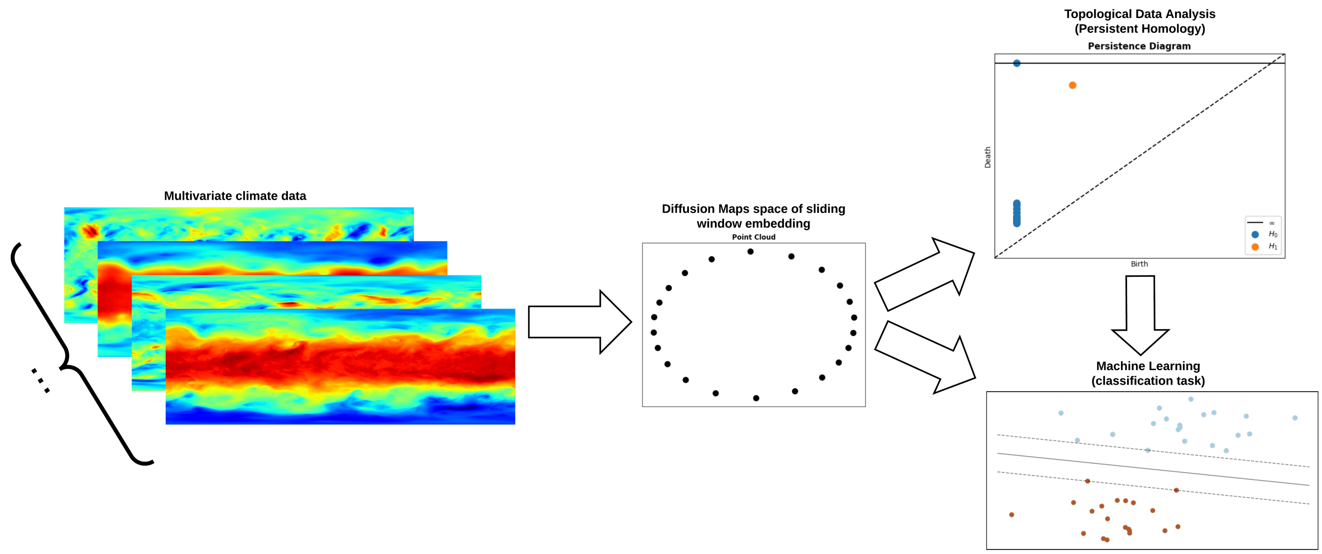

Fig. 1. The flowchart illustrates the approach to topological pattern detection in fluid or climate simulation data. The inputs are snapshots of a multivariate climate dataset. The sliding window coordinate embedding and diffusion maps are applied to represent the data as a low-dimensional point cloud. This embedding reveals a topological structure that is a discrete representation of repeating patterns in the data. The output is a set of persistence diagrams and their corresponding set of point clouds. Finally, there are two ways of detecting patterns: i) by measuring the 1-dimensional homology (orange dot marked as H_1 class in the diagram); ii) by directly applying a machine learning classifier to the point cloud or the persistence diagram suitably vectorized.

We start by describing the high-dimensional representation of the data produced by the sliding window coordinate embedding and which can then be compressed into a low-dimensional embedding using diffusion maps. Next, we explain the idea behind the topological data analysis, in particular, 1-dimensional persistent homology that can infer qualitative information about the structure and shape of the data. In addition, we mention where and how machine learning methods can be integrated into this approach.

A. Sliding Window Embedding and Manifold Learning

The evolution of a dynamical system on some attractors (i.e. its trajectory) can be reconstructed from measured or simulated time series data using Takens time-delay or sliding window embedding[6]. In our approach the sliding window embedding is applied to raw 2-dimensional snapshots of fluid flow or climate simulation data, which is an extension of the original embedding theorem applied to 1-dimensional time series [11].

If we denote fluid or climate simulation as a sequence of scalar fields indexed by the timesteps ($t \geq 0$). For given positive integers W (width) and H (height), a simulation with $W \times H$ grid points is a function

$$X : Z_+ \longrightarrow R^{W \times H} \tag{1}$$

Consider a sequence of scalar fields $\{X_i\}$, $i = 1, 2, 3, ..., n$ for a given integer $m \geq 0$ (the dimension), and a real number $\tau \geq 0$ (the delay / lag), the sliding window (SW) coordinate embedding of X at time t is defined as a vector

$$SW_{m,\tau}X(t) = \begin{bmatrix} X(t) \\ X(t+\tau) \\ \vdots \\ X(t+m\tau) \end{bmatrix} \in R^{W \times H \times (m+1)} \tag{2}$$

By applying manifold learning (i.e., the diffusion maps algorithm [7]) to the sliding window coordinate embedding space, a new low-dimensional representation (a point cloud) in the diffusion space (probabilistic space) is provided. The algorithm is a nonlinear technique for dimensionality reduction (or feature extraction) of data according to parameters of its intrinsic geometric structure. In general, the diffusion maps algorithm approximates the Laplace-Beltrami operator associated with the Gaussian kernel $k(X_i, X_j) = e^{\frac{-\|SW_{m,\tau}X_i - SW_{m,\tau}X_j\|}{\epsilon}}$, where ϵ is the kernel scale parameter. The operator is defined as follows

$$L\phi_j = \lambda_j \phi_j, \tag{3}$$

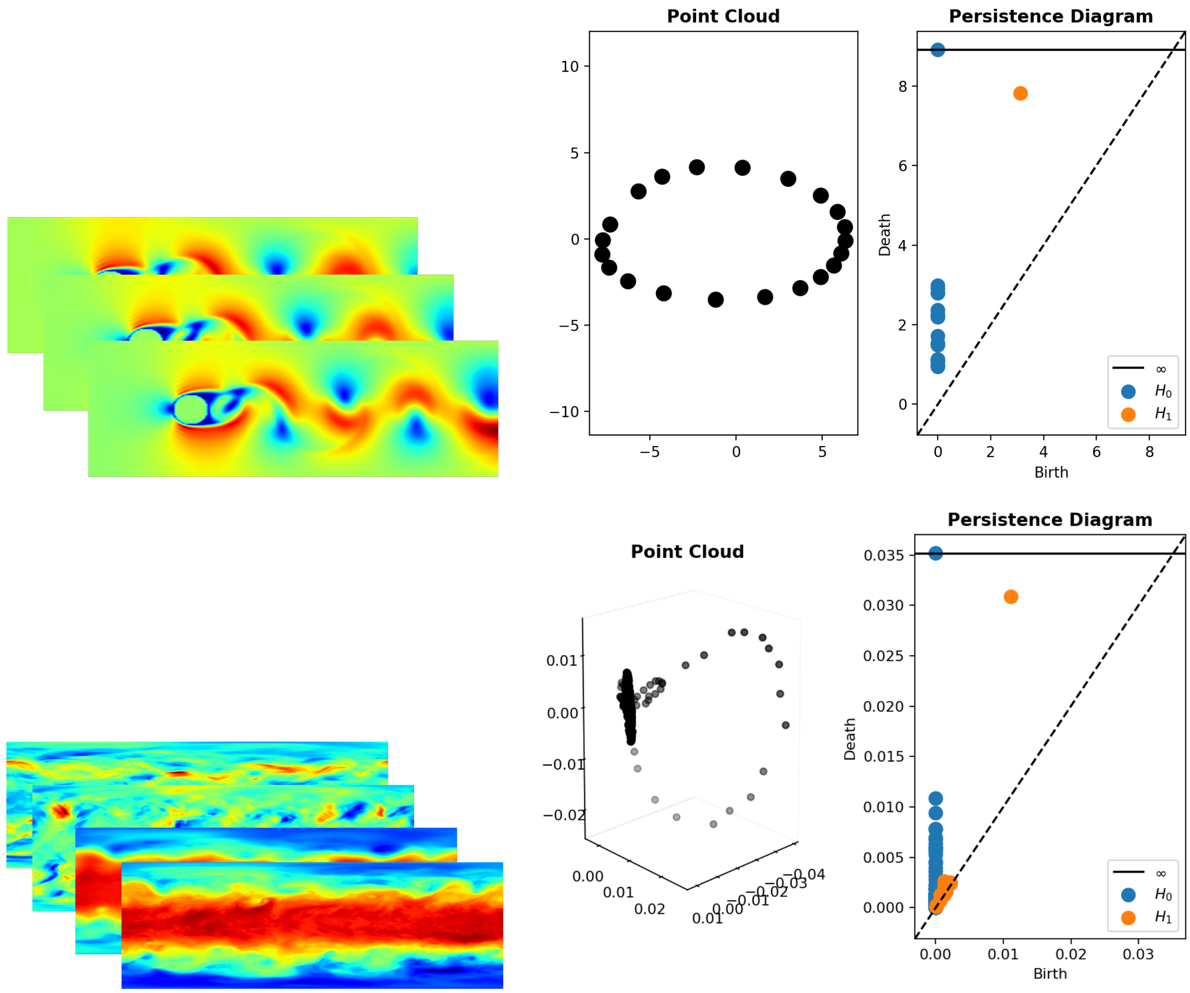

Fig. 2. Visualization of preliminary results for fluid flow (upper row; left) and Modern-Era Retrospective Analysis for Research and Application, Version 2 (MERRA-2) data (lower row; left) and their corresponding low-dimensional point cloud embeddings and persistence diagrams.

where L is a discrete Laplacian operator associated with the Gaussian kernel and ϕ_j are the eigenfunctions (eigenvectors) capturing temporal information of the coherent patterns and λ_j are their associated eigenvalues.

The obtained diffusion embedding space (Laplace-Beltrami eigenfunctions) is a more robust metric to "topological noise" in persistent homology computations [12].

B. Topological Data Analysis & Machine Learning

Topology is a branch of mathematics that studies properties of "shapes" (structures) that are preserved under continuous deformations (e.g., stretching, bending or folding but without tearing). In algebraic topology, homology is a general way of describing these properties (equivalence classes of loops). For example, one of the primary properties is the number of connected components in a topological space. Whereas the connected components are essential topological descriptors, the number of 1-dimensional "holes" (loops) is a property of higher dimensional structures. Topological data analysis can be broadly described as as a collection of data analysis techniques (including persistent homology) that infer qualitative information about the topological structure of data [8]. In recent years persistent homology has been successfully applied in many areas of science, including complex networks, signal processing and computer vision [5], [13]. Persistent homology is a tool that adapts the homology to finite metric spaces or point clouds (PC) in which 1-dimensional "holes" are tracked in a set of simplicial complexes that is constructed on top of the point cloud by adding edges and triangles. Let (PC, d_{PC}) be a point cloud, where d_{PC} is a distance function (in this case the diffusion distance [7]). The Rips complex $R_\alpha(PC)$ at scale parameter $\alpha \geq 0$ is the collection of subsets of PC such that $d_{PC}(PC_i, PC_j) \leq \alpha$. There is an inclusion $R_\alpha(PC)$ in $R_\beta(PC)$ for any $\alpha \leq \beta$. The sequence of inclusions is called a Rips filtration. The *Birth* is the distance in the Rips filtration at which a loop first appears and the *Death* is the distance at which it disappears. *Birth-Death* pairs are represented as points on a persistence diagram, as shown in Figure 1. The loop that dies later is geometrically larger and it usually means that those features are the most significant in the studied point cloud (data). Finally, one can apply a machine learning classifier to the persistence diagrams or directly to the point clouds in diffusion space [14], [15], [16].

III. TOWARDS REAL DATA - PROOF OF CONCEPT

Following the steps described above in our approach, we present results from two simple experiments with real data: i) canonical fluid flow simulations; ii) a climate reanalysis product (Modern-Era Retrospective Analysis for Research and Application, Version 2). First, small patches (100×100 pixels) have been extracted from each timestep of the data. Next, the sliding window embedding, diffusion maps algorithm and persistent homology (the Rips complex) have been applied to sequence of patches.

Figure 2 shows examples of the fluid flow and climate data and their corresponding low-dimensional representation (point clouds) with associated persistence diagrams. Note that in both examples there is a persistent repeating pattern represented by loop structures. The loop structures have been detected by the persistent homology algorithm, as shown by the furthest orange dots (H_1 class) from the diagonal lines in the persistence diagrams.

Thus we have shown that a combination of dynamical systems theory, manifold learning and topological data analysis can facilitate a new way of detecting persistent repeating patterns in fluid flow and climate simulation data. We are continuing to develop this approach for application to atmospheric wave phenomena and atmospheric blocking patterns.

ACKNOWLEDGMENTS

We thank Marielle Pinheiro and Paul Ullrich from the University of California, Davis for sharing their expertise on atmospheric blocking. Grzegorz Muszynski and Vitaliy Kurlin would also like to acknowledge Intel for supporting the IPCC at University of Liverpool, UK through the Big Data Center at Lawrence Berkeley National Lab, Berkeley, CA.

REFERENCES

[1] S. B. Prabhat, V. Vishwanath, E. Dart, M. Wehner, W. D. Collins, *et al.*, "Teca: Petascale pattern recognition for climate science," in *International Conference on Computer Analysis of Images and Patterns*, pp. 426–436, 2015.

[2] C. A. Shields, J. J. Rutz, L.-Y. Leung, F. M. Ralph, M. Wehner, B. Kawzenuk, J. M. Lora, E. McClenny, T. Osborne, A. E. Payne, *et al.*, "Atmospheric river tracking method intercomparison project (artmip): project goals and experimental design," *Geoscientific Model Development*, vol. 11, no. 6, pp. 2455–2474, 2018.

[3] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.

[4] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, and C. Pal, "Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," in *Advances in Neural Information Processing Systems*, pp. 3402–3413, 2017.

[5] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[6] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical systems and turbulence, Warwick 1980*, pp. 366–381, Springer, 1981.

[7] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and computational harmonic analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[8] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[9] D. F. Rex, "Blocking action in the middle troposphere and its effect upon regional climate," *Tellus*, vol. 2, no. 4, pp. 275–301, 1950.

[10] D. L. Swain, M. Tsiang, M. Haugen, D. Singh, A. Charland, B. Rajaratnam, and N. S. Diffenbaugh, "The extraordinary california drought of 2013/2014: Character, context, and the role of climate change," *Bull. Am. Meteorol. Soc*, vol. 95, no. 9, pp. S3–S7, 2014.

[11] C. J. Tralie and J. A. Perea, "Quasi periodicity quantification in video data, using topology," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1049–1077, 2018.

[12] P. Bendich, T. Galkovskyi, and J. Harer, "Improving homology estimates with random walks," *Inverse Problems*, vol. 27, no. 12, p. 124002, 2011.

[13] G. Carlsson, "Topological pattern recognition for point cloud data," *Acta Numerica*, vol. 23, pp. 289–368, 2014.

[14] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt, "A stable multi-scale kernel for topological machine learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4741–4748, 2015.

[15] G. Kusano, K. Fukumizu, and Y. Hiraoka, "Kernel method for persistence diagrams via kernel embedding and weight factor," *Journal of Machine Learning Research*, vol. 18, no. 189, pp. 1–41, 2018.

[16] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: A stable vector representation of persistent homology," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 218–252, 2017.