Climate Science at the Interface Between Topological Data Analysis and Dynamical Systems Theory

Davide Faranda, Théo Lacombe, Nina Otter, and Kristian Strommen

The authors are hosting an AMS sponsored Mathematics Research Community (MRC) on novel applications of topological data analysis (TDA) and dynamical systems theory to the study of climate change and weather forecasting. In this *Notices* article we introduce some of the big challenges in climate science, and describe how methods from TDA and dynamical systems theory can help tackle these. We hope to encourage applications to the MRC from mathematicians with a background in applied algebraic topology and scientists working on climate or meteorology, both from academia and industry.

Challenges in climate science. Understanding how the climate system is responding to increasing levels of anthropogenic emissions is one of today's key scientific challenges. Some aspects of this are very well understood: most fundamentally, the atmosphere as a whole will continue to warm up, and its capacity to carry moisture will therefore also increase. When the plethora of climate models (which numerically integrate the primitive equations describing the evolution of the climate system) are asked to simulate the future climate given continued greenhouse gas emissions, they all broadly agree on these points. Such changes to temperature and humidity are often referred to by climate scientists as the "thermodynamical" response. However, when it comes to regional changes, the details of which will dominate both the decision making for adaptation strategies and people's lived experience, there is considerably more uncertainty. There is perhaps no better example of this than the question of how the variability of the northern hemisphere jetstream will change in the future. North America and Eurasia have experienced dramatic examples of extreme weather in recent years, and these events are often related to atypical jetstream configurations, such as "blocking" events, where a persistent high-pressure system forces the jet to deviate from its usual trajectory. There is naturally intense interest in understanding how the frequency and severity of such events may change with global warming: these are often viewed as part of the "dynamical" response. The "dynamical" and "thermodynamical" response give a first-order approximation of the total response. Unfortunately, climate models can often disagree quite dramatically on the "dynamical" response, and theoretical understanding is still

Davide Faranda is directeur de recherche CNRS at Paris-Saclay University. His email address is `davide.faranda@cea.fr`.

Théo Lacombe is maitre de conférénce in the Laboratoire d'Informatique Gaspard Monge, Université Gustave Eiffel. His email address is `theo.lacombe@univ-eiffel.fr`.

Nina Otter holds an Inria Starting Faculty Position at Inria-Saclay. Her email address is `nina-lisann.otter@inria.fr`.

Kristian Strommen is a senior postdoctoral researcher in climate science at the University of Oxford. His email address is `kristian.strommen@physics.ox.ac.uk`.

DOI: https://doi.org/10.1090/noti2864

lacking, leading to pervasive uncertainty, especially with regards to the boreal winter months (December–January–February) [WBM+18]. The answers to simple sounding questions such as "will western Europe see increased risk of flooding?" or "will East Coast USA see increased risk of extreme cold snaps?" therefore remain unclear.

Climate as a dynamical system. One can view the climate as a dynamical system [GL20], and thus conceptually model it using a stochastic differential equation:

$$\frac{d}{dt}X = C(X) + f + \epsilon. \qquad (1)$$

Here X is a state vector encoding all the relevant physical variables (temperature, humidity, etc.) at every location in the atmosphere, land surface, and ocean; C is a smooth function which determines the evolution of X in time (and is therefore encoding things like the Navier–Stokes equations); and f is a forcing term, which in an equilibrium context would be 0, and in a global warming context would represent the effect of emissions. The ϵ term represents "noise." For a given fixed forcing term f, this equation determines the trajectory taken by X through its ambient phase space. For reasonable parameter choices, the climate system is stable, and the trajectory thereby settles onto its attractor, which in turn determines the statistics of the climate. The challenge of climate science can therefore be thought of as understanding how the attractor changes as a function of f = anthropogenic emissions. For example, does the new attractor contain a higher density of trajectories corresponding to extreme weather?

A key reason this is such a challenging question is that the dimensionality of the phase space (the size of the vector X with respect to a sufficiently accurate finite approximation of space-time) is in the billions. This is many orders of magnitude greater than both (i) the number of available independent observations; and (ii) the dimensionality of even the most state-of-the-art climate model simulations. This mismatch is exacerbated even further if one is interested in understanding extremes, which by definition make up only a tiny proportion of observed or simulated events.

Weather regimes and the Lorenz butterfly. Given the above discussion, a natural strategy is to try to artificially reduce the dimensionality in some way. In the context of the northern hemisphere jetstream, this is often done using the notion of "weather regimes" [HSF+17]. The basic idea is to find a small subset of dynamically relevant large-scale atmospheric circulation patterns (the "regimes") that dominate the low-frequency variability; one can then study the impact of forcing on the dynamical behavior of the regimes (e.g., their relative frequency of occurrence). An example is given in Figure 1, which shows the North Atlantic Oscillation, a large dipole of atmospheric pressure between Greenland and the Atlantic ocean, which is the dominant mode of variability in boreal winter. Knowing whether the NAO is in its positive or negative phase gives a good first-order approximation of the jetstream, and, consequently, of winter weather in Europe and eastern North America. It gives one possible regime-view of the Euro-Atlantic circulation. Strong NAO events are frequently linked to extreme surface weather, such as the exceptionally warm European winter of 2019–2020 [HDS+20].

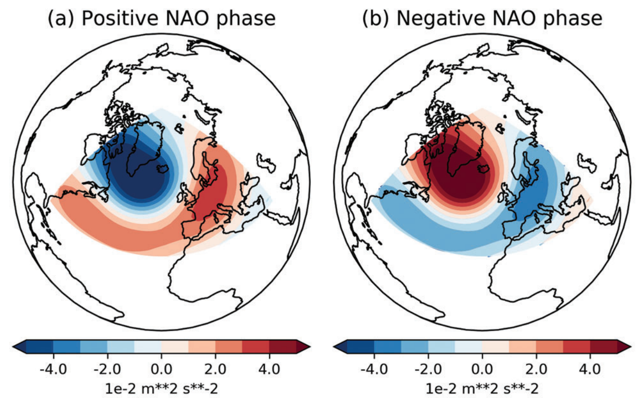

Figure 1. The positive (a) and negative (b) phases of the North Atlantic Oscillation (NAO), the dominant pattern of variability in the Euro-Atlantic wintertime circulation. The pattern is a dipole of atmospheric pressure anomalies, measured here as the first empirical orthogonal function of geopotential height at 500hPa. The two phases can be viewed as one of several possible ways to decompose the Euro-Atlantic circulation into distinct regimes.

A key reason why this perspective can be so powerful is that in some cases the response to a forcing f can be entirely understood in terms of the regime dynamics. The motivating example here is the "Lorenz '63" system [Lor63]. This system can be loosely thought of as an extremely severe truncation of Equation (1) to just 3 variables, with $f = \epsilon = 0$; it is famous for its "strange attractor," which, besides resembling a butterfly (see Figure 2a), encodes the first example of a chaotic dynamical system. Palmer, Corti, and Molteni [CMP99] showed that if the forcing vector f is made nonzero, the overall shape of the attractor does not change: the two "wings" remain in the same place. Instead, the system simply starts spending more time in whichever of the two wings the vector f is pointing towards, see Figure 2b. If the two wings are interpreted as two regimes, this says that in the Lorenz '63 system, the effect of forcing is determined by a single number, namely the relative frequency of occurrence of the two regimes, with the regimes themselves remaining invariant.

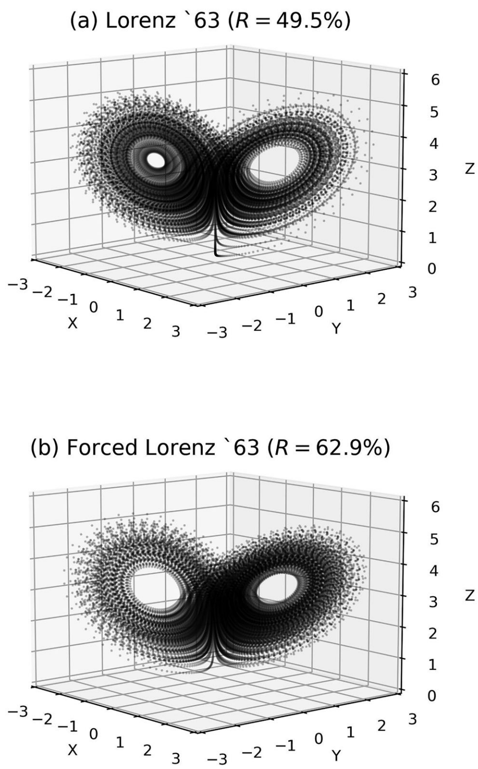

Figure 2. In (a), 30,000 samples from a long integration of the Lorenz '63 system. In (b), the same but with a nonzero constant forcing vector added which roughly points toward the right-hand "wing," following the set-up of [CMP99]. The value of R in each plot is the approximate proportion of time spent in the right-hand wing ($x > 0$). Lorenz '63 data kindly provided by J. Dorrington.

Palmer famously argued that the same may be true of real-world weather regimes [Pal99]. In other words, that the effect of global warming on the northern hemisphere jetstream may be to modulate the dynamical behavior of a small set of invariant regimes. If true, questions such as "will East Coast USA see increased risk of extreme cold snaps?" might be largely reducible to a determination of (a) how different weather regimes modulate the risk of cold snaps; (b) how the regime dynamics will change (their occurrence, temporal persistence etc.); and (c) a potential scaling factor to represent the much more well understood "thermodynamic" changes (to account for, e.g., the fact that temperatures will be warmer on average everywhere). Crucially, invariance of the regimes implies that point (a) can be estimated using *historical* observations. This strategy thereby offers a potentially dramatic simplification to the question of how midlatitude extreme weather will change.

Using persistent homology to diagnose weather regimes. It turns out that the main difficulty in implementing Palmer's strategy for understanding midlatitude climate change is actually determining what the "correct" weather regimes are. In the 3-variable Lorenz '63 system one can simply look at the attractor and identify the two distinct wings by eye. The dynamics of the jetstream is by contrast a very high-dimensional complex interplay between multiple physical variables confounded by considerable chaotic noise. A variety of strategies have been proposed [HSF+17], which typically involve using a clustering algorithm applied to a single physical variable whose dimensionality has been cut down using empirical orthogonal functions. The different strategies amount to inequivalent definitions of what a weather regime actually *is*. They all thereby suggest a different set of weather regimes that are not in general comparable and are far from being invariant under global warming [DSFM22]. Furthermore, when applied to different idealized representations of the atmosphere (of which the Lorenz '63 system is one example), they fail to identify the correct regime behavior in one or more cases [SCDO23]. It is likely that this is at least in part because these methods rely on a truncation of the true system which is far too severe. What is needed therefore is a method that allows one to efficiently "see" the structure in a more high-dimensional space. Enter persistent homology.

Persistent homology (PH) is one of the most successful methods in the field of topological data analysis [Car09, Oud15]. It is particularly useful for applications in which one is interested in studying data sets that depend on specific parameters of interest (e.g., studying points that have a specific density estimate, or thickenings of points), and for which one might not be able to accurately estimate the value of the parameter. In such cases, instead of

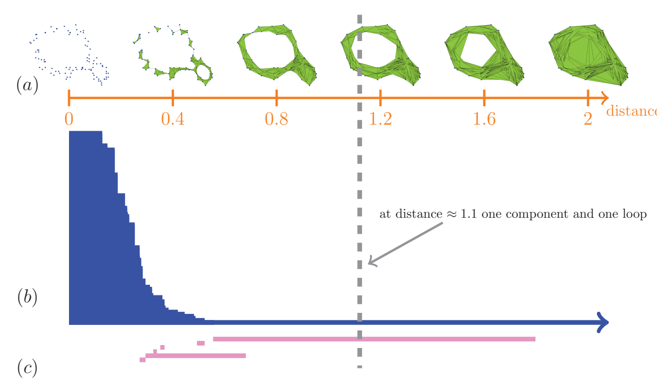

Figure 3. Persistent homology pipeline: (a) we associate a 1-parameter nested sequence of spaces ("filtration") to a metric space; barcode plots for (b) 0-cycles ("components") and (c) 1-cycles ("holes"). Each interval in a barcode plots describes the persistence of a component or hole.

estimating the optimal density value of the points, or the optimal thickening, PH allows us to study how the topological invariants (number of components, holes, voids) of the data set vary as a function of the density threshold or thickening parameter. The output of PH is a disjoint collection of intervals (collectively called the "barcode") with the length of each interval corresponding to the persistence of a specific topological invariant, see Figure 3.

PH has been shown to be able to capture local and global topological and geometric properties, as well as signals of different orders of magnitude. Crucial for applications to climate science is that main algorithms for PH do not depend on the dimension of the system that one is trying to study, and that PH is, in a precise sense, robust with respect to perturbations in the system. These properties of PH make it particularly suited for addressing the problem of determining weather regimes in a high-dimensional, noisy system, provided that the regimes correspond in some sense to topological features. A key observation of [SCDO23] is that this appears to be the case. This is already clear for the Lorenz '63 system, where the two "wings" are associated with the two holes, and PH is able to not just detect the existence of the two homological cycles but also compute "good" representatives of these that can be *visualized*, see Figure 4. Similarly, in other idealized models used to study weather regimes, and in low-dimensional truncations of atmospheric data, one finds that the regimes always correspond to a mixture of loops, holes and connected components. These can in all cases be efficiently and correctly identified with PH, whereas previously considered approaches would fail in one or more cases.

To summarize, PH represents a powerful tool for probing the high-dimensional dynamics of the jetstream, and there is good reason to believe that the topological features it might extract will encode physically meaningful weather regimes. Put informally, PH may be able to "see" the regimes of the real atmosphere as clearly as our human eyes can see the regimes in the Lorenz butterfly.

Open problems. There are several open problems that need to be addressed before techniques from topology can be used to study weather regimes in large data sets of high-dimensional climate data: these and others will be addressed at the MRC.

One such question concerns theoretical and computational improvements needed for developing algorithms for optimal PH representatives for weather data. For a recent survey on several approaches to define optimal representatives, see [LTHP+21]. Computing optimal representatives allows one to sensibly visualize the topological information that PH detects, essential for applications to climate science where one needs to verify that the

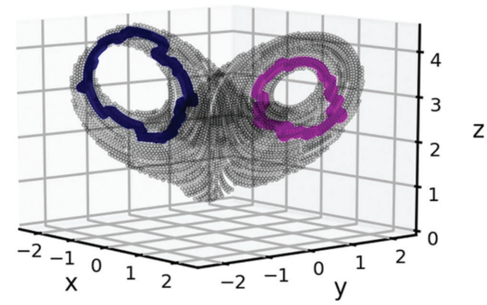

Figure 4. An illustration of two loops in the Lorenz '63 system. The homological cycles were identified using PH; optimal representatives of these were then computed, and are visualized in dark blue and purple thick lines.

topological cycles correspond to physically sensible phenomena. Having such a suitable notion of topological representatives then raises the question: how can these be used to study the dynamical properties of the weather systems? In particular, how can we best approximate the system in terms of transitions between a finite set of topologically defined states? What dynamical properties can one expect of regimes in general? Can we derive general principles for dynamical systems with arbitrary topological signals? Here we will make use of recent developments from dynamical systems theory, which allow dynamical features (such as the number of degrees of freedom or temporal persistence) to be analyzed robustly using instantaneous samples [FMY17].

Another question relates to statistical analysis. Our topological approach to regimes essentially equates the existence of regimes in a dynamical system with the existence of nontrivial topological structure. Here the term "nontrivial" is explicitly a statistical notion: a randomly drawn sample from a noisy system could produce nontrivial homology at some degree purely by chance. How can we assess the statistical significance of any regimes found using PH? Can we meaningfully perform statistical optimization across multiple topological descriptors to study weather systems? To address these questions we will rely on several different approaches developed in statistical topological data analysis, such as [BS22, BCCS+11, CFL+17].

Conclusions. We have highlighted how methods from computational topology represent a potentially very powerful tool for studying the effect of global warming on northern hemisphere weather, via the framework of weather regimes; this will be the focus for our MRC. However, there are many other potential applications of topological data analysis to climate science [MKK+18, TMD+20] and dynamical systems [YB20], including the use of more sophisticated topological information [GS23]. For example, many important components of the climate system exhibit oscillatory behavior, such as the El Niño Southern Oscillation, which has a big influence on year-to-year variations in global temperatures. Viewed appropriately, such oscillations correspond to loops through phase space which can be detected and studied using PH. The authors believe that there is great potential for the use of methods from topology in climate science and dynamical systems theory more broadly.

In order for this potential to be realized, important theoretical and algorithmic mathematical challenges in topological data analysis need to be tackled. The goal of our MRC is to begin this process.

References

[BCCS+11] Gérard Biau, Frédéric Chazal, David Cohen-Steiner, Luc Devroye, and Carlos Rodríguez, *A weighted k-nearest neighbor density estimate for geometric inference*, Electron. J. Stat. **5** (2011), 204–237, DOI 10.1214/11-EJS606. MR2792552

[BS22] Omer Bobrowski and Primoz Skraba, *On the Universality of Random Persistence Diagrams*, arXiv e-prints (July 2022), available at `arxiv:2207.03926`.

[Car09] Gunnar E. Carlsson, *Topology and data*, Bulletin of the American Mathematical Society **46** (2009), 255–308, `https://api.semanticscholar.org/CorpusID:1472609`.

[CFL+17] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman, *Robust topological inference: distance to a measure and kernel distance*, J. Mach. Learn. Res. **18** (2017), Paper No. 159, 40. MR3813808

[CMP99] S. Corti, F. Molteni, and T. N. Palmer, *Signature of recent climate change in frequencies of natural atmospheric circulation regimes*, Nature **398** (1999), no. 6730, 799–802.

[DSFM22] Josh Dorrington, Kristian Strommen, Federico Fabiano, and Franco Molteni, *CMIP6 models trend toward less persistent European blocking regimes in a warming climate*, Geophysical Research Letters **49** (2022), no. 24, e2022GL100811.

[FMY17] Davide Faranda, Gabriele Messori, and Pascal Yiou, *Dynamical proxies of North Atlantic predictability and extremes*, Scientific reports **7** (2017), no. 1, 41278.

[GL20] Michael Ghil and Valerio Lucarini, *The physics of climate variability and climate change*, Rev. Modern Phys. **92** (2020), no. 3, 035002, 77, DOI 10.1103/revmodphys.92.035002. MR4166837

[GS23] Michael Ghil and Denisse Sciamarella, *Dynamical systems, algebraic topology and the climate sciences*, Nonlinear Processes in Geophysics **30** (2023), no. 4, 399–434.

[HSF+17] Abdel. Hannachi, David M. Straus, Christian L. E. Franzke, Susanna Corti, and Tim Woollings, *Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere*, Reviews of Geophysics **55** (2017), no. 1, 199–234, `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000509`.

[HDS+20] Steven C. Hardiman, Nick J. Dunstone, Adam A. Scaife, Doug M. Smith, Jeff R. Knight, Paul Davies, Martin Claus, and Richard J. Greatbatch, *Predictability of European winter 2019/20: Indian Ocean dipole impacts on the NAO*, Atmospheric Science Letters **21** (2020), no. 12, e1005, `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.1005`.

[LTHP+21] Lu Li, Connor Thompson, Gregory Henselman-Petrusek, Chad Giusti, and Lori Ziegelmeier, *Minimal cycle representatives in persistent homology using linear programming: An empirical study with user's guide*, Frontiers in Artificial Intelligence **4** (2021), `https://www.frontiersin.org/articles/10.3389/frai.2021.681117`.

[Lor63] Edward N. Lorenz, *Deterministic nonperiodic flow*, J. Atmospheric Sci. **20** (1963), no. 2, 130–141, DOI 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2. MR4021434

[MKK+18] G. Muszynski1, Kashinath. K., V. Kurlin, M. Wehner, and Prabhat, *Towards topological pattern detection in fluid and climate simulation data*, 8th International Workshop on Climate Informatics, 2018.

[Oud15] S.Y. Oudot, *Persistence theory: From quiver representations to data analysis*, Mathematical Surveys and Monographs, American Mathematical Society, 2015, `https://books.google.fr/books?id=if8dCwAAQBAJ`.

[Pal99] Timothy N. Palmer, *A nonlinear dynamical perspective on climate prediction*, Journal of Climate **12** (1999), no. 2, 575–591.

[SCDO23] Kristian Strommen, Matthew Chantry, Joshua Dorrington, and Nina Otter, *A topological perspective on weather regimes*, Climate Dynamics **60** (2023), no. 5-6, 1415–1445.

[TMD+20] Sarah Tymochko, Elizabeth Munch, Jason Dunion, Kristen Corbosiero, and Ryan Torn, *Using persistent homology to quantify a diurnal cycle in hurricanes*, Pattern Recognition Letters **133** (2020), 137–143, `https://www.sciencedirect.com/science/article/pii/S0167865520300611`.

[WBM+18] Tim Woollings, David Barriopedro, John Methven, Seok-Woo Son, Olivia Martius, Ben Harvey, Jana Sillmann, Anthony R Lupo, and Sonia Seneviratne, *Blocking and its response to climate change*, Current climate change reports **4** (2018), 287–300.

[YB20] Gökhan Yalnız and Nazmi Burak Budanur, *Inferring symbolic dynamics of chaotic flows from persistence*, Chaos **30** (2020), no. 3, 033109, 11, DOI 10.1063/1.5122969. MR4071228

Credits

Figure 1 re-used with permission from [SCDO23], published originally in *Climate Dynamics* (Springer Journals) under a Creative Commons License: `http://creativecommons.org/licenses/by/4.0/`.

Figures 2 and 3 are courtesy of Kristian Strommen and Nina Otter.

Figure 4 adapted with permission from [SCDO23], published originally in *Climate Dynamics* (Springer) under a Creative Commons License: `http://creativecommons.org/licenses/by/4.0/`.