



Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets

Grzegorz Muszynski^{1,2}, Karthik Kashinath², Vitaliy Kurlin¹, Michael Wehner^{2,3}, and Prabhat²

¹Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, UK

²National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, CA 94720, USA

³Computational Research Division, Lawrence Berkeley National Laboratory, CA 94720, USA

Correspondence: Grzegorz Muszynski (gmszynski@lbl.gov)

Received: 22 February 2018 – Discussion started: 7 May 2018

Revised: 30 November 2018 – Accepted: 3 January 2019 – Published: 7 February 2019

Abstract. Identifying weather patterns that frequently lead to extreme weather events is a crucial first step in understanding how they may vary under different climate change scenarios. Here, we propose an automated method for recognizing atmospheric rivers (ARs) in climate data using topological data analysis and machine learning. The method provides useful information about topological features (shape characteristics) and statistics of ARs. We illustrate this method by applying it to outputs of version 5.1 of the Community Atmosphere Model version 5.1 (CAM5.1) and the reanalysis product of the second Modern-Era Retrospective Analysis for Research and Applications (MERRA-2). An advantage of the proposed method is that it is threshold-free – there is no need to determine any threshold criteria for the detection method – when the spatial resolution of the climate model changes. Hence, this method may be useful in evaluating model biases in calculating AR statistics. Further, the method can be applied to different climate scenarios without tuning since it does not rely on threshold conditions. We show that the method is suitable for rapidly analyzing large amounts of climate model and reanalysis output data.

- The identification process is critical in calculating statistics, including the frequency, location and intensity of extreme weather events under different climate change scenarios.
- It is the first step in evaluating how well a climate model captures physical features of extreme events and characterizing their changes under global warming.
- As high-performance computational technology continues to advance, there is an ever-increasing amount of data from climate model output, reanalysis products and observations that demands rapid and automated detection and characterization of extreme events.

This study is part of ongoing efforts to provide automated methods that are able to identify extreme weather and climate events in large climate datasets (Prabhat et al., 2015; Ullrich and Zarzycki, 2017; Shields et al., 2018).

Extreme precipitation events in midlatitudes are often associated with atmospheric rivers (ARs). Since the early 1990s, there has been a growing interest in studying ARs (Zhu and Newell, 1994). ARs are long and narrow filaments of high concentrated water vapor in the lower troposphere. They are responsible for more than $\sim 90\%$ of the total poleward water vapor transport outside of the tropics (Newell et al., 1992; Newell and Zhu, 1994; Zhu and Newell, 1998). Most of them are associated with extreme winter storms and heavy precipitation events on the western coast of North America (Dettinger et al., 2011) and along the Atlantic European coasts (Fragoso et al., 2012; Lavers and Villarini, 2013). Due to the large amount of water that can be transported by a single AR, they are potentially of high risk to society and

1 Introduction

The importance of understanding the behavior of extreme weather events in a changing climate cannot be overstated. A first step towards this challenging goal is to identify extreme events in large datasets. Identifying such events remains an important challenge for the climate science community for the following reasons:

often cause extreme flooding or have other devastating impacts when they make landfall (Ralph and Dettinger, 2011; Dettinger and Ingram, 2013; Ralph et al., 2016). On the other hand, ARs are critical in contributing to mountain snowpack and refilling reservoirs, thus mitigating drought, in areas such as the western United States, as in California (Guan et al., 2010; Dettinger, 2013). Figure 1 shows two examples of simulated ARs that deposit large amounts of rainfall on California and Washington state.

The first challenge in extreme event detection is to construct a quantitative definition of the event (Ullrich and Zarzycki, 2017). Once properly defined, developing a scheme to identify and track events in time and space can proceed. The American Meteorological Society (AMS) glossary defines an AR as follows: “A long, narrow, and transient corridor of strong horizontal water vapor transport that is typically associated with a low-level jet stream ahead of the cold front of an extratropical cyclone. The water vapor in atmospheric rivers is supplied by tropical and/or extratropical moisture sources. Atmospheric rivers frequently lead to heavy precipitation where they are forced upward – for example, by mountains or by ascent in the warm conveyor belt. Horizontal water vapor transport in the midlatitudes occurs primarily in atmospheric rivers and is focused in the lower troposphere” (AMS, 2018). Note that this definition is qualitative and numerous methods have been proposed to make this quantitative and use them to detect ARs in regional and global climate data (Sellars et al., 2017), but none of these are free from a subjective thresholding of some physical variable. Many existing techniques that have been designed for objective detection of ARs are based on a fixed threshold of more than 20 kg m^{-2} of integrated water vapor (IWV) in the atmospheric column (Ralph et al., 2004) or more than $750 \text{ kg m}^{-1} \text{ s}^{-1}$ of integrated water vapor transport (IVT) (Sellars et al., 2017). Selecting appropriate threshold values of IWV or IVT in various climate scenarios remains an open challenge (Shields et al., 2018).

Some recent efforts focus on alternative approaches to characterize and detect extreme events, such as deep learning methods for pattern recognition (Liu et al., 2016; Racah et al., 2017), which use underlying features of datasets. In particular, the inherent design of these methods circumvents a critical challenge of event detection schemes, choosing suitable thresholds for different variables.

In this paper, we present an alternative approach to AR pattern recognition based on topological data analysis (TDA) (Ghrist, 2008; Carlsson, 2009, 2014) and machine learning (ML) (Kubat, 2015). Our approach uses TDA as a first step, followed by training a ML model to perform binary classification. TDA provides feature extraction tools using techniques from topology and computer science to study topological features of data (Carlsson, 2009, 2014). Topological features provide a unique and threshold-free way of describing crucial shape characteristics of physical phenomena, including weather events, in large datasets. We use a partic-

ular type of topological feature descriptor called connected regions (Edelsbrunner and Harer, 2010), which are obtained from scalar fields on a latitude–longitude grid (see Fig. 2; Stage 1). The descriptors from positive and negative examples, i.e., events that are ARs and those that are not ARs, are then used in training a support vector machine (SVM) classifier (Cortes and Vapnik, 1995; Chang and Lin, 2011), which is a ML model used for binary classification. We note here that the training labels are generated by a heuristic algorithm that uses thresholds on IWV to classify events as ARs or non-ARs. In summary, the feature descriptors extract relevant topological information from a given scalar field, which is then used for training the ML classifier to perform the task of binary classification (see Fig. 2; Stage 2). To the best of our knowledge, this is the first framework based on TDA and ML that has been introduced for recognizing weather patterns in large climate datasets. In this study, we focus on ARs making landfall along the west coast of North America, but the method is easily extendable to other regions.

The key contributions of this paper are as follows: (i) we propose a novel method to identify ARs that is free from threshold selection, and (ii) we show that the framework of using TDA to extract topological feature descriptors and a ML classifier (SVM) provides high accuracy in recognizing AR patterns in both climate model output and reanalysis datasets across a range of spatial and temporal resolutions.

The rest of the paper is organized as follows: Sect. 2 describes datasets, the topological feature descriptors of ARs and non-ARs, the TDA algorithm and SVM classifier in more detail; Sect. 3 shows the results obtained with discussion; and Sect. 4 presents conclusions and future work.

2 Data and method

2.1 Data

In this study, we use both climate model simulation output generated by version 5.1 of the Community Atmosphere Model¹ (CAM5.1) (Eaton, 2011) and a reanalysis product from the second Modern-Era Retrospective Analysis for Research and Applications² (MERRA-2) (Gelaro et al., 2017), respectively.

The CAM5.1 climate model output is available at 25, 100 and 200 km spatial resolutions, and both 3-hourly and daily temporal resolutions, for the period of January 1979–December 2005 (Wehner et al., 2014). The MERRA-2 reanalysis data are at 50 km spatial resolution and 3-hourly temporal resolution (instantaneous snapshots), for the period

¹CAM5.1 data are provided by the Lawrence Berkeley National Laboratory (LBNL), Berkeley; National Energy Research Scientific Computing Center.

²MERRA-2 data are provided by the University of California, San Diego (UCSD); Center for Western Weather and Water Extremes (CW3E).

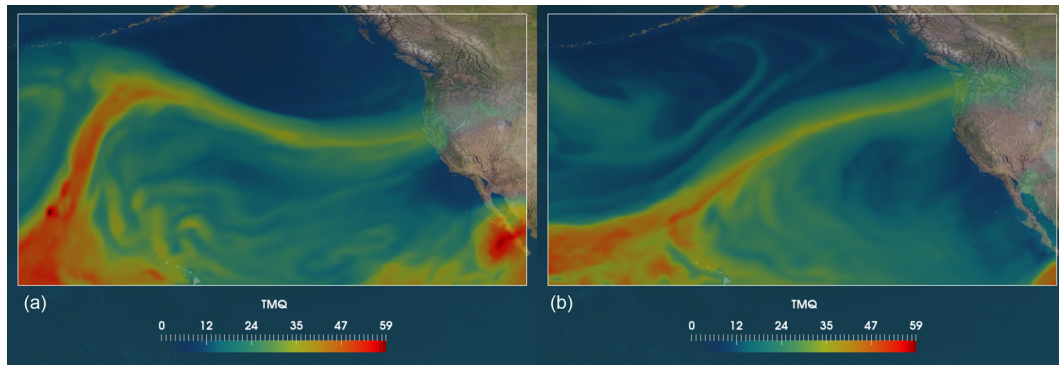


Figure 1. Sample images of atmospheric rivers – long filamentary structures reaching the west coast of the United States: **(a)** landfall in a region of northern California; **(b)** landfall in a region of Washington state. Shown is integrated water vapor (kg m^{-2}) from a simulation of version 5.1 of the Community Atmosphere Model (CAM5.1).

Table 1. Data sources (climate model and reanalysis datasets).

Climate model	Period	Temporal resolution	Spatial resolution
CAM5.1 (historical run)	1979–2005	3-hourly and daily	25 km
CAM5.1 (historical run)	1979–2005	3-hourly and daily	100 km
CAM5.1 (historical run)	1979–2005	3-hourly and daily	200 km
MERRA-2 (reanalysis product)	1980–2017	3-hourly	50 km

of January 1980–June 2017. A summary of datasets is listed in Table 1. For both the CAM5.1 output and MERRA-2 data, we use the total column IWV³. Note that this analysis could be performed on other relevant variables, including IVT, the vertically integrated vector product of wind and water vapor, which is another commonly used variable for identifying ARs (Sellars et al., 2017). However, we note that IWV is observable by satellite, whereas IVT is not. Although outside the scope of this paper, an AR identification algorithm based on IWV could offer an objective metric for evaluating both reanalysis products and climate models against observational data. We choose to use both 3-hourly and daily data because we anticipate the daily averages to smear out certain physical features of ARs. Further, 3-hourly data provide more event images labeled as ARs, which is useful for training in the machine learning model⁴.

Training a machine learning classifier, such as a SVM (see Sect. 2.2.2), requires labeled data of events that are ARs and non-ARs. In other words, each time step (snapshot) has to be tagged with a positive label (1 – if it contains an AR) or a negative label (0 – if it does not contain an AR). We use the parallel Toolkit for Extreme Climate events Analysis (TECA) (Prabhat et al., 2015) to obtain labels for training. The toolkit uses fixed threshold-based criteria (Ralph et al.,

³For the CAM5.1, this variable is called TMQ. For the MERRA-2 reanalysis data, it is called IWV. It is also called prw in the CF protocols.

⁴ML models tend to perform better when they have more training data.

2004) to determine if there is an AR in the given snapshot or not. The labels have been generated to for each dataset listed in Table 1. It is assumed that labels provided by TECA are “ground truth”.

2.2 Atmospheric river pattern recognition method

This subsection describes the two stages of the atmospheric river pattern recognition method (see Fig. 2) based on TDA (Carlsson, 2009, 2014) and ML (Kubat, 2015):

- *Stage 1:* This stage applies techniques from topology⁵ and algorithms from TDA to automatically infer relevant topological features from complex climate data including climate model output and reanalysis products. The TDA algorithm is based on the Union-Find data structure (Hopcroft and Ullman, 1973; Tarjan, 1975), which extracts topological feature descriptors of weather patterns, i.e., features of ARs and non-ARs, in a threshold-free way. These topological feature descriptors are called connected regions (Edelsbrunner and Harer, 2010) and are obtained from snapshots of global images on a latitude–longitude grid. The topological feature descriptors are provided as the input for the ML classifier in Stage 2.

⁵Topology is the branch of mathematics studying properties of geometric objects (e.g., 2-D grids) that are preserved under continuous deformations.

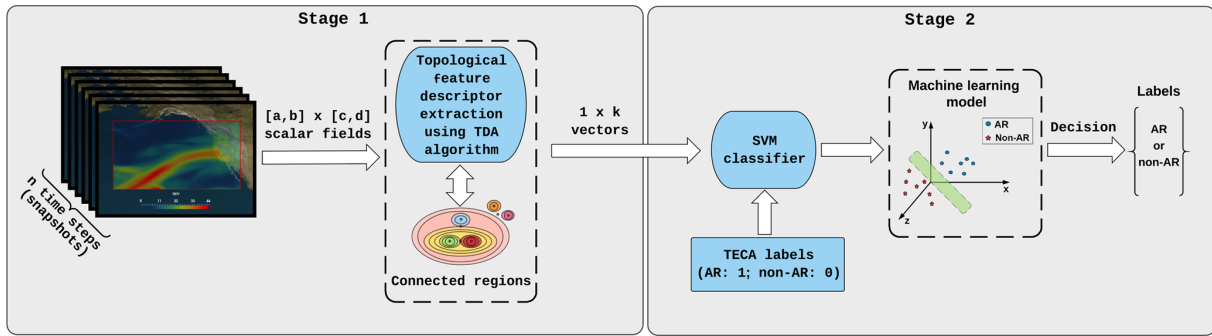


Figure 2. Block diagram of the AR pattern recognition method. Stage 1: the input of the method is a set of scalar fields (n snapshots) of size $[a, b] \times [c, d]$ on the latitude–longitude grid, where a , b , c and d are the dimensions of grid spatial extent in the real world. The topological data analysis (TDA) algorithm extracts connected regions from the snapshots of global images on the grid. The algorithm maintains the connected regions by varying IWV (referred to as TMQ in CAM5.1) values and dynamically keeps track of the regions in the grid. The connected regions are used as topological feature descriptors that characterize weather patterns (i.e., ARs and non-ARs). The descriptors are $1 \times k$ vectors, where k is greater than the maximal value of IWV/TMQ in a given set of scalar fields (in this case, $k = 60$). Stage 2: the vectors are stacked on top of each other to form a $n \times k$ matrix and are fed into the SVM classifier along with ground truth labels (i.e., AR: 1; non-AR: 0) provided by the Toolkit for Extreme Climate events Analysis (TECA) (Prabhat et al., 2015). Finally, the SVM finds a suitable hyperplane (the green surface shown in the figure) that can cleanly separate events into two groups (i.e., ARs and non-ARs). The output of the method is a set of n labels based on the decision made by the SVM classifier.

- *Stage 2:* In this stage, a binary classification task is performed using the ML classifier, SVM (Cortes and Vapnik, 1995; Chang and Lin, 2011). The classification task consists of two steps: (i) training the classifier to distinguish ARs from other weather events in the snapshots and (ii) testing the constructed SVM model on the unlabeled descriptors to separate events into two groups (i.e., ARs and non-ARs). The training process uses the topological feature descriptors (from Stage 1) and the ground truth labels (see Sect. 2.1) provided by TECA (Prabhat et al., 2015). The classifier performance is evaluated in the terms of accuracy, precision and sensitivity (see Sect. 2.3).

2.2.1 Stage 1: topological feature descriptors of ARs and non-ARs

The aim of this stage is to automatically characterize AR and non-AR events in raw climate data. Most existing methods have been designed to use thresholds for identification of ARs (Shields et al., 2018). In contrast, the approach proposed here is threshold-free by employing topological feature descriptors and, in particular, connected regions. This approach is a type of TDA that is inspired by persistence, which is a concept in applied topology that summarizes topological variations across all values of the scalar field under consideration (Ghrist, 2008; Edelsbrunner and Morozov, 2012; Carlsson, 2009, 2014).

Climate model output or reanalysis data may be represented as mapping from the grid to a set of real values, which in our case is the IWV over $[0, L]$, where L is the maximal value of the variable (here $L = 60 \text{ kg m}^{-2}$). It can be defined

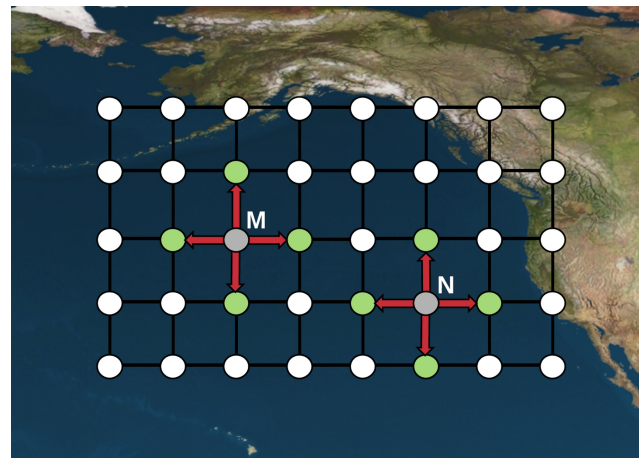


Figure 3. An illustration of four-connected neighborhood that is defined in the latitude–longitude grid in the plane with real coordinates. For example, each of the nodes M and N (gray points) has four neighbors, i.e., two nodes along the horizontal direction and two nodes along vertical direction (red arrows indicate neighboring nodes, i.e., green points).

as follows:

$$f : [a, b] \times [c, d] \rightarrow [0, L], \quad (1)$$

where a , b , c and d are the dimensions of the grid.

Every node (grid point) has four neighbors in the grid (except boundary nodes). In terms of point coordinates in the plane, the node at $(x, y) \in [a, b] \times [c, d]$ has four neighbors that have the coordinates $(x \pm 1, y)$ or $(x, y \pm 1)$. This is the so-called four-connected neighborhood, as shown in Fig. 3.

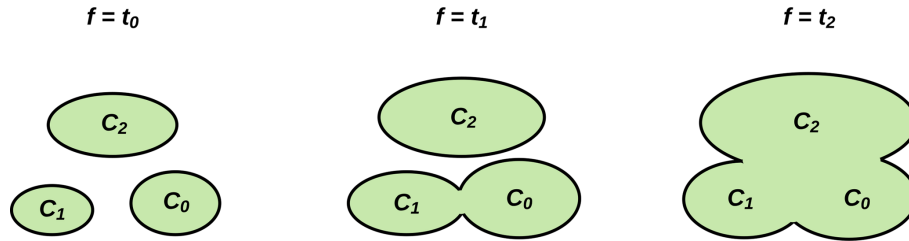


Figure 4. An illustration of the connected regions in the superlevel set (defined in Eq. 2) that are split into three pieces at value t_0 . They grow and merge first at value t_1 and then at t_2 when values of function f are systematically decreasing.

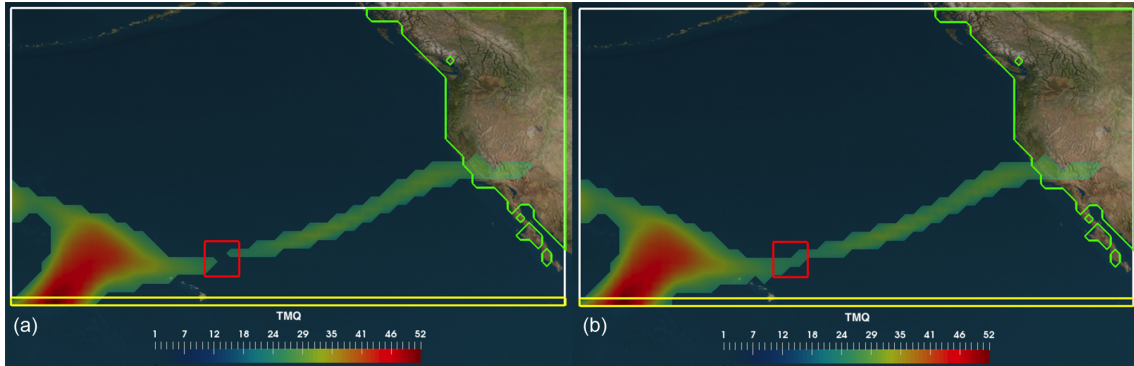


Figure 5. An illustration of finding connected AR regions over a specified search sector. In this example, the search for ARs is bounded by the latitude of the Hawaiian Islands (yellow line) and the west coast of North America (green line). (a) The red box indicates the location of two regions that are disconnected at some value ($IWV = t_1$). (b) At a new value ($IWV = t_2$), where $t_2 < t_1$, the two connected regions merge into one new connected region, forming a valid AR pattern. The IWV (kg m^{-2}) plotted in this example is from the CAM5.1 climate model.

Following the threshold-free approach in TDA, the evolution of connected regions in a superlevel set is monitored at every value t of function f . The superlevel set is a set of grid points in the domain of function f with scalar value greater than or equal to t . It is possible to mathematically express the superlevel set as follows:

$$f^{-1}[t, +\infty) = \{(x, y) \in [a, b] \times [c, d] : f(x, y) \geq t\}. \quad (2)$$

As t decreases, connected regions of $f^{-1}[t, +\infty)$ start to appear and grow and eventually merge into larger components.

Suppose there are three connected regions (C_0, C_1, C_2) at value t_0 in a superlevel set (defined in Eq. 2), as shown in Fig. 4. As values of f decrease, the component C_0 grows until eventually, at t_1 , it merges into the component of C_1 , which in turn merges into the component of C_2 at t_2 , and so on.

The above discussed approach of connected regions can be achieved by the TDA algorithm based on Union-Find data structure (U-F) (Hopcroft and Ullman, 1973). The algorithm determines connected regions of the grid by operating on sorted nodes by scalar values in decreasing order. The U-F data structure maintains the connected regions and keeps track of the evolution of these regions in the grid.

There are five main operations used in our TDA algorithm:

- i. form a new connected region and add the region to the data structure;
- ii. assign the right connected region to a given grid point;
- iii. check if the connected regions intersect a specified geographical location on the grid; e.g., we examine connected regions that intersect the west coast of North America and the latitude of the Hawaiian Islands, as shown in Fig. 5a;
- iv. merge two regions containing at least one same node into one new connected region, as shown in Fig. 5b; and
- v. track the evolution of a connected region (number of grid points in it) as IWV is varied.

The extracted information of evolution of connected regions is encoded in evolution plots. The plots show the recorded number of grid points in the region as values of IWV are systematically decreasing, as in Fig. 6. The horizontal axis t contains values of IWV in kg m^{-2} and the vertical axis $g(t)$ shows the number of grid points in the connected region. The vectors from these evolution plots are encoded as a matrix with n rows and k columns, where n is the number of time steps (snapshots) and k is the size of the topological feature descriptors returned by the TDA algorithm, as is

shown in Fig. 6. This $n \times k$ matrix serves as the input data to the support vector machine classifier, described in the next section.

2.2.2 Stage 2: applying SVM for classifying weather patterns

Support vector machine is a widely used machine learning method for binary classification (recognition) task (Cortes and Vapnik, 1995; Chang and Lin, 2011). The main objective of SVM classifier is to decide whether a particular pattern, an AR pattern, is present or not in a given snapshot extracted from global image. The SVM constructs a model based on the labeled topological feature descriptors in the training set and then uses it to predict the labels of the descriptors in the testing set. In general, the SVM finds the optimal hyperplane that separates two groups of patterns (ARs and non-ARs) by maximizing the margin between the separating boundary and the training points closest to it (support vector), as shown in Fig. 7.

Assume a training set of instance-label pairs (x_i, y_i) , $i = 1, \dots, N$, where $x_i \in \mathbb{R}^n$ and $y_i \in \{1, 0\}$. The solution of the optimization problem (finding the optimal hyperplane) is given by

$$\min_{w, b, \xi} \left(\frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \right), \quad (3)$$

subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0. \quad (4)$$

The penalty parameter of the error term takes only values greater than zero ($C > 0$) and $\xi_i \geq 0$ is a minimum error when two groups are not linearly separable (e.g., due to noise in training data). The samples $\{x_i\}$, where $x_i \in \mathbb{R}^n$, from the training set are mapped into a high-dimensional feature space F by means of the transformation $\phi(x_i)$, where $\phi(x) : \mathbb{R}^n \rightarrow F$. This transformation makes the samples of two groups (ARs and non-ARs) separable, as shown in Fig. 8. Then, the similarity between observations x_i and x_j is computed by a kernel function $K(x_i, x_j)$ that can be expressed as an inner product $\langle \phi(x_i), \phi(x_j) \rangle_F$ in the feature space F . Hence, it is sufficient to know $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_F$ rather than $\phi(x)$ explicitly (Burgess, 1998).

For this study, a radial basis function (RBF) kernel is chosen as it has been shown to achieve the best results in many applications. The RBF is defined as follows

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0, \quad (5)$$

where γ is the inverse of the standard deviation of the RBF kernel. The optimal configuration of parameters (C, γ) is found in the experiments by applying a loose grid search and fine grid search for these two parameters (Hsu et al., 2003).

Table 2. A confusion matrix (error matrix) is a way to present the performance of the method (typically, testing accuracy). The matrix reports the number of (i) false positives – cases when the model indicates that an AR exists, when it does not in the ground truth; (ii) false negatives – cases when the model indicates that an AR does not exist, while in fact it does in the ground truth; (iii) true positives – cases when the model indicates that an AR exists, when it does in the ground truth; (iv) true negatives – cases when the model indicates that an AR does not exist, when it does not in the ground truth.

	Label non-AR	Label AR
Predicted non-AR	True negatives	False positives
Predicted AR	False negatives	True positives

2.3 Evaluation metrics and preprocessing of data

In this subsection, we define the evaluation metrics that we use to assess the reliability of our AR pattern recognition method: classification accuracy score, confusion matrix, precision score and sensitivity score. Also, we explain the preprocessing step of the input to the SVM classifier to address the issues of imbalanced data (He and Garcia, 2009) and data normalization (standardization).

Classification accuracy score

Classification accuracy score is the ratio of correct predictions of ARs to total predictions made by the machine learning classifier (in percent). Training accuracy is the classification accuracy obtained by applying the classifier on the training data, while testing accuracy is the classification accuracy for the testing data. We present the classification accuracy scores for our method in Sect. 3.2.

Confusion matrix

A confusion matrix is a clear way to present the classification results of ARs with regard to testing accuracy of the machine learning classifier. The matrix has two rows and two columns, as shown in Table 2. The confusion matrices are shown in Sect. 3.3 and Appendix B for the SVM classifier.

Precision score

Precision score is a measure of the classifier's repeatability or reproducibility of ARs and can be computed using a confusion matrix. The score is the ratio of true positives to the sum of true positives and false positives. It is shown in Table 7 for the SVM classifier.

Sensitivity score

Sensitivity score is the proportion of actual ARs which are correctly identified as ARs by the classifier. The score is the

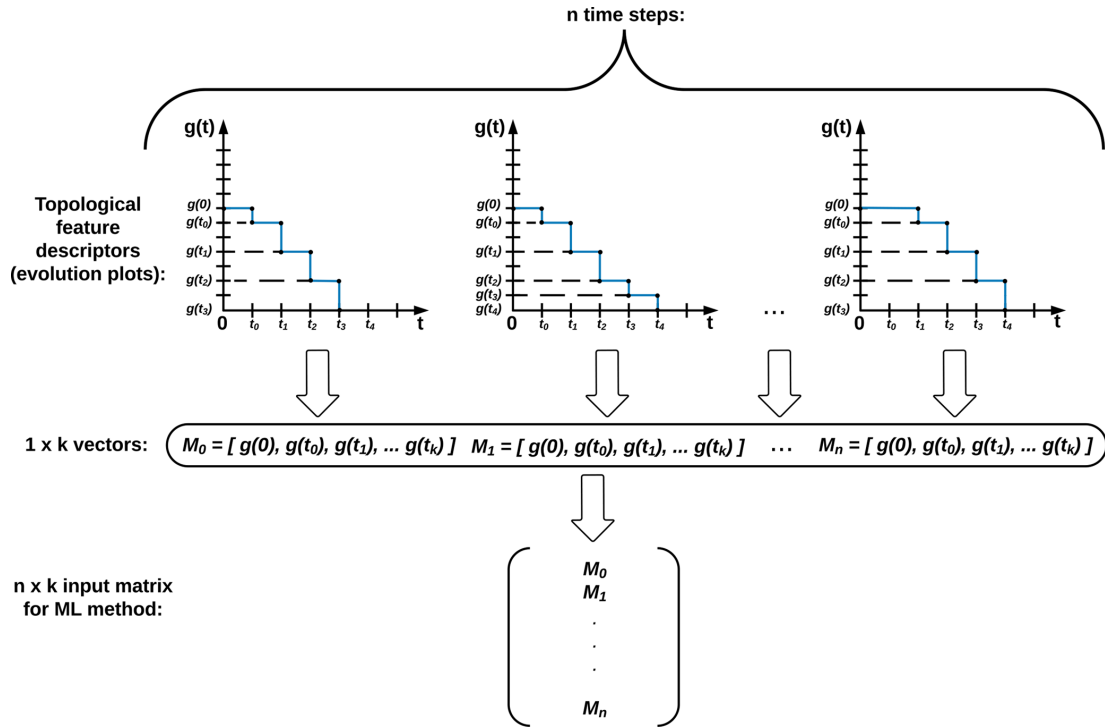


Figure 6. Creating an input matrix for the machine learning method: the mapped evolution plots into $1 \times k$ vectors (topological feature descriptors) are stacked on top of each other to construct $n \times k$ input matrix for the SVM classifier. The matrix is used by the SVM (see Sect. 2.2.2) along with labels provided by TECA (see Sect. 2.1) (Prabhat et al., 2015).

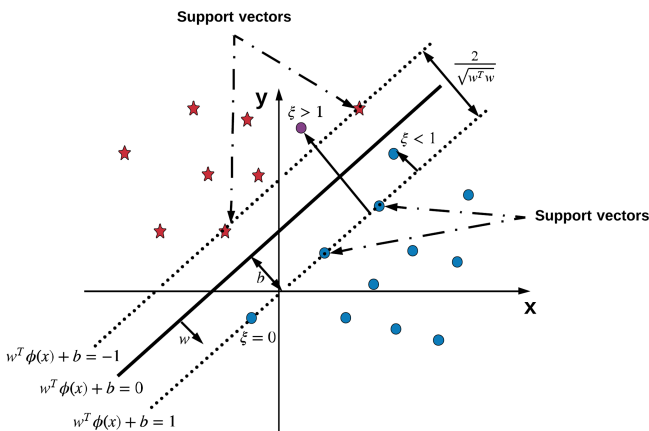


Figure 7. An example of linear SVM that finds the optimal hyperplane $w^T \phi(x) + b = 0$, its maximum-margin $\frac{2}{\sqrt{w^T w}}$ separating samples from two classes in data (blue dots and red stars), and all other quantities in Eqs. (3)–(4). ξ is a variable defining how much on the “wrong” side of the hyperplane a sample is: if it is $1 > \xi > 0$, the point is classified correctly but by less of a margin than the optimal hyperplane was found; otherwise, if it is more than $\xi > 1$, the point is classified incorrectly. The magenta dot indicates an example of the misclassified sample from the class of blue dots. Support vectors help to find the margin for the optimal linear hyperplane. $\phi(x)$ is a linear transformation in this case.

ratio of true positives to the sum of true positives and false negatives. It is shown in Table 7 for the SVM classifier.

Normalizing and balancing the data

Data normalization (standardization) is a way of adjusting measured values to a common scale (i.e., $[0, 1]$) by dividing through the largest maximum value in each feature (column of the matrix). This standardization allows for the comparison of corresponding normalized topological feature descriptors of different datasets. Also, standardization is a common requirement for many machine learning classifiers to avoid influence of outliers in the training process.

Balancing the data is motivated by the imbalanced class problem, which is that each class of event (ARs and non-ARs) is not equally represented in the dataset. This poses a problem because SVMs tend to overfit the majority class. We circumvent this problem by resampling (Lemaître et al., 2017). Resampling has been applied to all matrices created by the topological data analysis algorithm along with TECA labels.

3 Results and discussion

This section presents results from applying the proposed AR recognition method on test datasets. The tests have been done

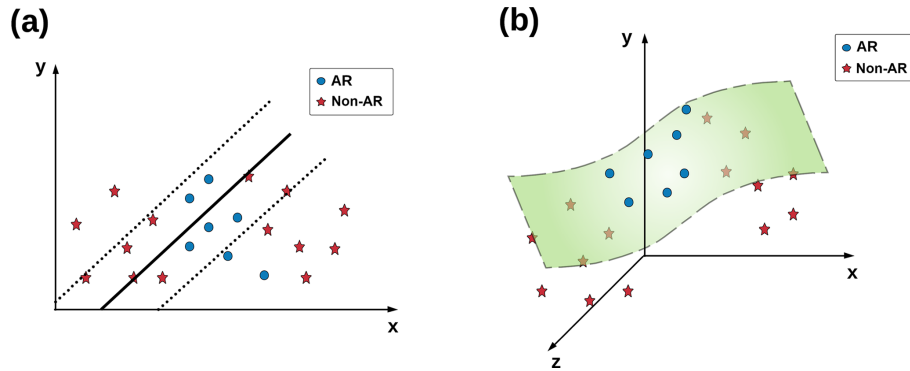


Figure 8. (a) An example of no clear linear separation between two classes (e.g., ARs and non-ARs) in data. This case cannot be solved using linear SVM. (b) In a situation where the set of two class samples is not linearly separable in the original space, the SVM introduces the notion of a “kernel-function-induced feature space” which casts the data into a higher-dimensional space where the data are separable.

on CAM5.1 simulation output and the MERRA-2 reanalysis product. A summary of the data and their spatial and temporal resolutions is in Table 1. First, we compare the topological feature descriptors of ARs and non-ARs based on the ground truth labeling provided by TECA (see Sect. 2.1). The descriptors have been normalized (see Sect. 2.3) to make the comparison of results to different datasets feasible. Second, we demonstrate performance and reliability of our method in the context of classification accuracy score obtained by the SVM classifier. Finally, we discuss some limitations of the method, its typical failure modes (using the confusion matrix) and its precision and sensitivity in recognizing ARs.

3.1 Topological feature descriptors representation

TDA provides a unique way of characterizing weather events in a dataset. Figure 9 shows an example of an evolution plot with two curves of averaged topological feature descriptors. The green and magenta curves correspond to ARs and non-ARs based on the TECA labels, respectively. Each curve represents the number of grid points in the connected region measured by the TDA algorithm. Note that the TDA algorithm records the evolution of the connected region as a function of the scalar variable (here TMQ). We observe that these two curves are close to each other; hence, visually distinguishing these two groups of climate events is a challenging task. However, one can train a machine learning model, such as a SVM, to perform this task with high accuracy.

Figures 10 and 11 show the evolution plots of topological feature descriptors obtained for both AR and non-AR events. Each plot consists of 100 randomly chosen examples of both categories of events from the raw 2-D snapshots from both the CAM5.1 climate model and the MERRA-2 reanalysis product. As in Fig. 9, we observe that it is hard to differentiate by eye the topological feature descriptor curves for ARs versus non-ARs. Yet the trained SVM can distinguish between AR and non-AR with fairly high accuracy by learning a suitable transformation of the feature descriptor curves

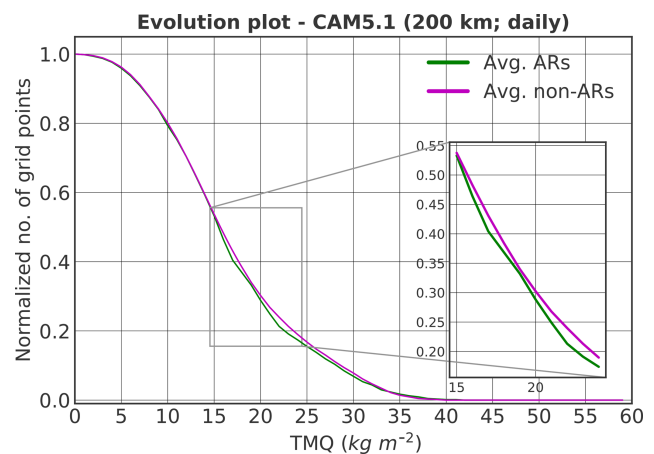


Figure 9. An example of normalized plots of averaged topological feature descriptors for 200 km spatial resolution and daily temporal resolution of the CAM5.1 simulation data. Note that the averaged plots for the ARs and non-ARs are very similar and it is hard to differentiate them by eye. However, a ML model can be trained to distinguish these two categories of events, by transforming the data into a high-dimensional space where a unique hyperplane exists that cleanly separates the two event categories (see Sect. 7).

into some high-dimensional space, where there exists a clean separation of the AR and non-AR groups with a suitable hyperplane (as shown in Fig. 7). This is typical in image recognition tasks; i.e., features that are difficult to distinguish by the human eye can be learned by a suitable ML method in order to perform the classification task with high accuracy.

The same analyses using topological feature descriptors have been done for all other datasets listed in Table 1; i.e., similar evolution plots have been prepared for daily temporal resolution and three different spatial resolutions of CAM5.1. We note that the curves look similar; hence, we only show one set of cases, and the others can be found in Appendix A.

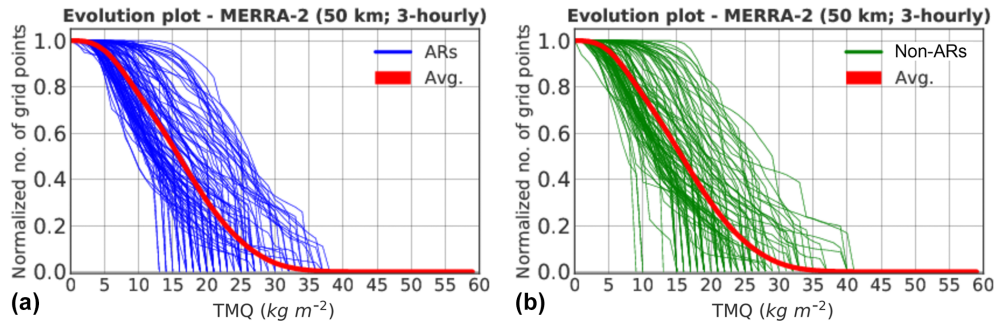


Figure 10. Normalized evolution plots of averaged (red curves) and 100 arbitrarily selected topological feature descriptors of ARs (blue curves; **a**) and non-ARs (green curves; **b**) for 3-hourly temporal resolution and 50 km spatial resolution of the MERRA-2 reanalysis data. The plots illustrate how topological descriptors vary versus IWV. They show the topological representation of both AR and non-AR events for MERRA-2 reanalysis data.

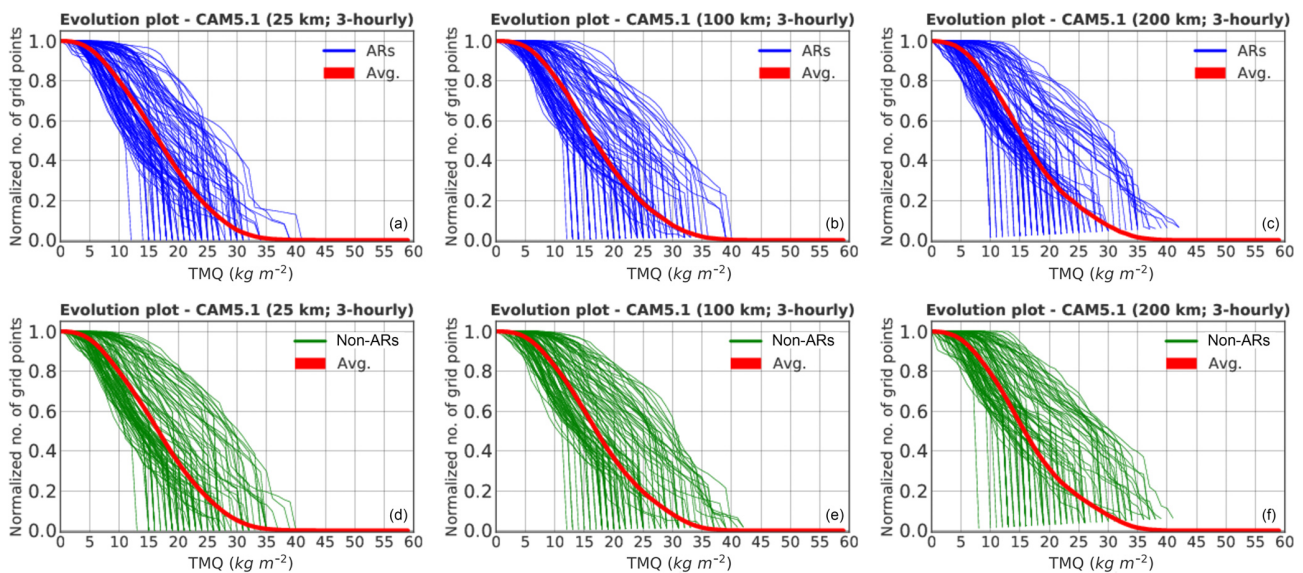


Figure 11. Normalized evolution plots of averaged (red curves) and 100 arbitrarily selected topological feature descriptors of ARs (blue curves; **a**, **b**, **c**) and non-ARs (green curves; **d**, **e**, **f**) for 3-hourly temporal resolution and 25 km (**a**, **d**), 100 km (**b**, **e**) and 200 km (**c**, **f**) spatial resolutions of the CAM5.1 simulation data. The plots illustrate how topological descriptors vary versus IWV. They show the topological representation of both AR and non-AR events for each resolution of CAM5.1 model output.

3.2 Classifier performance

We now evaluate the performance and reliability of the proposed AR recognition method by measuring the classification accuracy (as defined in Sect. 2). Tables 3, 4 and 5 summarize the classification accuracy of our method for the CAM5.1 climate model at different horizontal resolutions as well as for the MERRA-2 reanalysis product.

Training accuracy measures how well the model learns from training data (25 % of dataset), i.e., ground truth data labeled with ARs and non-ARs. Testing accuracy measures how well the method performs on a “held-out” dataset (75 % of dataset).

Table 3 shows that the SVM classifier is able to learn to better differentiate ARs from non-ARs when the spatial

resolution of the climate model is lower. We speculate that the reason for this is that despite the fact that the higher-resolution version of the model more realistically represents AR statistics (Wehner et al., 2014), the IWV fields tend to be noisier, leading to a less smooth topological representation and lower training accuracy. Further, despite a lower number of ARs to train on or classify due to resolution effects, the fairly high testing classification accuracy for the CAM5.1 (200 km) suggests that the SVM is able to capture key non-linear dependencies between topological feature descriptors.

In Table 4, we observe a similar trend with classification accuracy and model resolution as in Table 3. Also note that the number of snapshots is about 10 times smaller, but this does not affect testing accuracies (consistently above 80 %). This suggests that even though event boundaries may

Table 3. Classification accuracy score estimated by the SVM classifier for 3-hourly temporal resolution of the CAM5.1 model with three different spatial resolutions. The table also shows the number of snapshots (number of events for both categories: ARs and non-ARs).

Dataset	Training accuracy	Testing accuracy	No. of AR snapshots	No. of non-AR snapshots
CAM5.1 (25 km)	83 %	83 %	6838	6848
CAM5.1 (100 km)	77 %	77 %	7182	7581
CAM5.1 (200 km)	90 %	90 %	3914	3914

Table 4. Classification accuracy score estimated by the SVM classifier for daily temporal resolution of the CAM5.1 model with three different spatial resolutions. The table also shows the number of snapshots (number of events for both categories: ARs and non-ARs).

Dataset	Training accuracy	Testing accuracy	No. of AR snapshots	No. of non-AR snapshots
CAM5.1 (25 km)	78 %	82 %	624	624
CAM5.1 (100 km)	85 %	84 %	700	700
CAM5.1 (200 km)	89 %	91 %	397	397

Table 5. Classification accuracy score estimated by the SVM classifier for 3-hourly temporal resolution and 50 km spatial resolution of the MERRA-2 reanalysis. The table also shows the number of snapshots (number of events for both categories: ARs and non-ARs).

Dataset	Training accuracy	Testing accuracy	No. of AR snapshots	No. of non-AR snapshots
MERRA-2 (50 km)	80 %	80 %	13 294	13 434

Table 6. Confusion matrix of the method for testing set – the CAM5.1 data (3-hourly, 25 km), which show the number of correctly classified (diagonal) and incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	5047	391
Predicted AR	1432	4078

be more smeared out in daily averages, the topological feature descriptors encode sufficiently unique information about ARs and non-ARs that SVM is able to distinguish between the two categories with high accuracy. Finally, the SVM has highest training and testing accuracy for CAM5.1 (200 km), as in Table 3.

Table 5 reports the classification accuracy of SVM for MERRA-2 reanalysis product. Note that classification accuracies are about the same as for 3-hourly datasets from the CAM5.1 model. Hence, we conclude that the SVM classification method is robust to the source of maps of IWV.

In summary, the model has consistently high classification accuracy for ARs (77 %–91 %) across a broad set of spatial and temporal resolutions, illustrating that the combination of topological data analysis and machine learning is an effective and efficient threshold-free strategy for detecting ARs in large climate datasets. We note that the ML model is bi-

Table 7. Precision and sensitivity scores (described in Sect. 2) calculated for all datasets listed in Table 1. Both scores show the ability of the SVM classifier in assigning correct labels to the test instances.

Dataset	Precision	Sensitivity
CAM5.1 (25 km, 3-hourly)	0.91	0.74
CAM5.1 (100 km, 3-hourly)	0.83	0.67
CAM5.1 (200 km, 3-hourly)	0.95	0.85
CAM5.1 (25 km, daily)	0.87	0.77
CAM5.1 (100 km, daily)	0.86	0.83
CAM5.1 (200 km, daily)	0.97	0.85
MERRA-2 (25 km, 3-hourly)	0.84	0.74

ased by the ground truth data produced by TECA using the threshold-based criteria for AR identification. Characterizing the influence of using different ground truth data is beyond the scope of this study.

3.3 Limitations of our method

In this section, we examine some limitations of the proposed method. We investigate some typical failure modes further by examining snapshots of misclassified events. Then, we use the confusion matrix along with precision and sensitivity scores to quantify how accurately and precisely the model

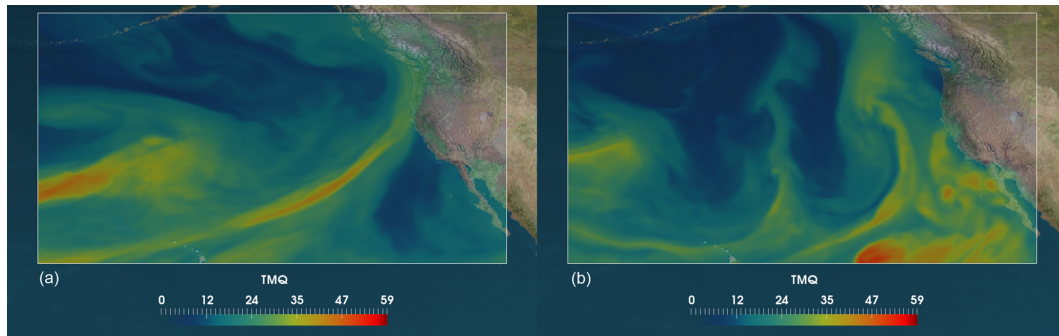


Figure 12. Sample images of events from the testing set showing a typical failure mode of the proposed method: examples of ARs misclassified as non-ARs (false negatives). The figure shows IWB (kg m^{-2}) in the CAM5.1 climate model (color map) and the land–sea mass as background (satellite image). (a) The model fails likely because there are two separate events in the figure; one is a fully formed AR and the other is the start of a new AR. (b) The method fails likely due to imperfect training data. The ground truth data from TECA label this image as an AR, although (visually) it does not appear to satisfy the definition of an AR. This illustrates how imperfect training data, due to limitations of the algorithm used to produce ground truth data, impact the performance of the ML model.

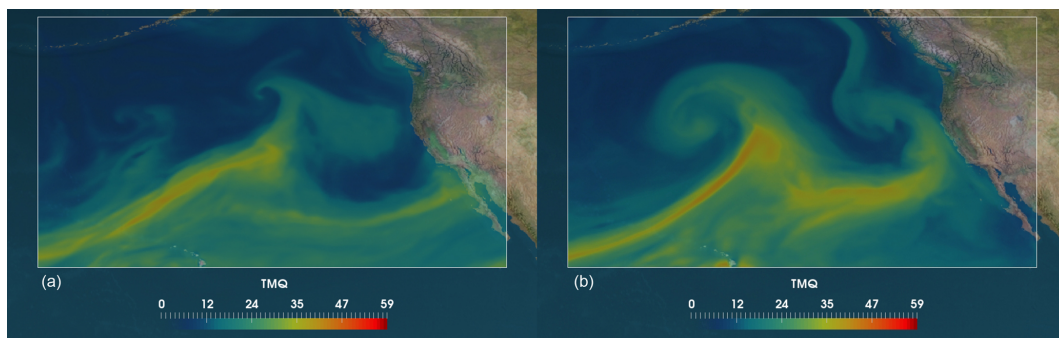


Figure 13. Sample images of events from the testing set showing a typical failure mode of the proposed method: examples of non-ARs misclassified as ARs (false positives). The figure shows IWB (kg m^{-2}) in the CAM5.1 climate model (color map) and the land–sea mass as background (satellite image). (a) The model likely fails because of the presence of two AR-like branches that are close to each other: one that has not yet made landfall and another that probably remained after the previous event. (b) The model fails likely due to the merging of two events, both with high concentrations of water vapor: one that appears to be an AR and the other likely an extratropical cyclone (ETC).

is able to classify events by comparing against ground truth data.

Figure 12 shows a typical failure mode of the proposed method: examples of AR misclassified as non-ARs, i.e., false negatives. We note that imperfect training data are a challenge in ML and high-quality ground truth data are essential for good model performance. However, in some cases, the process of feature abstraction that occurs while training the ML model may indeed produce a model that could outperform the algorithms used for producing the original ground truth training data.

Figure 13 shows another typical failure mode of the proposed method: examples of non-ARs misclassified as ARs, i.e., false positives. This failure mode is often related to the merging of multiple events, either of two ARs or of an AR (Fig. 13a) and some other event with high concentration of water vapor and similar topological structure, such as an extratropical cyclone (ETC) (Fig. 13b). We use the confusion matrix (described in Sect. 2) to give more insight into the

classification accuracy of the method and to quantitatively compare the types of correct and incorrect predictions made, as shown in Table 6 for CAM5.1 model output at 25 km spatial resolution and 3-hourly temporal resolution. Note that the model performs very well in classifying AR events correctly but has relatively poorer performance for non-AR events.

In Appendix B, we present confusion matrices of the method for different spatial and temporal resolutions of the CAM5.1 model and MERRA-2 reanalysis product.

Table 7 shows that the method has the highest precision and sensitivity scores for 200 km resolution of the CAM5.1 model for both 3-hourly and daily temporal resolutions. The scores are slightly lower for other spatial and temporal resolutions of CAM5.1 and reanalysis data.

4 Conclusions and future work

In this paper, we propose a novel and automated method for recognizing AR patterns in large climate datasets. The method combines TDA with ML, both of which are powerful tools that the climate science community often does not use.

We show that the proposed method is reliable, robust and performs well by testing it on a wide range of spatial and temporal resolutions of CAM5.1 climate model output as well as the MERRA-2 reanalysis product. The ground truth labels are obtained using TECA (Prabhat et al., 2015). The performance of the method is quantified by its classification accuracy in recognizing AR events, and precision and sensitivity scores.

Despite background noise, low-intensity AR signals and the existence of other events within the 2-D snapshots, our method is shown to work well. The method tends to perform better for lower-resolution data and we speculate that this is because high-resolution simulations tend to produce noisier spatial patterns, which tend to confuse the machine learning model more easily than low-resolution simulations.

The key advantage of the topological feature descriptors used in this work is that it is a threshold-free method that succinctly encapsulates the most important topological features of ARs. We anticipate that because the method is threshold-free (there is no need to determine any threshold criteria for the TDA step), when the spatial resolution of the climate model changes, there is no parameter retuning, unlike in the case of heuristic methods used by most other AR-detection methods. An application of this method to different climate change scenarios without any tuning will be explored in future work.

Further, it is a much faster method than, for example, using convolutional neural networks (Liu et al., 2016) (processing time of a couple of minutes versus a few days).

In future work, we will test our method on direct observations via satellite images. We also plan to test the proposed method in different climate scenarios, in order to test the method's sensitivity to biases in the training data. Further, we anticipate that the method can be made more robust by (i) employing a full "persistence" concept from TDA and (ii) training SVM on ground truth data that are not biased by fixed threshold criteria. This study shows that the TDA and ML framework could be an effective way to characterize and identify a wide range of other weather and climate phenomena, such as blocking events and jet streams. As the TDA step is not restricted to a 2-D scalar field on a grid, it is also possible to apply to higher-dimensional or multivariate fields. A similar TDA-based approach has successfully been applied to data skeletonization (Kurlin, 2015) and segmentation (Kurlin, 2016) problems. Hence, we believe that this method can be extended to be applied in a variety of other climate science problems where defining suitable thresholds remains a challenge.

Code and data availability. Source code is available at GitHub: <https://github.com/muszyna25/AR-Detection-Method-TDA-ML>. git (Muszynski et al., 2018a). Data are available at NERSC Science Gateway: https://doi.org/10.25342/GMD_2018 (Muszynski et al., 2018b).

Appendix A: Additional evolution plots for daily temporal resolution of CAM5.1 climate model output

This Appendix contains additional evolution plots mentioned in Sect. 3.1.

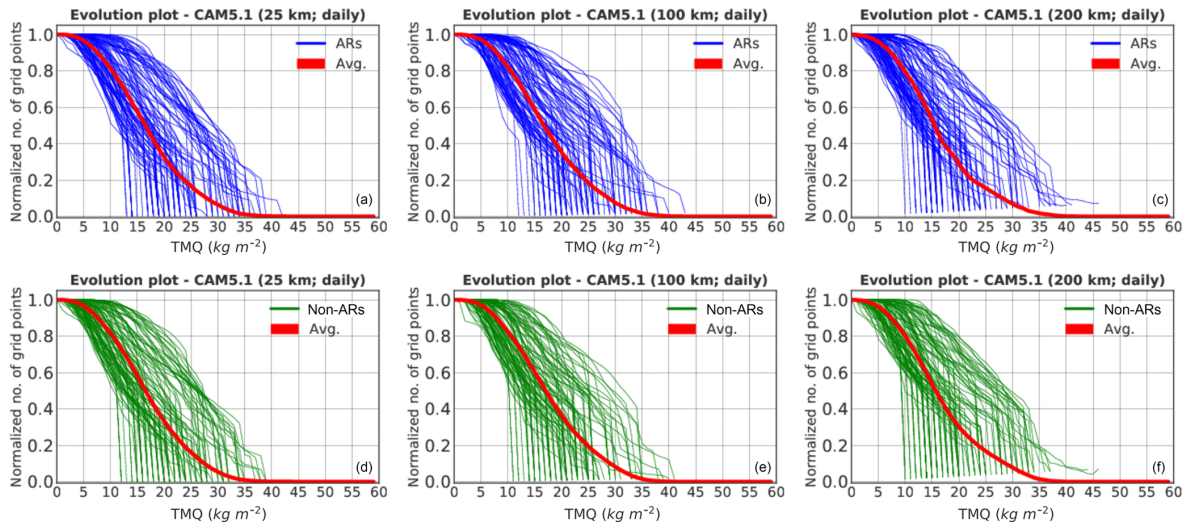


Figure A1. Normalized evolution plots of averaged (red curves) and 100 arbitrarily selected topological feature descriptors of ARs (blue curves; **a**, **b**, **c**) and non-ARs (green curves; **d**, **e**, **f**) for daily temporal resolution and 25 km (**a**, **d**), 100 km (**b**, **e**) and 200 km (**c**, **f**) spatial resolutions of the CAM5.1 simulation data. The plots illustrate the relationship of topological descriptor variation versus average. They show the topological representation of both AR and non-AR events for each resolution of CAM5.1 model output.

Appendix B: Additional confusion matrices for CAM5.1 and MERRA-2 testing sets

This Appendix includes the rest of the confusion matrices (tables) that were considered in Sect. 3.3. The presented tables allow for quantitative comparison of ML classifier performance to recognize ARs in CAM5.1 climate model outputs and the MERRA-2 reanalysis product.

Table B1. Confusion matrix of the method for the testing set – the MERRA-2 data (3-hourly, 50 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	9211	1489
Predicted AR	2782	7900

Table B2. Confusion matrix of the method for the testing set – the CAM5.1 data (3-hourly, 100 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	5258	808
Predicted AR	1887	3857

Table B3. Confusion matrix of the method for the testing set – the CAM5.1 data (3-hourly, 200 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	3020	137
Predicted AR	466	2639

Table B4. Confusion matrix of the method for the testing set – the CAM5.1 data (daily, 25 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	444	59
Predicted AR	116	379

Table B5. Confusion matrix of the method for the testing set – the CAM5.1 data (daily, 100 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	486	77
Predicted AR	97	460

Table B6. Confusion matrix of the method for the testing set – the CAM5.1 data (daily, 200 km). It shows the number of correctly recognized (the diagonal) and the number of incorrectly classified events (off-diagonal).

	Label non-AR	Label AR
Predicted non-AR	306	8
Predicted AR	48	273

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-12-613-2019-supplement>.

Author contributions. GM conceived the method, performed the computations, analyzed the data and wrote the manuscript. KK coordinated the work and assisted in the development of the overall content included in this article. MW assisted in the development of the overall content included in this article and contributed to the interpretation of the results. VK assisted in the development of the overall content included in this article and shared his expertise in topological data analysis. P supervised the work, assisted in the development of the overall content included in this article and shared his expertise in machine learning.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. This document was prepared as an account of work partially sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Acknowledgements. Grzegorz Muszynski and Vitaliy Kurlin would like to acknowledge Intel for supporting the IPCC at the University of Liverpool. We also thank Dmitriy Morozov and Burlen Loring from the Computational Research Division at Lawrence Berkeley National Laboratory for valuable discussions and sharing their expertise on computational mathematics.

Karthik Kashinath was supported by the Intel Big Data Center, and Michael Wehner was supported by the Regional and Global Climate Modeling Program of the Office of Biological and Environmental Research in the Department of Energy Office of Science under contract no. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

Edited by: James Annan

Reviewed by: Soulivanh Thao and one anonymous referee

References

- AMS: Atmospheric River, Glossary of Meteorology, available at: <http://glossary.ametsoc.org/wiki/Atmosphericriver>, last access: 10 January 2018.
- Burges, C. J.: A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Disc.*, 2, 121–167, 1998.
- Carlsson, G.: Topology and data, *B. Am. Math. Soc.*, 46, 255–308, 2009.
- Carlsson, G.: Topological pattern recognition for point cloud data, *Acta Numer.*, 23, 289–368, 2014.
- Chang, C.-C. and Lin, C.-J.: LIBSVM: a library for support vector machines, *ACM T. Intel. Syst. Tec.*, 2, 27, <https://doi.org/10.1145/1961189.1961199>, 2011.
- Cortes, C. and Vapnik, V.: Support-vector networks, *Mach. Learn.*, 20, 273–297, 1995.
- Dettinger, M. D.: Atmospheric rivers as drought busters on the US West Coast, *J. Hydrometeorol.*, 14, 1721–1732, 2013.
- Dettinger, M. D. and Ingram, B. L.: The coming megafloods, *Sci. Am.*, 308, 64–71, 2013.
- Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., and Cayan, D. R.: Atmospheric rivers, floods and the water resources of California, *Water*, 3, 445–478, 2011.
- Eaton, B.: User's Guide to the Community Atmosphere Model CAM-5.1, NCAR, 2011.
- Edelsbrunner, H. and Harer, J.: Computational topology: an introduction, American Mathematical Soc., ISBN-10 0-8218-4925-5, ISBN-13 978-0-8218-4925-5, 2010.
- Edelsbrunner, H. and Morozov, D.: Persistent Homology: Theory and Practice, Tech. rep., Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), 2012.
- Fragoso, M., Trigo, R. M., Pinto, J. G., Lopes, S., Lopes, A., Ulbrich, S., and Magro, C.: The 20 February 2010 Madeira flash-floods: synoptic analysis and extreme rainfall assessment, *Nat. Hazards Earth Syst. Sci.*, 12, 715–730, <https://doi.org/10.5194/nhess-12-715-2012>, 2012.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Parityka, G., Pawson, S., Putman, S., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *J. Climate*, 30, 5419–5454, 2017.
- Ghrist, R.: Barcodes: the persistent topology of data, *B. Am. Math. Soc.*, 45, 61–75, 2008.
- Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., and Neiman, P. J.: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements, *Geophys. Res. Lett.*, 37, L20401, <https://doi.org/10.1029/2010GL044696>, 2010.
- He, H. and Garcia, E. A.: Learning from imbalanced data, *IEEE T. Knowl. Data En.*, 21, 1263–1284, 2009.
- Hopcroft, J. E. and Ullman, J. D.: Set merging algorithms, *SIAM J. Comput.*, 2, 294–303, 1973.
- Hsu, C.-W., Chang, C.-C., and Lin, C.-J.: A practical guide to support vector classification, 2003.
- Kubat, M.: An Introduction to Machine Learning, Springer, 2015.

- Kurlin, V.: A one-dimensional Homologically Persistent Skeleton of a point cloud in any metric space, *Comput. Graph. Forum*, 34, 253–262, 2015.
- Kurlin, V.: A fast persistence-based segmentation of noisy 2D clouds with provable guarantees, *Pattern Recogn. Lett.*, 83, 3–12, 2016.
- Lavers, D. A. and Villarini, G.: The nexus between atmospheric rivers and extreme precipitation across Europe, *Geophys. Res. Lett.*, 40, 3259–3264, 2013.
- Lemaître, G., Nogueira, F., and Aridas, C. K.: Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning, *J. Mach. Learn. Res.*, 18, 1–5, 2017.
- Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., and Collins, W. D.: Application of deep convolutional neural networks for detecting extreme weather in climate datasets, arXiv preprint arXiv:1605.01156, 2016
- Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., and Prabhat: TDA and ML for AR pattern recognition, GitHub, available at: <https://github.com/muszyna25/AR-Detection-Method-TDA-ML>, last access: 25 November 2018a.
- Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., and Prabhat, M.: NERSC Science Gateway, https://doi.org/10.25342/GMD_2018, 2018b.
- Newell, R. E. and Zhu, Y.: Tropospheric rivers: A one-year record and a possible application to ice core data, *Geophys. Res. Lett.*, 21, 113–116, 1994.
- Newell, R. E., Newell, N. E., Zhu, Y., and Scott, C.: Tropospheric rivers?—A pilot study, *Geophys. Res. Lett.*, 19, 2401–2404, 1992.
- Prabhat, Byna, S., Vishwanath, V., Dart, E., Wehner, M., and Collins, W. D.: TECA: Petascale pattern recognition for climate science, in: *International Conference on Computer Analysis of Images and Patterns*, Springer, 426–436, 2015.
- Racah, E., Beckham, C., Maharaj, T., Ebrahimi Kahou, S., Prabhat, M., and Pal, C.: ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events, in: *Advances in Neural Information Processing Systems*, 3405–3416, 2017.
- Ralph, F. and Dettinger, M.: Storms, floods, and the science of atmospheric rivers, *Eos T. Am. Geophys. Un.*, 92, 265–266, 2011.
- Ralph, F. M., Prather, K. A., Cayan, D., Spackman, J. R., DeMott, P., Dettinger, M., Fairall, C., Leung, R., Rosenfeld, D., Rutledge, S., Waliser, D., White, A. B., Cordeira, J., Martin, A., Helly, J., and Intrieri, J.: CalWater field studies designed to quantify the roles of atmospheric rivers and aerosols in modulating US West Coast precipitation in a changing climate, *B. Am. Meteorol. Soc.*, 97, 1209–1228, 2016.
- Ralph, F. M., Neiman, P. J., and Wick, G. A.: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98, *Mon. Weather Rev.*, 132, 1721–1745, 2004.
- Sellars, S., Kawzenuk, B., Nguyen, P., Ralph, F., and Sorooshian, S.: Genesis, Pathways, and Terminations of Intense Global Water Vapor Transport in Association with Large-Scale Climate Patterns, *Geophys. Res. Lett.*, 44, 12–465, 2017.
- Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., Lora, J. M., McClenny, E., Osborne, T., Payne, A. E., Ullrich, P., Gershunov, A., Goldenson, N., Guan, B., Qian, Y., Ramos, A. M., Sarangi, C., Sellars, S., Gorodetskaya, I., Kashinath, K., Kurlin, V., Mahoney, K., Muszynski, G., Pierce, R., Subramanian, A. C., Tome, R., Waliser, D., Walton, D., Wick, G., Wilson, A., Lavers, D., Prabhat, Collow, A., Krishnan, H., Magnusdottir, G., and Nguyen, P.: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design, *Geosci. Model Dev.*, 11, 2455–2474, <https://doi.org/10.5194/gmd-11-2455-2018>, 2018.
- Tarjan, R. E.: Efficiency of a good but not linear set union algorithm, *J. ACM*, 22, 215–225, 1975.
- Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids, *Geosci. Model Dev.*, 10, 1069–1090, <https://doi.org/10.5194/gmd-10-1069-2017>, 2017.
- Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, Ch.-T., Paciorek, Ch., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., and Jablonowski, Ch.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, *J. Adv. Model. Earth Sy.*, 6, 980–997, 2014.
- Zhu, Y. and Newell, R. E.: Atmospheric rivers and bombs, *Geophys. Res. Lett.*, 21, 1999–2002, 1994.
- Zhu, Y. and Newell, R. E.: A proposed algorithm for moisture fluxes from atmospheric rivers, *Mon. Weather Rev.*, 126, 725–735, 1998.