Geophysical Research Abstracts
Vol. 20, EGU2018-10825, 2018
EGU General Assembly 2018

Topological Data Analysis and Machine Learning for Classifying Atmospheric River Patterns in Large Climate Datasets

Grzegorz Muszynski (1), Vitaliy Kurlin (2), Karthik Kashinath (3), Michael Wehner (4), and Mr Prabhat (5)

(1) University of Liverpool, Department of Computer Science, United Kingdom (muszyna23@gmail.com), (2) University of Liverpool, Department of Computer Science, United Kingdom, (3) Lawrence Berkeley National Laboratory, NERSC, United States, (4) Lawrence Berkeley National Laboratory, Computational Research Division, United States, (5) Lawrence Berkeley National Laboratory, NERSC, United States

We present novel approaches to classifying and characterizing extreme weather events, such as atmospheric rivers (ARs), in large high-dimensional climate datasets. ARs are narrow filaments of concentrated water vapour in the lower troposphere that bring much of the precipitation in many mid-latitude regions. Their presence has been connected to major flooding events along the west coast of the United States and western Europe.

In this study, we combine ideas from Topological Data Analysis (TDA) with Machine Learning (ML) for classifying and characterizing extreme weather events, like ARs. TDA is a new field that is on the interface between applied topology and computer science, that studies "shape" – hidden topological structure – in raw data. It has been applied successfully in many areas of applied sciences, including complex networks, signal processing, and image recognition.

Using TDA we provide ARs with a shape characteristic as a new feature descriptor for the task of AR classification. In particular, we track the change in topology in precipitable water (integrated water vapour) fields using the Union-Find algorithm. We use the generated feature descriptors with Support Vector Machines classifier to establish reliability and classification performance of our approach. We utilize the parallel toolkit for extreme climate events analysis (TECA: Petascale Pattern Recognition for Climate Science, Prabhat et al., Computer Analysis of Images and Patterns, 2015) for comparison (it is assumed that events identified by TECA are ground truth). Preliminary results indicate that our approach brings new insight into the study of ARs and provides quantitative information about the relevance of topological feature descriptors in analyses of large climate datasets. We illustrate this method on the CAM5.1 climate model output and MERRA-2 reanalysis datasets. Further, our method provides high accuracy in recognizing AR patterns for both different spatial and temporal resolutions. This work shows that TDA combined with ML can be an effective way to analyse extreme weather phenomena.