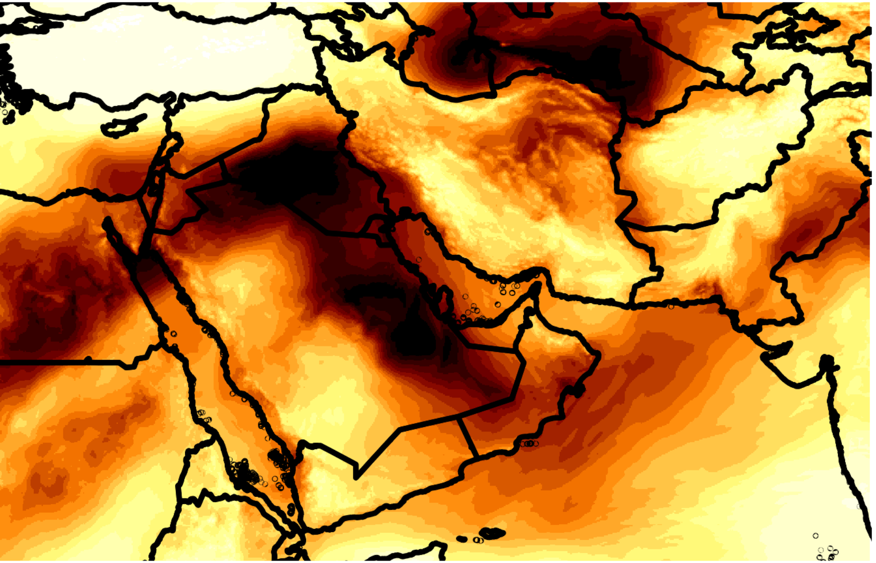
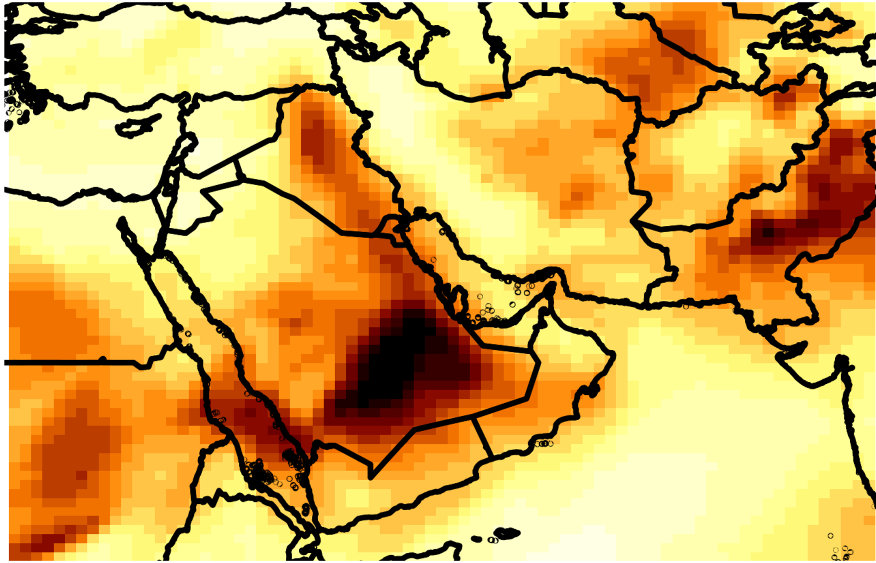
Nature Run



MERRA-2



Virtual Research Presentation Conference

Topological Data Analysis of Aerosol Optical Depth Observations

Principal Investigator: Huikyo Lee (398L)
Co-Is: Krzysztof Gorski (3268), Yulia Gel (University of Texas at Dallas)
Program: Spontaneous Concept

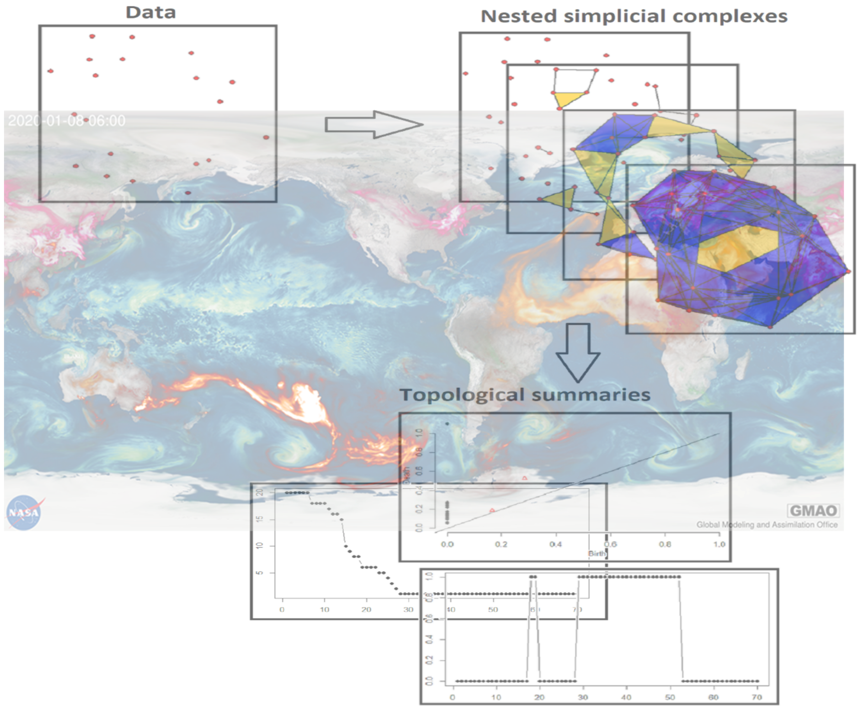
Assigned Presentation # RPC-038

Jet Propulsion Laboratory
California Institute of Technology

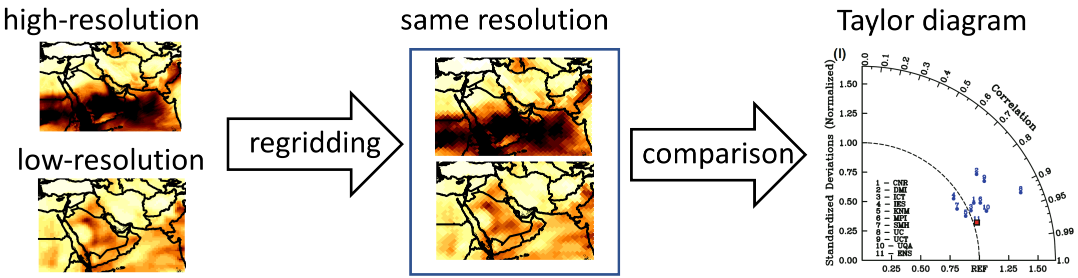
Tutorial Introduction

Abstract

Topological data analysis (TDA), a machinery that combines concepts from algebraic topology, machine learning, statistics and data science, allows to study datasets in terms of its latent shape. Despite TDA's popularity in statistical sciences and machine learning (ML), utility of geometric methods in assessing spatial characteristics of Earth Science datasets is yet untapped. For the first time, the current study aims at **developing novel topological approaches for comparing Earth science datasets at different resolutions.** We apply the machinery of TDA and, particularly, persistent homology (PH), to the aerosol optical depth (AOD) datasets to introduce a robust and reliable methodology based on systematic assessment of topology and geometry in AOD maps. In particular, we assessed two AOD datasets from the Goddard Earth Observing System, Version 5 (GEOS-5) model over the Middle East. Our application suggests that TDA can complement the conventional methodologies to compare spatial patterns from various observational and model datasets without regridding the datasets into common grids.



Problem Description



a) To the best of our knowledge, there is no quantitative metric to measure similarity/difference in spatial patterns between Earth science datasets at different spatial resolutions.

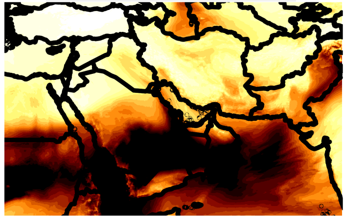
b) For example, the following three metrics have been widely used to quantitatively compare spatial patterns of climate variables of our interest: 1) a bias of a two-dimensional map against another map, 2) root mean square error (RMSE) between two maps, and 3) a pattern correlation coefficient. A Taylor diagram is also popular among climate scientists as the diagram displays both RMSE and pattern correlation coefficient thus can provides a concise summary of similarity in spatial patterns. The main challenge in using these conventional metrics or Taylor diagrams is that users need to regrid all of the datasets into common grid points prior to comparing them. The regridding process ignores differences in spatial resolutions for multiple datasets.

c) NASA centers including JPL has limited capability to conduct TDA with datasets from satellites and climate models. Our work demonstrates that TDA is applicable to comparing observational and model datasets at different spatial resolutions without remapping the datasets into common grids. Moreover, topological representations of Earth science datasets can be utilized in machine learning.

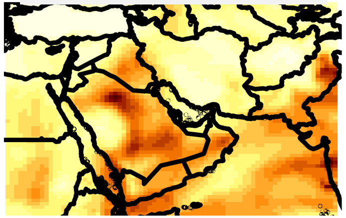
Methodology

- Using persistence homology (PH), we have tested and built a methodology that compares two aerosol optical depth (AOD) datasets with two different resolutions (G5NR: 7 km, MERRA-2: 0.625° x 0.5)

G5NR AOD map

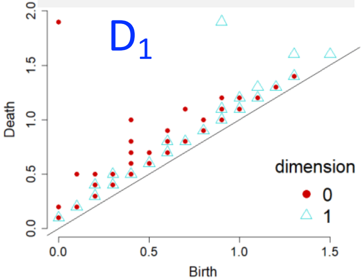


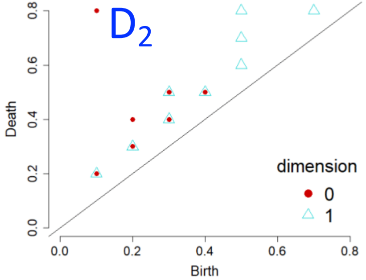
MERRA-2 AOD map



Capture the topological features of AOD maps based on the formation of cubes (cubical complex).

Persistence Diagrams



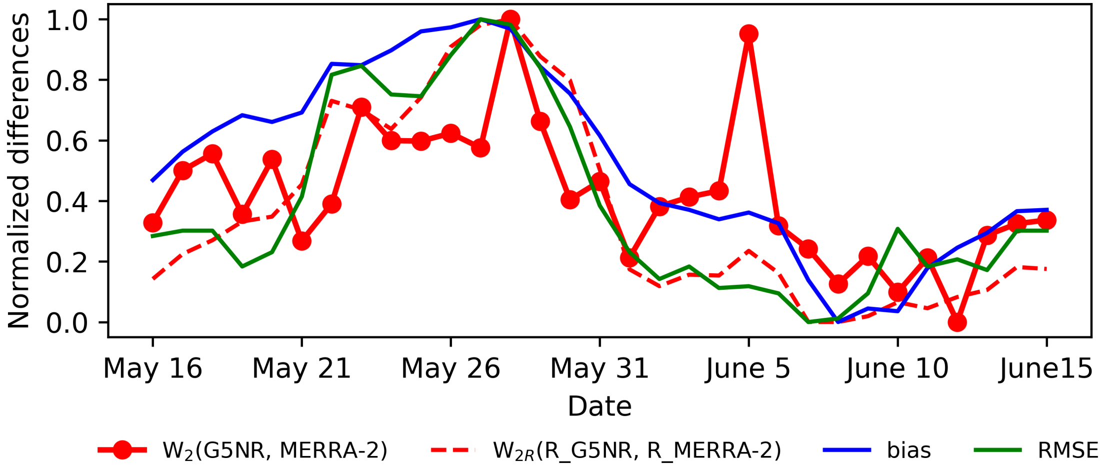


Wasserstein distance between D_1 and D_2

$$W_r(D_1, D_2) = \left(\inf_{\gamma} \sum_{x \in D_1} ||x - \gamma(x)||_{\infty}^{r} \right)^{1/r}$$

In terms of latent topology, two AOD maps from G5NR and MERRA-2 can be compared by calculating Wasserstein distance between their respective persistence diagrams (D_1 and D_2).

Topological difference (W_2) vs. conventional metrics (biases/RMSEs)



a) Although we have not applied TDA to satellite observations, Wasserstein distance (W_2) between persistent diagrams represents topological difference between different AOD maps.

b) While W_2 is consistent with the bias and RMSE, W_2 is sensitive to the difference in key spatial features even when both bias and RMSE are low.

c) We will develop a machine learning-based framework to evaluate topological features in NASA's satellite and suborbital observations.

Publications and References

Ofori-Boateng, D., H. Lee, K. M. Gorski, and Y. R. Gel (2020): Multi-resolution data matching through lenses of topological data analysis (manuscript in preparation)