Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-53-AC1, 2018


Geoscientific Model Development Discussions

# *Interactive comment on* "Topological Data Analysis and Machine Learning for Recognizing Atmospheric River Patterns in Large Climate Datasets" *by* Grzegorz Muszynski et al.

**Grzegorz Muszynski et al.**

muszyna23@gmail.com

Received and published: 22 August 2018

C: Since the method is so novel and different with respect to relatively large number of method available now in literature (e.g., Shields et al 2018), it would be nice to see some figures and analysis of the frequency of ARs that make landfall on the west coast of North America according to this method. These extra figures may constitute a reference point to compare the occurrence of AR between this new and novel method and previous ones in other studies, especially those that use IWV field to detect ARs (Neiman et al 2008, Dettinger et al 2011, Wick et al 2013). A comparison with previous climatology is expected, as well as between the different gridded datasets used here



and the possible cause of differences.

A: We thank the reviewer for this suggestion. While it would indeed be nice to some figures of the statistics of ARs on the west coast of North America using our novel approach, a part of this analysis has been done for MERRA-2 data for the Atmospheric River Intercomparison Project (ARTMIP) (https://www.geosci-model-dev.net/11/2455/2018/). Further analysis of the statistics of ARs using this method is underway and in preparation for a subsequent paper with ARTMIP. Furthermore, we are currently repeating this analysis for CAM5.1 data also for ARTMIP. We intend to publish our results on the frequency analyses and other relevant statistics for the next ARTMIP paper, which is in preparation. This paper intends to describe in detail the novel method proposed and validate its usefulness in recognizing AR patterns.

C: It is not clear for me what labels are uses as "ground truth". It should be stated explicitly in the text despite that a related reference has been added.

A: We thank the reviewer for pointing this out. We included the definition of the "ground truth" on page 4, line 24 in the paper. The "ground truth" is a set of binary labels (AR: 1, Non-AR: 0) generated for all datasets listed in Table 1 by a heuristics-based detection method implemented in the Toolkit for Extreme Climate events Analysis (TECA) (Prabhat et al., 2015). TECA includes an AR detection method that uses geometrical constraints and a fixed threshold parameter on IWV based on threshold value provided in (Dettinger et al 2011). We are also aware that the ground truth data used for training comes from a method that uses thresholds, and hence there is the possibility of a bias, but this is a classic problem in machine learning and we plan to address the sensitivity to choice of ground truth in future work.

C: It is suggested in page 5 line 20-23 that the TDA approach works with scalar field. If so, this is the main reason to not use the vectoral IVT variable for identifying ARs, rather than IWV being observable by satellite. Please explain further about this point, especially because IVT variable is now being used more than IWV variable for detecting

ARs (e.g., see ARTMIP paper, Shields et al 2018).

A: We agree with the reviewer that the precise reason for using IWV should be clarified. While the TDA method implemented in this novel approach uses scalar fields, the method itself is applicable to vector fields (like IVT) and point clouds. Hence a scalar field is neither a requirement nor is it a restriction of the choice of TDA method used here. We plan to incorporate an extension of this method to vector fields (IVT) and point clouds in future work. As mentioned in the paper, we wish to emphasize that the choice of using IWV is because of the intention for this method to be usable across any type of dataset, including direct observational data, where IVT is very hard to measure directly. While we have not tested the method on direct observational data, it can be done without any extensions or modifications to the current method in its current form.

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-53, 2018.