Title: Identifying Atmospheric Rivers and their Poleward Latent Heat Transport with Generalizable Neural Networks: ARCNNv1

Authors: Ankur Mahesh, Travis A. O'Brien, Burlen Loring, Abdelrahman Elbashandy, William Boos, William D. Collins

Summary:

This work introduces various different CNN-based AR tracking methods with the goal of addressing the issue of AR detection uncertainty and flexibility of AR detection with different datasets and resolutions. An index composed of ARs detected by various other common methods is defined to validate the approach. A semi-supervised learning approach based on image-style transfer is applied. The result shows robust consistency with other common AR detection methods. The results also demonstrate that the ARCNNs also have consistent amounts of AR-induced latent heat transport with other common detection methods.

Overall Comments:

Mahesh et al. bring forward a highly effective method of addressing several existing challenges in AR-related research. The writing is clear and easy to follow. The authors provide easily accessible code and data to reproduce the results and apply them to other future studies. The results show high performance of the method. My largest concern about this work is the choice of ARTMIP methods used for validation. Out of seven methods used for validation (when calculating IoU), three of them are taken from the same group (Tempest). Choosing nearly half of detection methods in the validation set that are almost identical could cause the results to be misleading. There were also repeated ARTMIP methods used in the ARCI. Once that and several other comments are addressed, I recommend publication.

Specific Comments:

Line 66: I'm not convinced that different datasets would require new training labels for the purpose of detecting ARs. Re-gridding the training data could allow the user to have some flexibility with other datasets.

Line 134: You could justify the claim of strong performance with Wu et al. 2019

Wu, T., Tang, S., Zhang, R., & Zhang, Y. (2019). CGNet: A Light-weight Context Guided Network for Semantic Segmentation. *ArXiv:1811.08201 [Cs]*. Retrieved from http://arxiv.org/abs/1811.08201

Line 299: The language here ("its detected AR probabilities are too low") could be improved. Instead, I would suggest changing this to something along the lines of "its detected AR probabilities are lower than the ARCI".

Line 372-373: "CNNs have millions of tunable parameters" It could be useful to the reader to include a source for this claim

Line 631-632: Mundhenk is mentioned twice here. The first mention of Mundhenk does not include a reference so it is unclear if Mundhenk is being used twice, if there are two different versions used, or if this was a typo.

The ARDTs used for the AR Consensus Index include multiple algorithms from the same group (Lora, Mundhenk, CONNECT). While there are slight variations between different algorithms created by the same groups, some justification of the choice to weight algorithms from some groups more heavily than others in the ARCI could be useful.

Three different versions of Tempest are used to calculate IoU. Some justification for this could be useful.

It is unclear which version of Tempest is used in the ARCI

In Figure 8, it is unclear if the calculated IoU scores only representing grid points in which ARs are detected or is the background class IoU factored into the calculation as well.

I suggest referencing Higgins et al. 2023 to establish some precedent to using a variety of different ARTMIP labels to validate ARCNNs

Higgins, T. B., Subramanian, A. C., Graubner, A., Kapp-Schwoerer, L., Watson, P. A. G., Sparrow, S., et al. (2023). Using Deep Learning for an Analysis of Atmospheric Rivers in a High-Resolution Large Ensemble Climate Data Set. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003495. https://doi.org/10.1029/2022MS003495