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 paper is divided into two parts. First, Mahesh et al. describe a new machine learning methodology to identify atmospheric rivers (ARs) based on convolutional neural networks which utilizes a semi-supervised framework and image-style transfer learning. Uncertainty is quantified as well as the performance of the new methodology using a multi-pronged approach which includes observations, simplified models, and complex models. Second, latent heat transport attributable to ARs is quantified and demonstrated to have large uncertainty due to detection technique.

Overall Comments:

Mahesh et al. demonstrate a deep understanding of the problems and issues surrounding AR detection and the subsequent consequences on characterization of a physical process, which in this case, focus on latent heat transport. The paper provides thorough and robust quantification of uncertainty by leveraging many different datasets such as the ARTMIP database, reanalysis products, satellite data, and both climate and idealized simulations. Additionally, they provide details on computational resources, code, and datasets, all of which is necessary for reproducibility and is appreciated. I find the results presented here quite convincing and robustly vetted. Their conclusions on the spread of uncertainty due to detectors for latent heat transport is an important contribution and nicely updates and improves upon the current state of the literature on the topic. I recommend publication after a few minor comments and questions are answered. It was a pleasure to read and review.

Specific Comments:

Line 21: I think this statement has been demonstrated by all the main ARTMIP papers (Rutz et al, 2019, Collow et al., 2021, Shields, Payne et al., 2023).

Line 32: Shields, Payne et al. 2023 should also be added to the list for climate change ARDT comparisons.

Shields, C. A., Payne, A. E., Shearer, E. J., Wehner, M. F., O'Brien, T. A., Rutz, J. J., Leung, L.R., Ralph, F. M., Collow, A. B. M., Ullrich, P. A. Ullrich, Dong, Q., Gershunov, A., Griffith, H., Guan, B., Lora, J. M., Lu, M., McClenny, E., Nardi, K. M., Pan, M., Qian, Y., Ramos, A. M. Ramos, Shulgina, T., Viale, M., Sarangi, C., Tomé, R., Zarzycki, C. (2023). Future atmospheric rivers and impacts on precipitation: Overview of the ARTMIP Tier 2 high-resolution global warming experiment. Geophysical Research Letters, 50, e2022GL102091. https://doi.org/10.1029/2022GL102091

Line 95: Are there biases in the models that would impact AR detection? IWV, IVT? If so, a sentence or two addressing these would be helpful. Why not ERA5?

Line 105: Rutz et al. 2019 should be included in the ARTMIP list.

Figure 3: Confidence index (plot y label), or Consensus index (Figure caption)?

Section 2.1: I really like the ARCI and think its application in your paper is appropriate given you are looking at heat transport via mid-latitude ARs. However, one limitation I see is for regions such as the poles, where the majority of ARDTs actually don't capture ARs reaching either into the Arctic, or on the Antarctic continent correctly compared to ARDTs designed for high latitudes (Shields et al., 2022). The ARCI might not be that useful here because many of the globals (with no polar constraints) are not "fit for purpose". I'd recommend a qualifying statement on the use of ARCI for middle latitudes versus polar regions.

Shields, C. A., Wille, J. D., Marquardt Collow, A. B., Maclennan, M., & Gorodetskaya, I. V. (2022). Evaluating uncertainty and modes of variability for Antarctic atmospheric rivers. Geophysical Research Letters, 49, e2022GL099577 . https://doi.org/10.1029/2022GL099577.

Line 134: I know you include this in the discussion, but I think a line about how this is different from ClimateNet is needed here as well. My guess is there will be readers that are undoubtedly familiar with ClimateNet, given your use of the same underlying CAM5 data and DeepLabv3+ code? (i.e., your use of ARCI vs ClimateNet's hand drawn labels, or perhaps I am misunderstanding something)?

Line 280: For my clarification: Is the reverse also true? I.e., if the neural network is trained on *model* data, then applied to *reanalysis*, the same problem would exist? Isn't this what ClimateNet's ARTMIP contribution does? Maybe you don't have access to that answer, but if this is the case, would the ClimateNet ARTMIP catalogues have these same problems? Is this concerning that they are included in your ARCI?

Lines 451-454: For 11b, the 80% consensus line is actually bigger in the NH. Maybe move your explanation of this in lines 465-471 after this initial hemispheric asymmetry statement.

Figure 11: I really like this figure! Why not add Figure C1 as Figure 11c? It is a nice demonstration of the validation of your ARCNN. Have you looked at other energy transport quantities like sensible heat?

Figure C1 label: Do you mean Figure 11a, rather than top row?