Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-72-AC2, 2020

Geoscientific Model Development Discussions

Interactive comment on "ClimateNet: an expert-labelled open dataset and Deep Learning architecture for enabling high-precision analyses of extreme weather" *by* Prabhat et al.

Prabhat et al.

kkashinath@lbl.gov

Received and published: 4 June 2020

We thank the reviewer for their kind acknowledgement of the efforts and achievements described in our manuscript. We also thank the reviewer for their very positive impression of our work. Furthermore, we are grateful that the reviewer explicitly urges the atmospheric science community to contribute to the expansion of the dataset. Finally, by highlighting some key aspects of this work, such as "the NN trained on this data set actually performs better than those trained on heuristic labels" and "provides to the community a dataset that I believe will accelerate research progress regarding research in tropical cyclones, atmospheric rivers, and other atmospheric phenomena"

the reviewer has brought to the forefront the importance of this work in their review.
==========

Comments and Questions:

The article has many references, but would benefit from a more thorough analysis of existing work on labeling / detecting ARs and tropical cyclones, be that using heuristics or DL. Please expand that section. Here are some references that come to mind: 1) Bonfanti, C., Trailovic, L., Stewart, J., & Govett, M. (2018, July). Ma- chine Learning: Defining Worldwide Cyclone Labels for Training. 2018 21st In- ternational Conference on Information Fusion (FUSION) (pp. 753-760). IEEE. https://ieeexplore.ieee.org/document/8455276 2) C Bonfanti, J Stewart, S Maksimovic, D Hall, M Govett, L Trailovic, I Jankov Detecting Extratropical and Tropical Cyclone Regions of Inter- est (ROI) in Satellite Data using Deep Learning AGU abstract Dec 2018 https://ui.adsabs.harvard.edu/abs/2018AGUFM.H31H1992B/abstract #1 is a good demonstration of how labels are difficult to obtain and #2 is complimentary to your methods of region detection.

We thank the reviewer for pointing out additional references that we missed out. We will certainly include these references with suitable descriptions in the body of the revised manuscript.
==========

P. 5, Line 14. You say "The placement of vertices ceases when a convex hull is created, i.e. when the last vertex coincides with the first vertex." Do you really mean to say "convex hull", or maybe "closed polygon"? Shapes, especially for bounding ARs, are usually not convex (see also Fig. 1).

We thank the reviewer for catching this nuance – yes, indeed, we mean closed polygons and agree that these boundaries are not always convex. We will correct this in the revised manuscript.
==========

Fig. 2: The caption speaks of "yellow masks" for TC labels. In my print-out they look white.

We will update this to “light-colored masks” to avoid ambiguities from printed versus online colors.
==========

Section 3.1.1: I know the model in Section 3.1.1. is neither new, nor the emphasis of this paper. Nevertheless, for the average reader it would be nice to have one more paragraph that explains the functionality of its different elements a bit more intuitively.

We thank the reviewer for pointing this out. We will certainly include a para describing the model with some intuition in the revised manuscript.
==========

Section 3.1.2: You really just use 5 epochs? I guess with so few training samples...

Yes, indeed. We suspect that the large image size (768x1152) and the multiple examples of the same class (TC or AR) in each global snapshot provides more information to the NN than one would normally expect from a single image (say in computer vision examples). Perhaps this aids faster convergence of the NN...
==========

Fig. 4: It’s hard to see the labeling and compare it across the let and right column. Could you use a different color theme? Fig. 6: How about choosing colors that are more different between Expert 1 and Expert 2?

We thank the reviewer for pointing out these issues with clarity. We will attempt to fix these in the revised manuscript.
==========

Section 4.3: Great section that nicely demonstrates the benefits of - and potential way of utilizing - the new data set, and corresponding DL model. I would have liked to see in the tables also the overall increase in precipitation, etc., to see how much that differs

from increase in precipitation due to ARs/TCs. But that’s not crucial.

We thank the reviewer for suggesting additional characteristics of precipitation to examine. This is part of an ongoing detailed investigation of extreme precipitation changes using the DL model for segmentation and full 3-dimensional fields of several other variables (omega, q, T etc.), including attributing changes in extreme precipitation to thermodynamic and dynamic contributions.
==========

Section 5: I really like this section. It has lots of excellent thoughts on limitations and different methods to apply, from active learning (may I suggest Claire Moneleoni as a potential collaborator on that topic?) to transfer learning.

We thank the reviewer for these kind remarks and suggestion. We will follow up with Claire Monteleoni regarding active learning.
==========

I have one comment for the paragraph on Spatio-Temporal Segmentation. I agree that the temporal persistence of weather events could be an excellent criterion you could utilize. However, rather than acquiring expert labels for more datapoints, as you propose in that paragraph, couldn’t you just make this a constraint for your DL method? The simplest solution - Generate labels for several consecutive time steps using your DL method, then compare them, and only report labels that are fairly consistent across time steps? There are many ways to incorporate such constraints. Would be happy to send REFs (e.g., Vipin Kumar’s group at U Minn has done a lot of work in that area, e.g., to detect water bodies from satellite images), but I suspect you already have plenty of ideas of your own.

We thank the reviewer for suggesting ways to incorporate temporal persistence of events into the DL model. We would appreciate additional references and suggestions. We are also considering several other approaches and extensions to improve the performance of the DL model, including reducing false positives and true negatives, by

incorporating ideas from persistent homology and using 3D space-time convolutions.
==========

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-72, 2020.