Geoscientific

Model Development

Discussions

GMDD

Interactive

comment

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 insightful comments and clear elucidation of the strengths and weaknesses of our submission. We also thank the reviewer for their very positive impression of our work. By highlighting the key motivations and objectives of this work, such as "An implicit definition comprised of expert labelled examples avoids the fragility and arbitrariness of hand-crafted heuristics" and "The advantage of pixel-level segmentation of extreme events is made abundantly clear in section 4.3 where conditional precipitation events under a half-degree additional warming scenario are quantified" the reviewer has brought to the forefront the importance of this work in

their review.

In response to the reviewer's comments regarding the weaknesses of this submission, we fully agree with the reviewer that the current curated dataset size of approximately 500 expert-labeled images is indeed quite small for training a state-of-the-art deep learning model such as DeepLabv3+, which is complex and has many parameters. We also resonate with the reviewer's comment (Furthermore, the limited dataset size reflects the fact that hand-labelled expert data is extremely scarce) that obtaining high-quality expert labels is very tedious and time-consuming. Furthermore, the reviewer's recommendation on expanding the size of the ClimateNet dataset and efficient ways of expanding the dataset is very well-received. In fact, that is exactly what we have embarked upon following the submission: instead of having experts label many images from scratch, we are using the trained DeepLabv3+ model's predictions in inference mode (on new images different from the original training data), and then using expert time to correct those predictions. Indeed, this enables experts to correct segmentation masks that are already of high quality, and in particular, delete the false positive TCs and ARs, instead of labeling images from scratch. As the reviewer has rightly pointed out, this "human-in-the loop training" is an effective way to better make use of an expert's time. We anticipate that our dataset size will increase to a few thousand high-quality labeled images in a short period of time. We are confident that with a larger training dataset that will be achieved from the "human-in-the loop training" we will be able to significantly reduce false positives and eradicate the problem of fragmented events.