

Reply on RC1

Tarek Beutler et al.

Author comment on "Deep Learning Approach Towards Precipitation Nowcasting: Evaluating Regional Extrapolation Capabilities" by Tarek Beutler et al., EGU sphere, <https://doi.org/10.5194/egusphere-2022-440-AC1>, 2022

We would like to thank the anonymous referee #1 for their constructive comments and suggestions. We are in the process of revising the manuscript with the referee's suggested changes. Point by point answers to the referee's comments can be found further below.

- **Page 1 Line 6, "...(ConvLSTM) is applied to Radar data of the German...". The experiment is based on TrajGRU. The description of this manuscript is inconsistent. which model was used for the study?** This is indeed a mistake, it should say TrajGRU. This has been revised.
- **There may be significant differences between multiple datasets. The authors may discuss whether this method leads to non-convergence of training.** While we think that the features and the prediction problem are similar enough across datasets to avoid this, we will add some thoughts about datasets with significant property differences in the revised manuscript. For example in a preliminary draft we also took a look at COSMO-REA2, which is a regional reanalysis dataset with a significantly lower temporal resolution. This didn't cause non-convergence, but resulted in significantly worse scores across the board.
- **Page 2 Line 56. What is the "notework"?** This is a spelling error and has been revised.
- **Page 2 Line 57, "...the Critical Success Index (CSI) and the Heidke Skill Score (HSS)...". The evaluation is incomplete without also including the False Alarm Ratio = $FP/(TP + FP)$ and the Probability of Detection = $TP/(TP+FP)$ scores.** The revised manuscript will contain the False Alarm Ratio and Probability of Detection skill scores for all experiments in the appendix.
- **Page 4 Line 100. It is suggested to provide the number of pretrained iterations, which helps to more objectively compare the random initialization model and the pretrained model.** The number of pretrained iterations will be provided in the revised manuscript.
- **Page 7 Line 161. It is suggested to explain why the generalization of features performs better for heavy rainfall than non-heavy rainfall** This is explained on Page 7 Line 164 f. We have added a more detailed explanation in the revised manuscript: The HKO-7 data set is larger and has a significantly higher amount of heavy rainfall events compared to RADOLAN, while having a similar distribution of non-heavy rainfall events (cf. Table 1). Yosinski et al. (2014) show that transferring features (like having learned heavy rainfall events) between networks can improve generalization on data, even after finetuning the network.

- **The results of experimental analysis have been verified in other papers. Did the author draw any other new conclusions?** We believe the anonymous referee is asking if we came to new conclusions when compared to other finetuning studies. While other papers indeed already verified conclusions like improved generalization for smaller datasets, the main conclusion of this paper is that finetuning can be a helpful approach to enable regional extrapolation of neural networks for precipitation nowcasting.