Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-85-RC2, 2019

Atmospheric Chemistry and Physics Discussions
Open Access

Interactive comment on "Technical Note: Deep Learning for Creating Surrogate Models of Precipitation in Earth System Models" *by* Theodore Weber et al.

Anonymous Referee #1

Received and published: 28 May 2019

General comments: 1. I am not sure how meaningful this study is. This study used a sliding window approach to predict global precipitation, i.e., using CNN to simulate the relationship between the precipitations from the most recent K time steps and that at next time step. First, the mapping becomes useless when we need to predict more than one step into the future or to use more/less than K previous time steps. Secondly, the sliding window approach used the fixed window size, incapable of learning the temporal dependence in a dynamic form. 2. I found the description of methodology and numerical experiments is confusing. After reading the manuscript, I am not sure how many network architectures and how many numerical experiments the authors considered. A table listing all of this information would be very helpful. 3. The compar-

Printer-friendly version

Discussion paper

ison with persistence forecasting is not enough to demonstrate the effectiveness and advantages of the deep neural networks. I think a comparison with other advanced time series forecasting methods is necessary, such as autoregression, moving average, and their combinations, and even the more advanced long short-term memory. 4. What is the computational cost to build the surrogate model, such as the number of training samples, the training time, the hyperparameter tuning time? When comparing the methods, besides accuracy, computational costs should be another factor to be considered.

Specific comments: 1. Page 5, Line 1, whether a deep network is needed depends on the problem, i.e., adding depth to the network can improve the model performance of this study. The reason should not be that deep models were successful in recent studies in image classification. As problems are different and the training data size is different, the deep network might not be a good choice of this work. I would like to see a better justification for using the deep network in this work. 2. Page 6, Line 1, If I understand correctly, the training data are 3D images with size m*n*p. What do the authors mean by saying that "The distribution of training data was heavy-tailed and positively skewed"? 3. Page 7, Lines 8-11, I do not understand why not using the ground truth all the time, as errors made in early forecasts would accumulate in later forecasts if the predicted values are used. 4. Page 7, lines 18-19, the comparison is not fair because the baseline CNNs used the best hyperparameters of the residual network. The best set of hyperparameters tuned for the residual network could be a bad choice for the baseline CNNs.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-85, 2019.