Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-278-RC1, 2019

Geoscientific Model Development Discussions

Interactive comment on "Configuration and Intercomparison of Deep Learning Neural Models for Statistical Downscaling" *by* Jorge Baño-Medina et al.

Matteo De Felice (Referee)

matteo.de-felice@ec.europa.eu

Received and published: 21 November 2019

In literature we can find several works about downscaling using statistical methods or machine learning methods: on this topic there are many questions still to answer. This paper contributes with two main positive points: 1. it's part of a rigorous project/experiment framework (VALUE) amd 2. it is fully reproducible (from the data to the algorithms, available on Github and Zenodo). I have a few comments that I think would improve the quality of the paper:

1. The authors should say something about the computational effort of the proposed methods, saying if there is any trade-off between performances (RMSE, correlation,

Printer-friendly version

Discussion paper

etc.) and complexity/computation time. I expect that a linear model should run much faster than a CNN, can the authors say something about this? 2. Possibly related to the point 1. probably: the authors use different CNN setups but then they analyse only the best one (CNN1), can they say something about the others? Why they do not work well? Why they were supposed to work well? Why they are considered in the paper? 3. For the precipitation the authors use a probabilistic score (ROCSS) in addition to the common ones (RMSE, Correlation, etc), it's not clear how the output of a linear model or a CNN could be considered a probabilistic forecast. They should clarify this point. 4. Can the authors comment (or provide reference) on how they decided the best configuration for the CNN? Number of layers, etc. This could be beneficial especially considered that the journal is for a community that, as you say, does not really trust deep learning models. 5. Regarding the comment about deep learning and distrust in climate community, I have the impression that the problem is not just about the extrapolation capabilities, but in general about the impossibility to really know how a black-box model operates. The extrapolation is only a part of it. You can not really assess the capability to "extrapolate" for a complex model like a CNN because any assessment would be 1. configuration specific and 2. data specific. Then I think the problem is a conceptual one: the difficulty in generalising the behaviour of a very complex and highly-nonlinear model. (This point is just a personal comment, I think that this paper is not the right place to for this kind of discussion however I have really appreciated that comment) 6. Can the authors provide a map with the difference between metrics (RMSE or correlation) between GLM4 and CNN1? Can they say something about the areas where CNN/GLM outperforms the other method? 7. The DOI at line 72 does not work 8. There is a typo in the first panel of Figure 1, in the caption title.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-278, 2019.