Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-278-RC2, 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.

Anonymous Referee #2

Received and published: 17 December 2019

This study investigates the ability of convolutional neural networks (CNN) for downscaling daily temperature and precipitation over Europe. The added value is found for extreme temperature and most metrics for precipitation. The paper is well written and the results demonstrate the importance of accounting for non-linearity for downscaling precipitation. I have a few concerns that the authors should be able to address.

Major comments 1. It would be good to say something more specifically about European applications and the related data needs. Climate change studies are mentioned in the text. But nothing is said about the types of applications/users in the Introduction – and this influences the types of information required – e.g., whether spatial consistency

is important, the types of extremes that are relevant.

2. Line 33-39: This study focuses on the deep learning techniques in the context of perfect prognosis SD. However, it's not clear what the difference between classical SD methods and machine learning techniques is. This needs to be mentioned in the introduction.

3. A warm validation period is selected as surrogate of possible future climate conditions to investigate the suitability of CNN in climate change studies. It should be better to clearly state how the models produce extremes which are larger than those in the calibration data, and the ability of models to account for changes in the statistics in the future (related to the stationarity assumption).

4. Different network architectures of CNN have been evaluated and intercompared in this study. However, the authors should provide more interpretations on the impact of these configuration on model performance. There are a few examples where this is currently done (e.g., lines 215-218, 238-244) but this needs to be done more systematically, and highlighted in the conclusion section.

5. The skill of the various downscaling methods is assessed mostly on spatial variability. How could the CNN reproduce the temporal variability of the local climate? You may want to validate the ability of CNN to represent dry/wet spells and interannual variation.

Minor comments 1. Line 13: What does 'classic ones' refer to? Need to make them clear. 2. Line 79: 'such' → 'such as' 3. Line 111: 'vale' should be 'value'. 4. Figure 2: The label 'bias' is misleading here, since the map shows the differences between the test and train periods based on observations. 5. Figure 4 & 6: The best method is in fact different for each metric, but the same best method (CNN10 for temperature and CNN1 for precipitation) for all metrics is indicated in the figure. How do you choose the best performing method, may be based on one metric? 6. Figure 6: Please explain 'DET'(e) and 'STO'(f). 7. Traditional statistical downscaling methods generally require high-resolution obserbations for model training, thus it is difficult to provide downscaled

cliamte simulations for the regions with little observation data. Is the skill of CNN sensitive to the resolution of observations?

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