Review of "Downscaling Multi-Model Climate Projection Ensembles with Deep Learning (DeepESD): Contribution to CORDEX EUR-44" by Baño-Medina et al., submitted to *Geoscientific Model Development*

Recommendation: Major revision

General comments:

Recently, the rapid development of deep learning (DL) technologies has provided novel approaches to construct empirical-statistical models for downscaling GCM outputs. DL can identify and extract complex nonlinear relationships that exist in large data sets and model them to develop the downscaling technique. There are progressively more studies exploring a proof of the concept for developing empirical-statistical models based on the DL technique, as in this work.

In this study, the deep learning algorithm of convolutional neural networks (CNN) is used as a statistical downscaling technique. The CNN-based model was trained using daily predictors from the ERA-Interim reanalysis and predictands from E-OBS v20 observations. And then, the well-trained model was applied to downscale eight GCMs outputs for the historical and RCP8.5 periods. The downscaled projections by the CNN-based model were compared with results from RCMs driven by the same GCMs. The authors found that the CNN-based model can provide comparable climate changes signals to those obtained with the RCMs, with a smaller uncertainty for precipitation.

As RCMs require massive computational resources, the CNN-based model may be plausible to be used as an alternative for downscaling GCM outputs. So, please add some discussions about the computational efficiency of the CNN-based model.

This study shows promising results and the manuscript is well written. However, before being accepted, I think the manuscript could be improved as commented below. I believe my comments/suggestions are not very critical and do not require hard work, and this manuscript would be suitable for publication after appropriate revision.

Specific comments:

Line 12: "… but a similar uncertainty for temperature"
More precisely, it is a larger uncertainty for temperature, as shown in Figure 3.

Line 71−73: The CNN-based model provides smaller uncertainties for precipitation. Does the harmonization process contribute to this respect? Please add discussions about the role of the harmonization process.

Figure 1. Please add units (°C) to the color bar of the temperature bias.

Line 102. The CNN-based model is trained with the observations. So it can be expected that DeepESD exhibits a largely unbiased spatial pattern. The CNN-based model is also used to downscale the GCM outputs for the historical period. How about the downscaled historical GCM simulations by the CNN-based model? Please add discussions about the bias of the CNN-based model for the historical period.

L106. Figure 2 shows that DeepESD shows good performance in reproducing the variability and extremes. I think this capability is important. Please add more discussions about Figure 2.

L114−116: I do not quite understand this sentence. Could you please explain it?