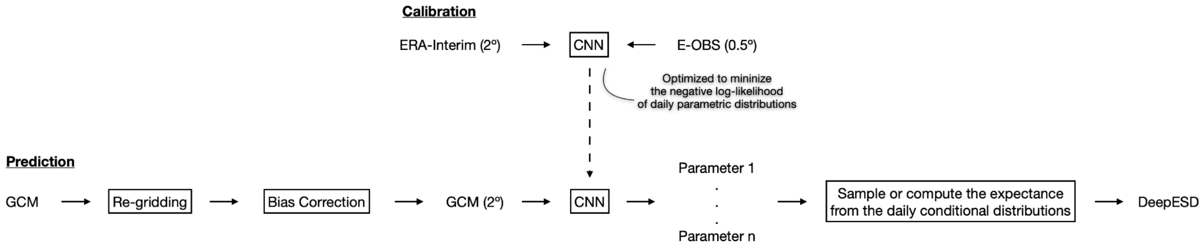
We thank the reviewer for his/her fruitful comments and sincerely acknowledge his/her time for reviewing the manuscript.

- Could you make flowchart of your workflow and include as a figure in the methods section? It is a bit hard to follow your exact procedure.

We will included an schematic flowchart in the revised manuscript. A draft is included below.


Calibration
ERA-Interim (2º)
CNN
E-OBS (0.5º)
Optimized to mininize the negative log-likelihood of daily parametric distributions
Prediction
GCM
Re-gridding
Bias Correction
GCM (2º)
CNN
Parameter 1
Parameter n
Sample or compute the expectance from the daily conditional distributions
DeepESD

- Isn‘t the comparison to observations between unconstrained mechanistic models (i.e. GCMs) and CNNs trained on observations "unfair"? If you did some nudging procedure with GCMs you would also end up with model output better fitting observations. For the CNN training, did you split the observational data into train (validation) and test set (only train on 20 years and show performance for 10 years)? Again, a flowchart would help to understand what you did. If you show the performance of DeepESD for the test set and compare that to GCM output, it'd be "more fair", but still, just by design we would expect that the CNN reproduces observations better than GCMs.

  The flowchart indicated in the previous question probably sheds light on this one as well. Indeed, the train and test sets differ in the predictor datasets (ERA-Interim for training and GCMs for testing) rather than in the temporal period. As the reviewer points out, the comparison between GCM and DeepESD is still not totally "fair" in the historical period. However, in the manuscript we do not intend to establish an argument in this line, but rather to compare the climate change signals between ensembles whilst showing DeepESD a good reproducibility of the local scale in the historical period.

- You show that the CNN learns the "necessary" dynamics based on predictors of the historical period and extrapolates reasonably well using predictors from GCM output for projections. That is a very interesting point. I wonder if this simple bias correction for GCMs really does the trick, as the models considerably diverge over the climatic time-scales and very model specific regional biases emerge. Can you comment on whether other bias-correcting measures were tested?

Overall, there is certainly a long list of potential further evaluation and testing steps that could be undertaken, but maybe it is enough for this model description paper.

We agree with the reviewer that there are many factors that could have some influence in the downscaled results obtained with our CNNs which may be worth analyzing further. For instance, we are currently exploring the influence that the application of different bias correction techniques over the predictor space may have in the downscaled climate change signal. However, in order to not add more complexity to the present manuscript, we plan to publish the results from this analysis in a future paper..

For the moment we have compared the downscaled projections that are obtained based on two different approaches for bias correcting the GCM predictors —(1) bias correction of the mean at a monthly scale following [1], and (2) bias correction of the mean and standard deviation at a monthly scale (DeepESD),— with no major differences found.

[1] *Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. "On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections." Climate Dynamics 57.11 (2021): 2941-2951.*

- L5: What is DeepESD standing for? Please introduce acronym before first usage.

  It stands for Deep learning Empirical Statistical Downscaling (DeepESD). We have introduced this acronym in the manuscript.

- LL33-34: The "perfect prognosis" approach is based on the assumption that GCMs don't have systematic biases with respect to the observations that were used for training, right? Maybe you should include a short sentence here that addresses this aspect.

  We will include a comment on this in the revised paper.

- L55: I recommend to use another more static hosting platform for your code, e.g. Zenedo (https://zenodo.org/).

  The code is already hosted in Zenodo. However in the manuscript we point to both the GitHub repository (https://github.com/SantanderMetGroup/DeepDownscaling) and Zenodo (DOI:10.5281/zenodo.3461087) which might be confusing. Also, based on the comment from the chief editor we have removed all references to GitHub and just stick to Zenodo.

- L60: Why did you use ERA-Interim reanalysis? It is outdated for quite some time now.

This study builds on previous ones which use ERA-Interim data to deploy CNNs over Europe [1,2]. For consistency with these studies and also with reference statistical downscaling experiments in the continent [3,4,5,6] which also build on ERA-Interim data, we decided to use this dataset for the predictors. However, we plan to move to ERA5 to downscale CMIP6 GCMs in future work.

[1] *Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. "Configuration and intercomparison of deep learning neural models for statistical downscaling." Geoscientific Model Development 13.4 (2020): 2109-2124.*
[2] *Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. "On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections." Climate Dynamics 57.11 (2021): 2941-2951.*
[3] *Bedia, Joaquín, et al. "Statistical downscaling with the downscaleR package (v3. 1.0): contribution to the VALUE intercomparison experiment." Geoscientific Model Development 13.3 (2020): 1711-1735.*
[4] *Maraun, Douglas, et al. "VALUE: A framework to validate downscaling approaches for climate change studies." Earth's Future 3.1 (2015): 1-14.*
[5] *Maraun, Douglas, Martin Widmann, and José M. Gutiérrez. "Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment." International Journal of Climatology 39.9 (2019): 3692-3703.*
[6] *Gutiérrez, José Manuel, et al. "An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment." International journal of climatology 39.9 (2019): 3750-3785.*

- L62: I don’t understand your use of dashes (—) in the manuscript. Please check whether the make sense throughout the manuscript.

  This character is going to be checked across the entire notebook as the reviewer suggests.

- LL62-65: What about adding high-resolution orography description as static predictor?

  This is done in other studies (e.g., super-resolution, where high-resolution predictors are used to downscale target variables) but is not a standard approach for perfect-prognosis downscaling.

- L85: Why did you analyze both and can you provide the reason why you settled with the deterministic one?

  Sampling from the conditional distributions permits reproducing the temporal variability of the local time-series. However, since sampling is performed at each gridpoint individually, there is a loss in the spatio-temporal structure of the downscaled fields. Since local temperature is largely explained by the large-scale predictors, there is no need to sample from the inferred conditional distributions. Conversely, local precipitation is not completely explained by the predictor set, and a stochastic downscaled version is needed to recover the temporal variability of the target predictand. These aspects were already analyzed in a prior study [1] and for this reason we do not delve into too much detail in this manuscript.

[1] *Baño-Medina, Jorge, Rodrigo Manzanas, and José Manuel Gutiérrez. "On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections." Climate Dynamics 57.11 (2021): 2941-2951.*

L88: Please stick to the tenses (in this paragraph you mix present and past tense), i.e. do not switch between present and past tense when describing your results or methods. I recommend that you always use present tense when talking about your study, i.e. when describing your methods, your results etc., and use past tense when referring to already published studies.

Solved

- L137: “contribute to increasing”

  Solved.

- Figure 1: Add unit at lower right colorbar. Also, it’d be useful if you could include letter characters as pointers to subplots, e.g. a,b,c,d. This comment applies for all Figures.

  We will consider this suggestion in the revised version of the manuscript.

- Figure 4: Please be more specific about the numbers in the plots. Please provide more detailed information in the caption.

  Solved.

- Figure 5: The mid-column misses a time axis. DeepESD is not “yellow” but “green”, no?

  It is green. This typo has been corrected in the new version of the manuscript.