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

- With which criteria were chosen the predictand fields?
  We chose to downscale air surface temperature and daily total precipitation since they are two variables highly demanded by the climate, impact and adaptation communities. These are also the variables with best observations available and, therefore, provide reliable training data sets. It would be interesting to extend this study to other relevant predictand variables (e.g., wind) and we leave this possible continuation to future work.

- How expensive in computer resources is the method?
  We have included the following sentence in the manuscript:
  "We lean on a 2x Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz (16 cores, 32 threads) with 60 GiB of memory RAM. The computational time employed to calibrate the model and generate the projections for e.g., one GCM is approximately 5-6 hours, being considerably lower than the one needed to run a RCM."

- Why eobs was used? It is too smooth, what can be seen in the results, specially in places with high topography. Why did not used a regional, high resolution reanalysis as predictands?
  We have added the following paragraph to the manuscript:
  "E-OBS is a high-resolution observational dataset generated through an interpolation procedure of the European Climate Assessment & Dataset (ECA&D, [1]) station network. Whilst national and sub-national datasets exist, E-OBS accurately represents the regional climate over the entire European continent [2] and it is commonly employed in statistical downscaling experiments on a continental level [3,4,5,6]. We chose version 20 (v20, release date October 2019) since it was the latest one at the beginning of this study."

  [1] Klok, E. J., and A. M. G. Klein Tank. "Updated and extended European dataset of daily climate observations." *International Journal of Climatology: A Journal of the Royal Meteorological Society* 29.8 (2009): 1182-1191.
  [2] *Bandhauer, Moritz, et al. "Evaluation of daily precipitation analyses in E-OBS (v19. 0e) and ERA5 by comparison to regional high-resolution datasets in European regions." International Journal of Climatology 42.2 (2022): 727-747.*
  [3] *Maraun, Douglas, et al. "VALUE: A framework to validate downscaling approaches for climate change studies." Earth's Future 3.1 (2015): 1-14.*
  [4] *Vrac, Mathieu, and Pradeebane Vaittinada Ayar. "Influence of bias correcting predictors on statistical downscaling models." Journal of Applied Meteorology and Climatology 56.1 (2017): 5-26.*
  [5] *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.*
  [6] *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.*

- It seems that the use of more output layers for the precipitation than in temperature makes the biases in the downscaling of precipitation as small as for temperature, but reduces the standard deviation in the downscaling (Figure 3) . I think that this fact is related to the methodology and should be commented on by the authors.

  Both temperature and precipitation topologies have the same number of filter maps and hidden layers in their topologies and, thus, they can achieve the same degree of nonlinearity. The CNNs deployed contain different number of, n, output neurons (2*n for temperature and 3*n for precipitation), representing each of the statistical parameters of the parametric distributions estimated per gridbox. For temperature, we learn Gaussian daily conditional distributions parameterized by 2 parameters (mean and standard deviation) per predictand site, while for precipitation there are 3 parameters (probability of rain, shape and scale factor) corresponding to the Bernoulli-Gamma distribution. Having more output layers does not add non-linearity to the network, and therefore this aspect does not have an influence on the future estimates and indices —either biases or standard deviation. For more details on the topology we refer to [1].

  [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.*

- Also, the fact that the simulation of R01 in DeepESD is closer to the RCMs that to the GCMs shows the importance of a good simulation of orographic precipitation, while SDII and Mean temperature in DeepESD and GCM are closer, probably reflecting the tuning of the GCMs (which usually is not made in RCMs) and the training with observations in DeepESD. The exception for temperature in ED looks strange for me and would be nice if you explain this behavior.
  This aspect is related to what is explained in lines 126-127: “In the case of the RCMs, some recent studies attribute these differences to the lack of time-varying anthropogenic aerosols in the RCM formulation (Boé et al., 2020; Gutiérrez et al., 2020)”. Therefore, there is an on-going analysis by the dynamical downscaling community to analyze the differences mentioned by the reviewer in Eastern and Central Europe of the climate change signal of temperature between the GCMs and RCMs, investigating whether it is due to an added value of dynamical downscaling or to deficiencies in the model formulation of the RCMs.

- Results in figures 3,4 and could be also contributed by the use of stochastic (deterministic) approaches for the precipitation (temperature) specific comments
  The choice for either stochastic or deterministic downscaled fields is mainly relevant to the reproduction of extremes (Figure 2), but they do have a negligible influence on the results of Figure 3 and 4, which display only the mean of temperature and precipitation. This comparison among stochastic and deterministic fields can be encountered in [1].

[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.*

- A more detailed description of the methodology for not specialists (most readers, I guess) should be interesting. Can be added as an appendix

  We agree with the reviewer that some concepts related to the deep learning terminology (e.g., convolutional layer, filter map and kernel), which are unfamiliar for non-machine-learning researchers, were hardly explained in the manuscript, difficulting the understanding of the proposed topology. For this reason we have added the following sentences:
  "In particular, we deploy the best performing topologies developed in [1], a recent study which intercompares different CNNs over Europe in ``perfect" conditions to downscale precipitation/temperature. They consist of 3-layers with 3x3 kernels of 50, 25 and 1/10 filter maps followed by a dense connection which links all the neurons in the last hidden layer to the output neurons at each land gridpoint in E-OBS. For precipitation (temperature), these CNNs are trained to optimize the negative log-likelihood of a Bernoulli-Gamma (Gaussian) distribution, yielding thus daily estimates of 3 (2) parameters per predictand site (e.g., for precipitation $n^o_{\text{output neurons}} = n^o_{\text{predictand sites}} * 3$) representing the probability of rain, shape and scale (mean and variance). We refer the reader to [1] for a detailed computational analysis and more details in the topology."

- How does the interpolation method influence the results?
  To analyze this issue we used two different approaches to re-grid the GCM predictor fields to a common 2º latitude-longitude: nearest-neighbour and bilinear interpolation. We found no remarkable differences in the climate change signals obtained for these two interpolation methods. We will add a comment on this to the revised manuscript.

- In the Iberian Peninsula and the Scandinavian peninsula the climate change signal in DeepESD is similar to that of the global models, while the opposite is true in central Europe. Could you elaborate on this?
  We agree with the reviewer that this aspect is very interesting. To understand the nature of these similarities and differences between the GCM and DeepESD ensembles —beyond a comparison with RCMs, which is what is done in this manuscript,— we have designed a few experiments that we plan to publish in a future paper (outlined in lines 185-189). These experiments consist on analyzing (1) the adaptability of the statistical model learned in perfect conditions (i.e., with reanalysis data for the predictors) to the climate model space, (2) the extrapolation ability of CNNs on pseudo-reality experiments, and (3) the influence of each predictor in the climate change signal.