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

- Line 32: “perfect prognosis” please explain the term in detail
  We have extended the explanation of “perfect-prognosis” resulting into the following paragraph:
  “Under the “perfect-prognosis" (PP) approach, the statistical models are trained using observations (both for predictors and predictands) in a historical period and afterwards applied to GCM predictors from global projections to obtain the regional/local downscaled results. In particular, reanalysis data are used for the predictor set during calibration. Several assumptions need to be fulfilled in any PP downscaling setup [1]; (a) predictors have to be realistically simulated by GCMs and project with certain plausibility the climate change signal, (b) these predictors to be informative to the regional scale, and (c) the statistical models have to be flexible enough to learn the complex patterns inherent in the predictor-predictand link.”

  [1] *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.*

- Lines 35-36: please include 1-2 sentences to introduce dynamic downscale as a comparison to statistical downscale
  We have included the following sentence: “As compared to DD, PP lacks explicit physics in the model formulation, but overcomes systematic biases present in RCM products, since the model is trained using observations. Regarding computational requirements PP has smaller requirements avoiding the need for large computational infrastructures [1,2].
  [1] Le Roux, Renan, et al. "Comparison of statistical and dynamical downscaling results from the WRF model." Environmental modelling & software 100 (2018): 67-73.
  [2] 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.

- Line 61: Please provides more information about “E-OBS v20”
  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.*

- Line 62: I assume "—" is a typo?
  Yes, thank you for noticing it.

- Line 70: can you explain the "harmonization process" further?
  We have included the following paragraph:
  "Moreover, we have applied a harmonization process suggested in previous works [1,2,3], to increase the distributional similarity between the GCM and reanalysis fields. This post-processing consists of bias adjusting the seasonal cycle of the historical and RCP8.5 predictor variables towards the one described by the reanalysis fields over the reference period 1979-2005. A simple adjustment (mean and variance) applied at a monthly scale is used to keep this harmonization as simple as possible.

  [1] *Brands, Swen, et al. "Validation of the ENSEMBLES global climate models over southwestern Europe using probability density functions, from a downscaling perspective." Climate Research 48.2-3 (2011): 145-161.*
  [2] *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.*
  [3] *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.*

- Lines 75-77: The method of DL should be explained in further detail. "They consist of … (one per each gridpoint in E-OBS)". I found many terms in these sentences that might be barriers to fully understand the method. Can you rephrase it?
  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:
  (introduction) A 2-D convolutional layer convolutes a set of k = 1, 2, ..., K 2-D parameters called kernel over the input space (or previous hidden layer), generating K filter maps, which are spatial representations of the K patterns optimized by the network (we refer the reader to [1] for more details)
  (Data and Methods) 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 (one per 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_{output\ neurons}$ = $n^o_{predictand\ sites}$ * 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.

[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.*

- Line 97-98: Merge the single sentence to the following paragraph
  Done.

- Line 120: It is interesting that the DeepESD has the smallest ensemble spread over the historical period (Fig. 2) but has the largest one over the future. Any explanation for that?
  In the historical period the predictor fields are all standardized with the same mean and standard deviation —regardless of the GCM considered,— thus leading to a small ensemble spread. Nevertheless, in the future the ensemble spread of DeepESD is the result of 1) the divergence in the trends/evolution of the predictor fields, and/or to 2) the extrapolation ability of the CNN.

- 125: what do "these differences" refer to?
  It refers to the differences in the climate change signal between the GCM and the RCM/DeepESD, which are described in the preceding paragraph. In particular, the differences between the GCM and the DeepESD are displayed in rows 1-2 of Figure 4.

- Line 130: "These differences are quite systematic for the case of precipitation indicating a robust CNN extrapolation fingerprint." Can you explain in further detail?
  Thank you for this comment. We agree that this sentence may be misleading. We have decided to replace it by the following one: "For the case of precipitation, DeepESD modifies the signal differently for each GCM, pushing the values closer across GCMs and thus reducing the inter-model uncertainty (Figure 3)."

- Figure 4: How do you produce Row 3? Is it the difference (DeepESD minus E-OBS v2.0) shown in each model in row 2 minus the mean of the difference? In that case, the mean of 8 panels in row 3 should be 0, but clearly they are not.
  In row 2 we display the difference (Diff) between the climate change signal for mean temperature projected by DeepESD ($Delta_{CNN}$) and the one projected by the GCM ($Delta_{GCM}$). In row 3, for each GCM (i.e., column) we compute the difference between Diff and the spatial average of Diff ($\overline{Diff}$). Since Diff is a spatial field (row 2), whilst $\overline{Diff}$ is just a number, the operation Diff - $\overline{Diff}$ can be different from 0.