Downscaling Multi-Model Climate Projection Ensembles with Deep Learning (DeepESD): Contribution to CORDEX EUR-44

Jorge Baño-Medina[1], Rodrigo Manzanas[2,3], Ezequiel Cimadevilla[1], Jesús Fernández[1], Jose González-Abad[1], Antonio S. Cofiño[1], and José Manuel Gutiérrez[1]

[1]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander,Spain
[2]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, Santander, Spain
[3]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain

Correspondence: Jorge Baño-Medina (bmedina@ifca.unican.es)

Abstract. Deep Learning (DL) has recently emerged as an innovative tool to downscale climate variables from large-scale atmospheric fields under the perfect prognosis (PP) approach. Different Convolutional Neural Networks (CNN) have been applied under present-day conditions with promising results, but little is known about their suitability for extrapolating future climate change conditions. Here, we analyze this problem from a multi-model perspective, developing and evaluating an ensemble of CNN-based downscaled projections (hereafter DeepESD) for temperature and precipitation over the European EUR-44i (0.5°) domain, based on eight GCMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5). To our knowledge, this is the first time that CNNs have been used to produce downscaled multi-model ensembles ~~of downscaled projections~~based on the perfect-prognosis approach, allowing to quantify inter-model uncertainty in climate change signals. The results are compared with those corresponding to an EUR-44 ensemble of regional climate models (RCMs) showing that DeepESD reduces distributional biases in the historical period. Moreover, the resulting climate change signals are broadly comparable to those obtained with the RCMs, with similar spatial structures. As for the uncertainty of the climate change signal (measured on the basis of inter-model spread), DeepESD ~~yields a smaller uncertainty for precipitation, but a similar uncertainty for temperature.~~ preserves the uncertainty for temperature and results in a reduced uncertainty for precipitation.

To facilitate further studies of this downscaling approach we follow FAIR principles and make publicly available the code (a Jupyter notebook) and the DeepESD dataset. In particular, DeepESD is published at the Earth System Grid Federation (ESGF), as the first continental-wide PP dataset contributing to CORDEX (EUR-44).

1 Introduction

The Coupled Model Intercomparison (CMIP) initiative produces periodic multi-model ensembles of centennial global climate projections under different future scenarios using Global Circulation Models (GCMs). The two latest ensembles available are CMIP5 (Taylor et al., 2012) and CMIP6 (Eyring et al., 2016), with typical resolutions of around 200 and 100km, respectively. These results are widely used by the impacts and adaptation communities in different sectors (e.g., energy, agriculture and health, among others). However, the biases and spatial resolution of these global projections ~~hampers~~ hamper their use in regional applications and different downscaling approaches and methods are routinely applied to produce actionable information at the regional and local scales (Maraun and Widmann, 2018).

Dynamical downscaling is based on the use of Regional Climate Models (RCMs) over a limited region driven by GCM outputs at the boundaries (Giorgi, 2019; Gutowski et al., 2020). Different regional initiatives provide high-resolution, physically consistent downscaled simulations over continental-wide domains. In particular, the Coordinated Regional Climate Downscaling Experiment (CORDEX, https://cordex.org) provides multi-model ensembles of regional climate projections driven by CMIP5 model outputs over 14 continental domains. These regional projections are highly demanding in terms of computational resources and the resolution of the available regional projections ranges from 50 to 10km, depending on the domain.

The empirical-statistical downscaling approach (ESD) is based on empirical/statistical models translating the coarse-resolution information provided by the GCMs (*predictors*) to the regional/local scale provided by the available historical observations (*predictands*), typically temperature or precipitation fields (Gutiérrez et al., 2019). Under the “perfect-prognosis~~"~~” (PP) approach, the statistical models are trained ~~using observations (reanalysis for the predictors)~~ in a historical period ~~and afterwards~~ to learn a predictor-predictand link using simultaneous observed and reanalysis (quasi-observations) values (daily in this work) for predictands and predictors, respectively. The resulting models are then applied to GCM ~~predictors from global projections~~ predictor values (from present climate or future scenarios) to obtain the regional/local downscaled results. ~~These methods~~ This approach is based on a number of assumptions. For example, predictors have to be realistically simulated by GCMs (e.g. exhibiting small systematic biases), so large-scale fields in upper levels (less affected by orography and model resolution) are typically used as predictors (perfect prognosis assumption); moreover, the statistical models trained in present climate conditions should remain valid under modified (out-of-sample) climate conditions (generalization assumption) (see Gutiérrez et al., 2019, for more details). Compared to dynamical downscaling, ESD lacks explicit physics in the model formulation and typically does not ensure full multivariate (intervariate and spatial) consistency. However, these methods overcome the systematic biases present in RCM products (as the model is trained using observations) and are not computationally demanding~~and therefore could be extensively~~ , avoiding the need for large computational infrastructures (Le Roux et al., 2018) . Therefore, these methods could be widely used to downscale global multi-model ensembles providing ~~continental-wide regional projection fields~~results at continental scales, e.g. ~~over the~~ in CORDEX domains.

Recently, deep learning methods based on Convolutional Neural Networks (CNNs) have become very popular as a statistical downscaling technique due to their ability to achieve an automatic selection of predictors in the form of data-driven spatial features (Baño-Medina, 2020). Although they have shown promising results for continental-level climate downscaling under “perfect” conditions (Pan et al., 2019; Baño-Medina et al., 2020; Sun and Lan, 2021; François et al., 2021), there is little knowledge on whether these statistical models are able to generalize to out-of-sample climate change conditions. Some preliminary work using a single GCM shows that CNNs can accurately reproduce the local climate variability and provide plausible climate change projections over Europe as compared to well-established statistical downscaling approaches (Baño-Medina et al., 2021). However, further analysis along these lines is needed to assess the suitability of CNNs for climate change applications.

Here we provide a multi-model perspective by applying a CNN model (Baño-Medina et al., 2021) to downscale daily precipitation and temperature over Europe from the historical and future projections (RCP8.5 scenario) provided by an ensemble of eight GCMs. We evaluate the consistency of the downscaling approach across models and analyze the uncertainty of the resulting climate change signals. Moreover, we follow previous downscaling literature (Vrac et al., 2007; San-Martín et al., 2017; Quesada-Chacón et al., 2021) and compare the resulting projections with an ensemble of RCMs, which are used as “pseudo-observations". In order to facilitate further analysis, this dataset (referred to as ~~DeepESD~~Deep learning Empirical Statistical Downscaling (DeepESD)) is made publicly available on the Earth System Grid Federation (ESGF), as a contribution to the EUR-44i domain (0.5° horizontal resolution), so it can be downloaded together with the ensemble of available RCMs. To our knowledge, this is the first continental-scale climate change projection dataset produced using statistical downscaling methods contributing to CORDEX and published in ESGF, following the standard procedure for RCMs. Moreover, following FAIR principles (Wilkinson et al., 2016), the code used to generate the dataset along with guidelines on how to access the data is available on ~~GitHub~~ Zenodo (see the section on code and data availability).

2 Data and Methods

~~DeepESD has been produced based on the application of CNN-based downscaling methods to an ensemble of eight CMIP5 GCMs (Table 1).~~ Following the PP approach, the ~~employed CNNs~~ CNN models have been trained over the period 1979-2005 using daily predictors from the ERA-Interim reanalysis (Berrisford et al., 2011)~~— upscaled~~ , upscaled from its original 0.75° resolution to a reference 2° regular ~~grid —~~ grid, and predictands from E-OBS v20 (Cornes et al., 2018)~~— originally~~ , originally at 0.25° but upscaled to 0.5° for consistency with previous works (Baño-Medina et al., 2020; Baño-Medina et al., 2021). ~~—~~ E-OBS is a high-resolution observational dataset generated by spatially interpolating the European Climate Assessment & Dataset (ECA&D) network of stations (Klok and Klein Tank, 2009). Although national and sub-national datasets exist, E-OBS accurately represents the regional climate over the entire European continent (Bandhauer et al., 2022) and it is commonly used in continental-wide statistical downscaling experiments (Maraun et al., 2015; Vrac and Ayar, 2016; Baño-Medina et al., 2020; Baño-Medir . We chose version 20 (v20, release date October 2019) since it was the most recent at the start of this study. Following previous studies (Gutiérrez et al., 2019; Baño-Medina et al., 2020), air temperature, specific humidity, geopotential, meridional and zonal wind velocity at 500, 700, 850 hPa (i.e., a total of 15 variables per gridbox) have been used as predictors ~~— covering~~ covering the domain $34^\circ N$-$76^\circ N$, $8^\circ W$-$34^\circ E$, ~~—~~ resulting in a 22×22×15 (longitude×latitude×variable) high-dimensional input grid. To avoid potential artifacts derived from the different scale of the distinct variables~~the~~ , ERA-Interim predictors are standardized at the gridbox level (Baño-Medina et al., 2021).

~~The trained models are applied to downscale the projections from the ensemble of~~ For downscaling we used an ensemble formed by the eight CMIP5 GCMs described in Table~~ 1, for the~~ 1, whose ability to reproduce the large-scale dynamics has already been assessed for PP studies (Brands et al., 2013) and have also been used in EURO-CORDEX to drive RCMs (Vautard et al., 2021). Therefore, we apply our trained models to downscale the projections from the this ensemble for the historical (1975-2005) and RCP8.5 scenario (2006-2100) periods. We follow previous work in this field (Baño-Medina et al., 2021; Olmo et al. and select the RCP8.5 scenario, which shows the strongest climate change signal (especially for temperature) and therefore allows the generalization capability of CNNs to be optimally explored. Due to their different spatial resolutions, all GCM data

Name	Institution	Spatial resolution
CanESM2 (Christian et al., 2010)	Canadian Centre for Climate Modelling and Analysis	(2.81°, 2.79°)
CNRM-CM5 (Voldoire et al., 2013)	Centre National de Recherches Météorologiques and Centre Européen de Recherche et de Formation Avancée	(1.4°, 1.4°)
MPI-ESM-MR (Müller et al., 2018)	Max-Planck Institut für Meteorologie	(1.87°, 1.87°)
MPI-ESM-LR (Müller et al., 2018)	Max-Planck Institut für Meteorologie	(1.87°, 1.87°)
NorESM1-M (Bentsen et al., 2013)	Norwegian Climate Center	(2.5°, 1.9°)
GFDL-ESM2M (Dunne et al., 2013)	National Oceanic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory	(2.5°, 2.02°)
EC-EARTH (Doblas Reyes et al., 2018)	European-wide consortium	(1.12°, 1.12°)
IPSL-CM5A-MR (Dufresne et al., 2013)	Institut Pierre Simon Laplace Climate Modelling Center	(2.5°, 1.27°)

Table 1. The different CMIP5 models used in this study.

have been interpolated to the reference 2° grid ~~—the nearest gridboxwas considered—~~ (considering the nearest gridbox) to match the predictor space used for ERA-Interim. ~~Moreover, we have applied a harmonization process suggested in previous works (Brands et al., 2011; Vrac and Ayar, 2016; Baño-Medina et al., 2021), bias adjusting the~~ No differences in the downscaled results were found by employing other interpolation techniques (e.g., bilinear). Moreover, in order to reduce potential systematic biases in GCM predictors which may affect the perfect prognosis assumption, we bias adjust GCM predictors towards the corresponding reanalysis values. ~~A simple scaling~~ As suggested in previous studies we use a change-preserving method (Vrac and Ayar, 2016) in order to avoid introducing artificial trends/changes in future GCM predictor values. In particular we use a simple scaling method (mean and variance) applied at a monthly scale~~is used to keep this harmonization as simple as possible.~~ ; for future periods, the climate change signal is removed from the data before bias adjustment and added to the results. We want to remark that we tested both signal-preserving and standard bias adjustment obtaining substantial differences in the climate change signals for temperature; signal-preserving yields more plausible results (as compared with GCM and RCM climate change signals). As in the case of the reanalysis, GCM predictors are standardized at the gridbox level for their use in the CNN (the same standardization parameters used for the reanalysis data is applied here).

The above pre-processing steps are illustrated in Fig. 1.

~~In particular~~For the CNN models used in this work, we deploy the best performing topologies developed in ~~Baño-Medina et al. (2020)~~ (Baño-Medina et al., 2020), a recent study which intercompares different CNNs over Europe ~~in "perfect" conditions to downscale precipitation/temperature~~ to downscale temperature (precipitation). They consist of ~~3-layers with 3×3 kernels of~~ three convolutional layers (LeCun et al., 1995) with 50, 25 and ~~1~~/10 ~~filter maps~~ (1) spatial kernels (3x3 gridboxes) followed by a dense connection ~~which links~~ linking the last hidden layer ~~with~~ to the output neurons (~~one per each gridpoint~~ corresponding to the land gridpoints in E-OBS). ~~For precipitation (temperature), these CNNs~~ As in (Baño-Medina et al., 2020) we apply a distributional downscaling approach and use the network to estimate daily predictor-conditioned Gaussian (Bernoulli-Gamma) distributions for temperature (precipitation). This is implemented for each land gridbox using two (three) output neurons corresponding

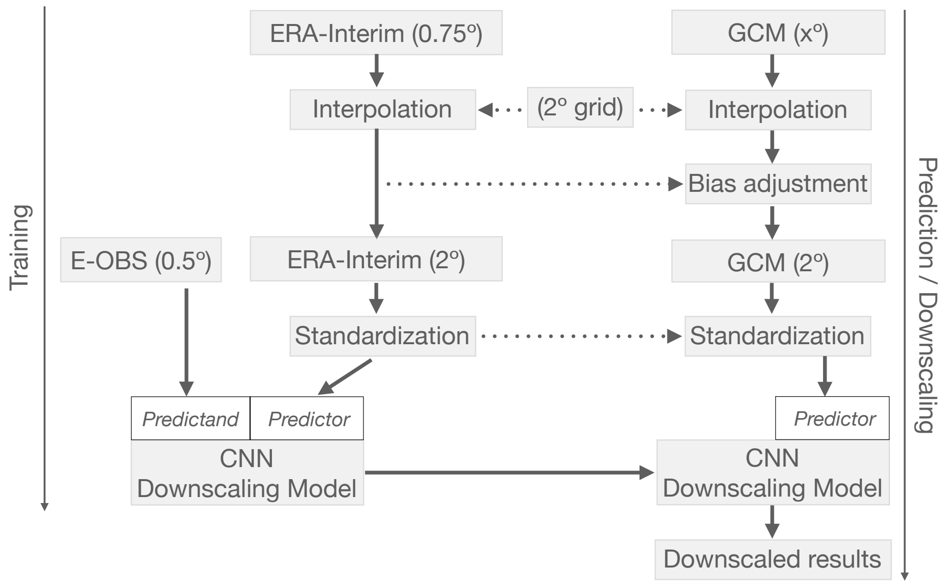

Figure 1. Workflow of pre-processing steps applied to reanalysis and GCM data in this work.

GCM	member	RCM	Institution
CanESM2	1	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
CNRM-CM5	1	CLMcom-CCLM5-0-6	Climate Limited-area Modelling Community
CNRM-CM5	1	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
MPI-ESM-LR	1	CLMcom-CCLM4-8-17	Climate Limited-area Modelling Community
MPI-ESM-LR	1	MPI-CSC-REMO2009	Max Planck Institute for Meteorology
NorESM1-M	1	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
GFDL-ESM2M	1	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
EC-EARTH	12	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
EC-EARTH	12	CLMcom-CCLM5-0-6	Climate Limited-area Modelling Community
IPSL-CM5A-MR	1	SMHI-RCA4	Swedish Meteorological and Hydrological Institute, Rossby Centre
IPSL-CM5A-MR	1	IPSL-INERIS-WRF331F	Institut Pierre-Simon Laplace

Table 2. Details of the EURO-CORDEX (EUR-44 domain) RCM simulations used in this study. The first two columns show the GCM and ensemble member driving the RCM.

to the distributional parameters: mean and variance (probability of rain, shape and scale factors). The resulting networks are trained to optimize the negative log-likelihood of ~~a Bernoulli-Gamma (Gaussian) distribution, yielding thus daily estimates of its probability of rain, shape and scale parameters (mean and variance) .~~ the Gaussian (Bernoulli-Gamma) distribution. We refer the reader to Baño-Medina et al. (2020) for more details. During calibration, we use a test set (randomly selected 10% of the data) to perform early-stopping, and stop training when the test error stops decreasing after 30 epochs.

The computations performed in this work were executed on a single node 2x Intel(R) Xeon(R) E5-2670 0 @ 2.60GHz CPU (16 cores) with 60 GiB of RAM. The computational time taken to calibrate the model and generate the projections for a GCM was less than six hours, which is considerably less than the time required to run a similar experiment with an RCM (for instance, the EUR-44 simulations performed with the WRF model for a single GCM in Fernández et al., 2019, lasted six months usin . This approach can provide either deterministic predictions, by considering the expected value of the distribution for each day and grid-point, or stochastic ones, by simulating a random value from the distribution. Note that the deterministic approach typically result in an underestimation of the variability (and the extremes), since the explained variance may be significantly smaller than the observed one (Williams, 1998; Cannon, 2008; Baño-Medina et al., 2020). This is especially relevant for precipitation, whose local variability is often influenced by local phenomena which are not captured by the chosen predictors (Schoof and Pryor, 2001; Maraun and Widmann, 2018). We analyzed both deterministic and stochastic approaches and finally used the stochastic (deterministic) version of the precipitation (temperature) downscaled fields. For the stochastic version we tested the results for different realizations and found robust results for historical biases and climate change signals.

We use a set of CORDEX RCMs (EUR-44 domain, Table 2) to analyze the generalization to out-of-sample climate change conditions of the CNN-based regional projections. Using RCM simulations as “pseudo-observations" is a common procedure adopted in the literature to validate ESD downscaled projections for future scenarios (Vrac et al., 2007; San-Martín et al., 2017; Quesada-Chacón et al., 2021). Nevertheless, note that RCMs still suffer from deficiencies in their model formulations that may affect their futures estimates (Boé et al., 2020; Gutiérrez et al., 2020), and therefore they should not be considered as purely true values for the CNN projections but rather as plausible trajectories. For a direct comparison, ~~these RCMs were interpolated~~ we interpolate these RCMs from their original spatial resolution (0.44°) to the predictand 0.5° regular grid.

Finally, we ~~tested~~ test the sensitivity of CNN training on the results by repeating the downscaling experiment ten times and ~~evaluating~~ evaluate historical biases and future climate change signals as shown below without finding appreciable variations.

3 Results

Figure 2 shows mean daily precipitation and temperature over the historical period 1975-2005 (and biases relative to E-OBS) for the multi-model means provided by the GCMs, RCMs and DeepESD ensembles.

~~Annual mean daily precipitation (left block) and temperature (right) for the 31-year period 1975-2005, as obtained from the GCMs, RCMs and DeepESD ensemble means (left, middle and right columns, respectively). The bottom row displays the corresponding biases with respect to E-OBS v20.~~

For precipitation, the raw GCM results show a smooth spatial pattern which does not capture the strong local-to-regional variability of this variable, and both GCM and RCM overestimate rainfall over most of the domain. As expected, DeepESD exhibits a largely unbiased spatial pattern over the entire continent, which is a result of being trained directly with observations. For temperature all approaches capture the latitudinal gradient, but both GCM and RCM results exhibit important biases over vast regions of the continent with predominant negative biases for RCM results. Again, DeepESD yields a mostly unbiased spatial pattern ~~. The reduction of biases in the mean is~~ as a consequence of the training process (Casanueva et al., 2016)~~, but this reduction is also notable~~. Besides these results for the mean, Figure 3 compares the entire precipitation and temperature distributions for the GCM, RCM and DeepESD ensembles over the historical period 1979-2005, for three different illustrative regions (the Alps, Iberian Peninsula and Eastern Europe). The reduction of biases for DeepESD is noticeable along the entire distribution (including the extremes) for both precipitation and temperature~~as shown in Figure 3.~~. Note that for precipitation these results are due to the use of the stochastic nature of the method, sampling from the inferred conditional distributions.

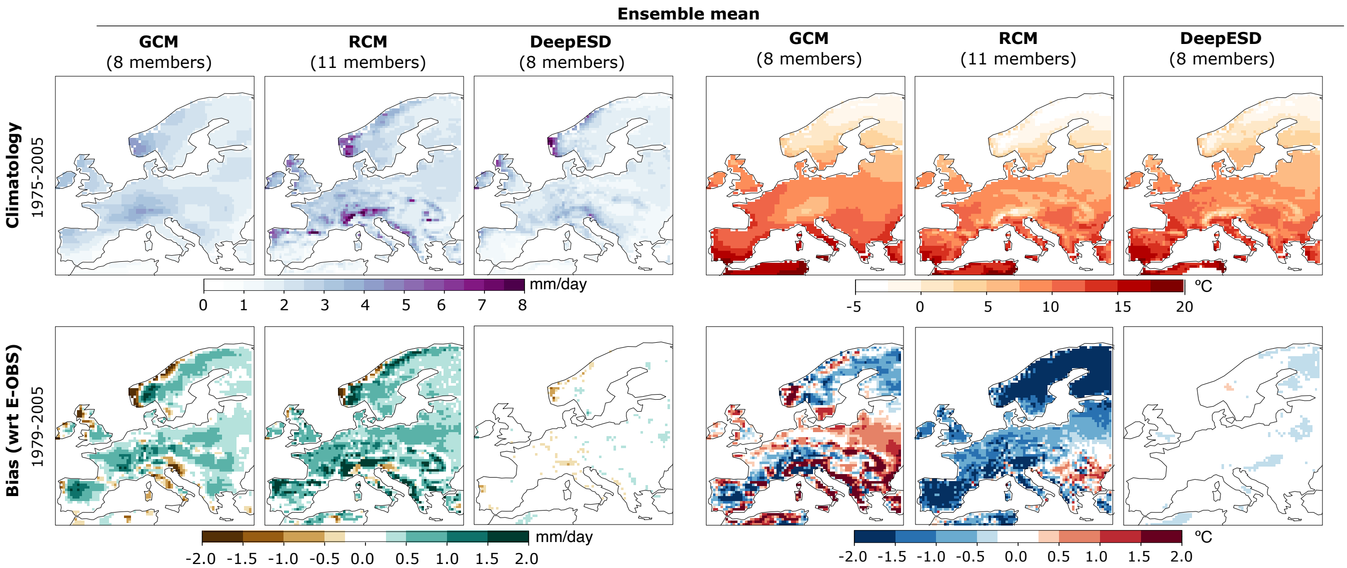

Figure 2. Annual daily precipitation (left block) and temperature (right) for the historical period 1975-2005, as obtained from the ensembles of GCMs, RCMs and DeepESD GCM-downscaled results (left, middle and right columns, respectively). The first row shows the ensemble mean climatological values and the second row displays the corresponding biases with respect to E-OBS v20.

Figure 4 shows the mean climate change signal resulting from the GCM, RCM and DeepESD ensembles, as well as the underlying uncertainty (characterized by multi-model dispersion). In particular, the right (left) panel in this figure shows the values for precipitation (temperature) for near, mid and far future periods (rows 1-3) relative to 1975-2005, as projected by the GCM, RCM and DeepESD ensembles (in columns).

Overall, the spatial pattern of future precipitation changes is similar for the three ensembles, with precipitation decreasing over Southern Europe and increasing over the Northern part of the continent. Slight regional differences exist among the three ensembles, with DeepESD presenting weaker (decreasing) signals of change over the Iberian Peninsula but stronger (increasing) ones over some parts of Northern and Eastern Europe, especially when compared with GCMs. Interestingly, both the climate change signal and the multi-model uncertainty spatial patterns of DeepESD are more similar to the downscaled

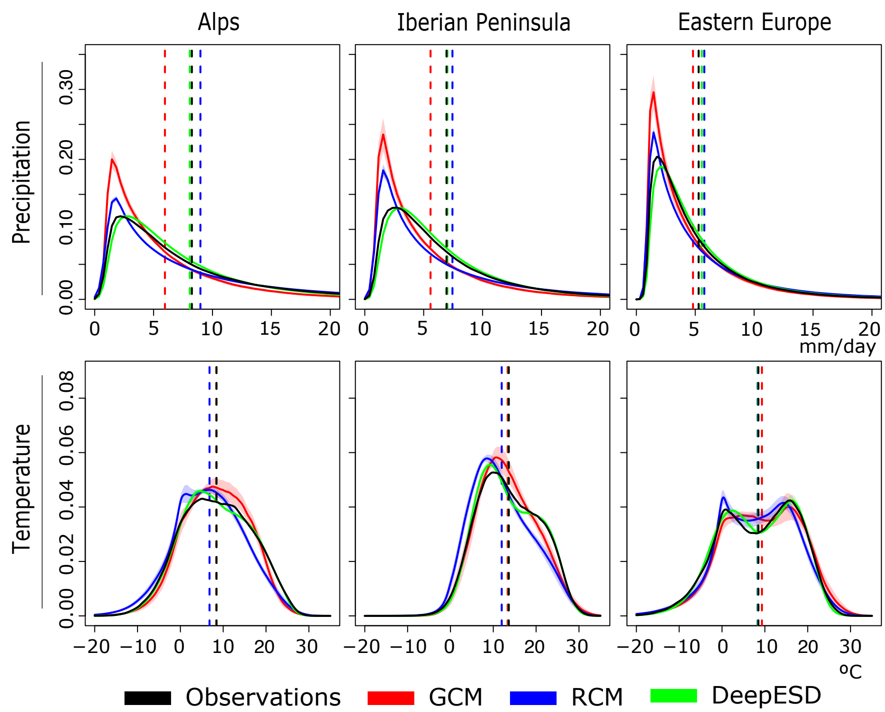

Figure 3. Probability Density Functions (PDFs) of the GCM (red), RCM (blue) and DeepESD (green) ensembles of precipitation and temperature over the historical period 1979-2005, plus E-OBS (black) for the Alps, the Iberian Peninsula and Eastern Europe as defined in the PRUDENCE regions (Christensen and Christensen, 2007). The solid line represents the ensemble mean and the shadow encompass two standard deviations. The dashed line indicates the distributional mean of each PDF.

RCM than to the GCM ensemble. Moreover, DeepESD projects lower uncertainty than both physical-based ensembles across most of the European continent.

Regarding temperature, the spatial patterns are broadly consistent among the three ensembles, with the highest warming located over Northern Scandinavia, Eastern Europe and the Mediterranean basin, and the lowest one for the British Isles and ~~western and central~~ Western and Central Europe. As in the case of precipitation, ~~the spatial patterns of the climate change signal are similar for RCM and DeepESD (as compared with GCM results), with larger differences over Iberia~~ some regional differences exist among ensembles, especially over Central and Eastern Europe ~~. However, the uncertainty (ensemble spread) is higher in DeepESD than in the GCM and RCM ensembles in most regions. In addition, both RCM and CNN downscaling reduce~~ where both RCMs and DeepESD project lower signals of change than the GCMs, reducing the warming signal by about 0.5~~-1.5~~-1°C ~~over most of Europe, as compared with the global GCM signal, especially for the far-future.~~ by the end of the century. Finally, the GCMs' ensemble spread range between 0.5-1.5°C with higher values in Southern and especially Northern Europe than in the rest of the continent. The RCMs (DeepESD) project a similar spatial pattern than the GCMs with a lower spread over Central and Eastern Europe (Scandinavia).

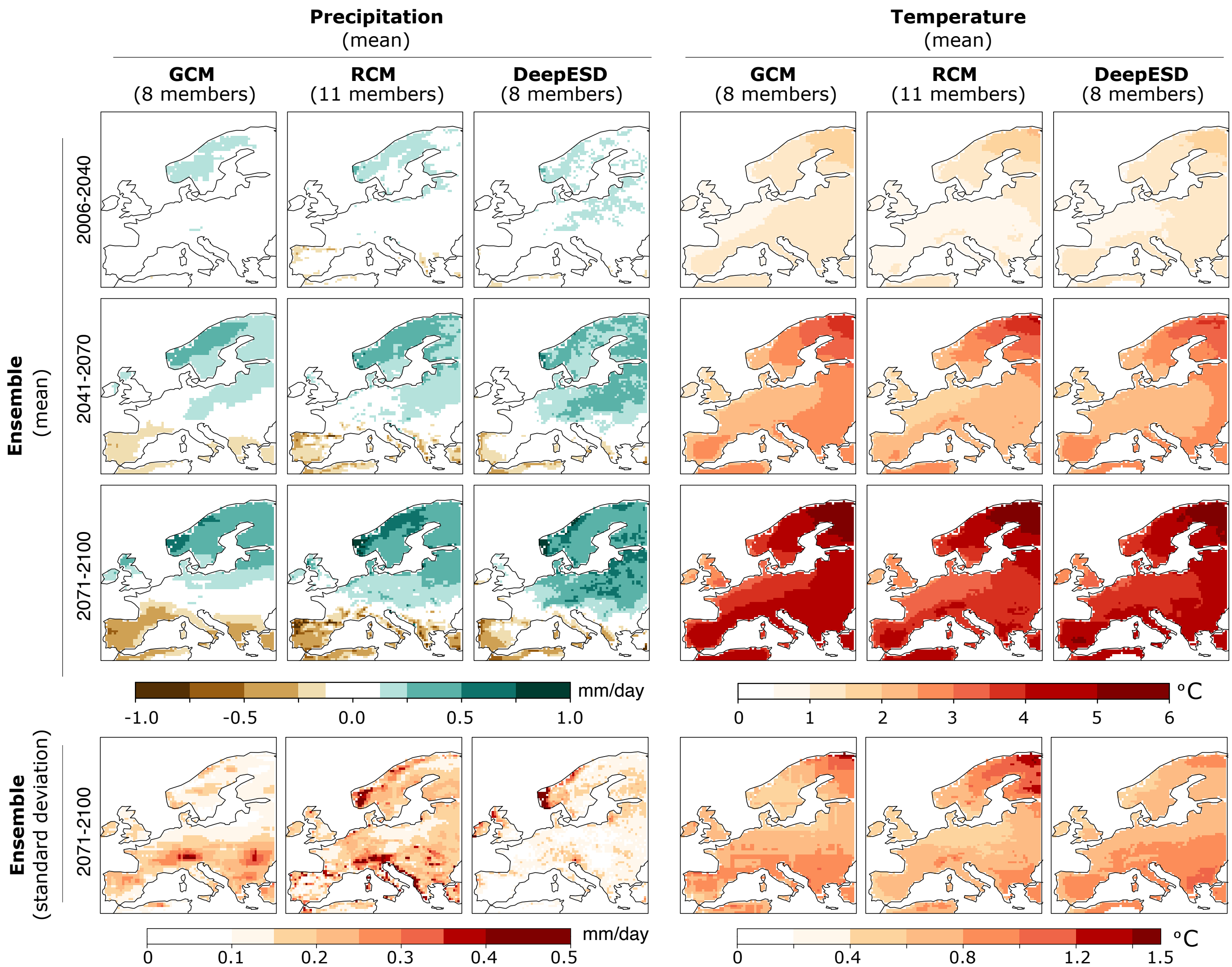

Figure 4. Climate change signal for annual mean precipitation (left) and temperature (right) for near- (2006-2040), mid- (2041-2070) and far-future (2071-2100) periods, in rows, relative to 1975-2005 as projected by the GCM, RCM and DeepESD ensembles (in columns). The last row shows the uncertainty of the far-future signal, as measured by the standard deviation of the results across models.

Further research is needed to assess whether ~~these differences~~ the differences between GCM and RCM/DeepESD signals are due to an added value of downscaling or to deficiencies in the models. In the case of the RCMs, some recent studies attribute ~~these differences~~ them to the lack of time-varying anthropogenic aerosols in the RCM formulation (Boé et al., 2020; Gutiérrez et al., 2020). ~~For the DeepESD dowscaled results, this issue is yet to be explored.~~ To further analyze ~~this~~the results for DeepESD, Figure 5 shows ~~the differences (~~(in columns) the climate change signals (2071-2100 with respect to 1975-2005) of the eight CMIP5 climate models considered and the corresponding DeepESD downscaled fields for precipitation (rows 1-2) and temperature (rows 4-5). Rows 3 and 6 show the differences between the DeepESD downscaled and raw climate change

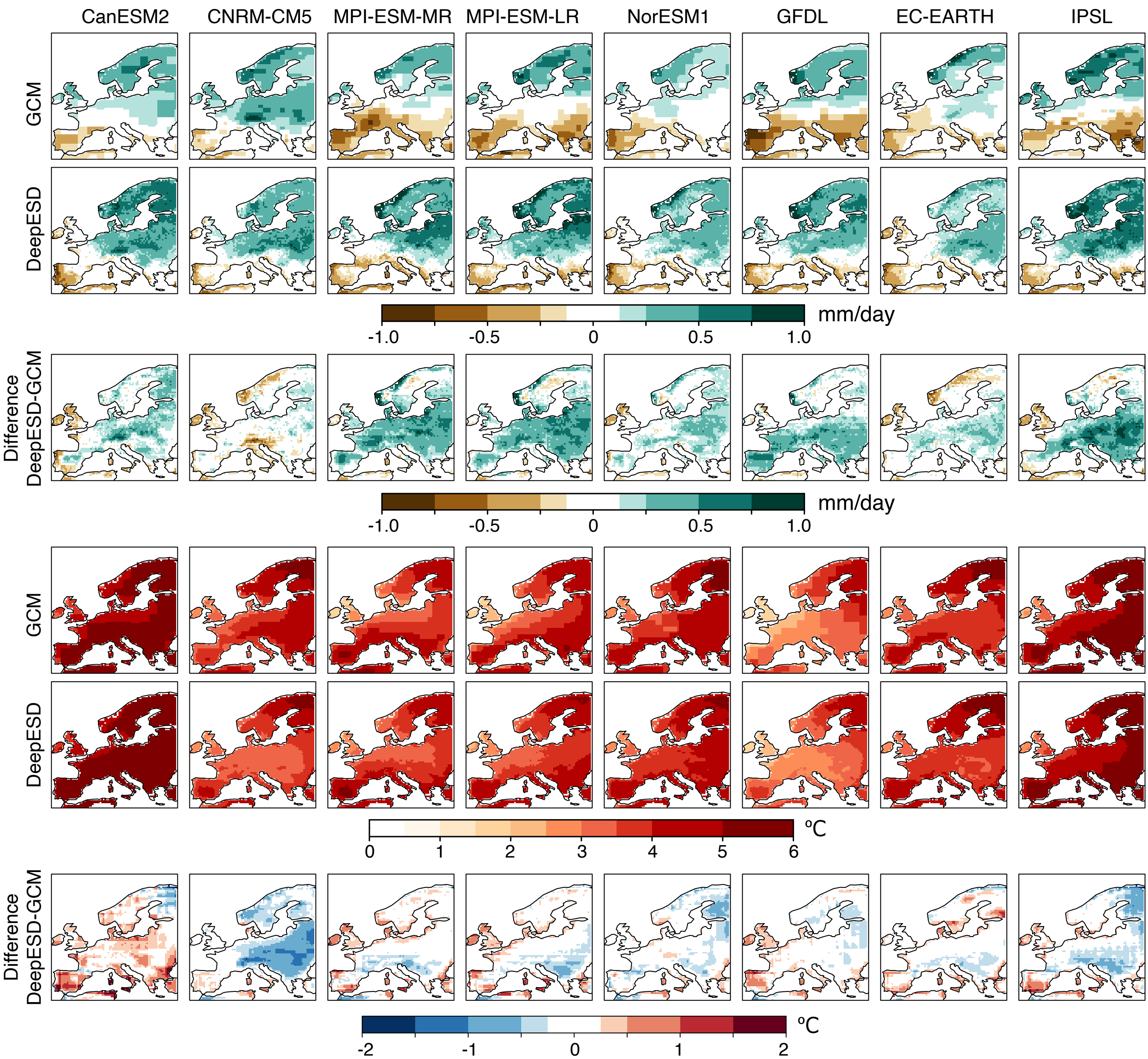

Figure 5. ~~Differences between DeepESD and raw GCM~~ The climate change signals (~~delta) of the mean precipitation and temperature (in rows) for the far future (~~2071-2100 ~~), relative~~ with respect to 1975-2005~~, for~~) of the eight ~~GCMs~~ CMIP5 climate models considered and the equivalent DeepESD downscaled fields for precipitation (~~in columns~~rows 1-2) ~~. For~~ and temperature ~~, "centered" fields are also displayed by substracting the spatial mean of each panel indicated in the upper left corner~~ (~~row 3~~rows 4-5). ~~The averaged~~ Rows 3 and 6 show the difference between the DeepESD downscaled and raw climate change signals ~~are also indicated in~~ for the ~~same corner.~~ different GCMs

signals for the different GCMs~~(in columns). These differences are quite systematic for the case of precipitationindicating a robust CNN extrapolation fingerprint. On the contrary, the differences exhibit different magnitudes and patterns as a function of the GCM , alternating from mostly positive (CanESM2 and GFDL), mostly negative (MPI-ESM-LR/MR, NorESM1~~. As per the climate change signal of precipitation, ~~EC-Earth) or a combination of both (~~we observe a South-to-North gradient with different intensities depending on the GCM (i.e., MPI, GFDL and IPSL present the lowest values over the Mediterranean basin). Differently to the GCMs signal, DeepESD provide a more homogeneous spatial pattern explaining the low inter-model spread of Figure 4. As per temperature, all GCMs project a positive climate change signal with subtle spatial patterns which vary across GCMs that are well captured by the DeepESD downscaled fields. This similarity in the climate change signals between the GCM and DeepESD fields explain the similar inter-model spread of Figure 4. Also, we observe that the CNRM-CM5 ~~, IPSL). Nevertheless, the spatial patterns present a similar structure (a robust CNN fingerprint) when substracting their spatial means (row 3 in Figure 5), with lower (higher) values in Eastern Europe (Scandinavia and the Iberian Peninsula). This shows that the CNN extrapolation yields different scaling factors for different GCMs but a similar patter of regional changes (when compared to the GCMs). Moreover, as shown by the numbers included in the figures, these scaling factors are not related to the magnitude of the GCM warming signal and thus, they contribute increasing the uncertainty of the ensemble. Further analysis of the ensemble to detect potential problems and reduce uncertainty would require a detailed analysis of the signal-carrying predictors for the different GCMs and their effect in the relationships learnt in present-day conditions using reanalysis data, but this is beyond the scope of this paper~~is the one model responsible for the reduced warming signal over Eastern Europe described in Figure 4.

To examine the behaviour of CNNs beyond climatological fields, Figure 6 shows the yearly time-series for precipitation and temperature averaged over Eastern the Alps, the Iberian Peninsula and Eastern Europe domains, as defined in the PRUDENCE regions (Christensen and Christensen, 2007), which are broadly representative of the different European climate regimes —mountainous, mediterranean and continental, respectively.— Namely, we focus on the frequency of rainy days (R01), i.e. those receiving at least 1mm of rain, the average precipitation in rainy days (SDII) and the mean of temperature. For every indicator, the ensemble of GCMs (red), RCMs (blue) and DeepESD (yellow) for the total period 1975-2100, plus the observational reference, E-OBS (black), for the period 1979-2008, are shown. In all cases, the solid lines represent the multi-model ensemble mean whilst the shadows encompass all the models contributing to the ensemble.

Figure 6.a shows that both GCMs and RCMs overestimate the frequency of wet days with respect to the observational reference —consequence of the drizzle effect (Dai, 2006).— For the SDII, RCMs present mostly unbiased fields whilst GCMs underestimate this metric across all regions, remarking the added value of RCMs to reproduce regional precipitation. In contrast to GCMs and RCMs, DeepESD provides in general more robust estimates for both R01 and SDII under the historical scenario. In terms of future changes, the three ensembles project an increase in in the SDII across all regions, and a decrease (increase) of the number of wet days in the Southern (Northern) regions consistent with the results of Figure 2.

For temperature, Figure 6.c shows that the three ensembles perform similarly across all regions, with some systematic underestimation of mean temperatures by the RCMs, and DeepESD exhibiting nearly unbiased results under the historical scenario. Note that the GCMs time-series are mostly unbiased which is the result of averaging out the positive and negative

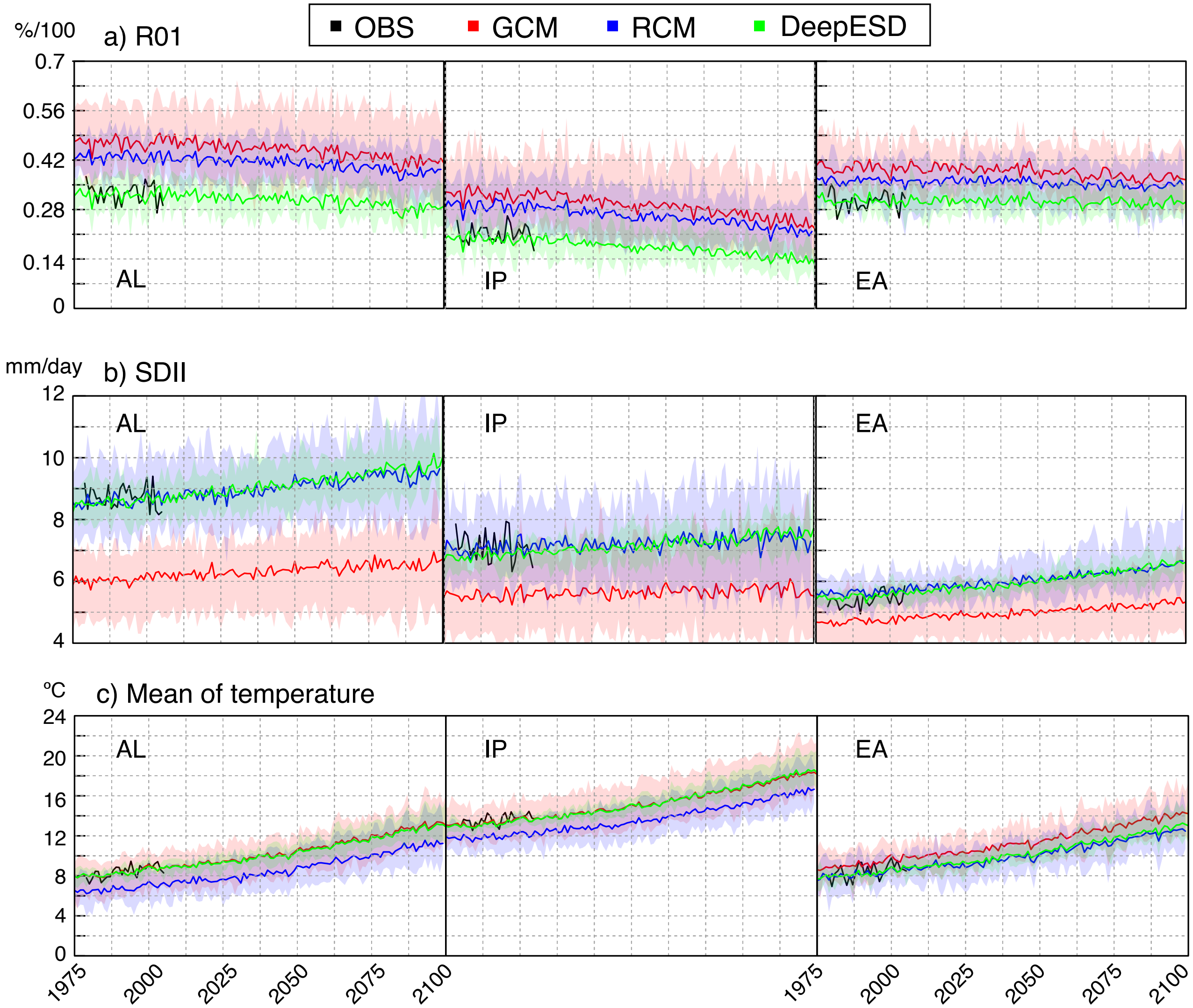

Figure 6. Annual time-series for R01, SDII and the mean of temperature, averaged over the eight PRUDENCE regions. For every indicator, the ensemble of GCMs (red), RCMs (blue) and DeepESD (yellow) for the total period 1975-2100, plus the observational reference, E-OBS (black), for the period 1979-2008, are shown. In all cases, the solid lines represent the multi-model ensemble mean whilst the shadows encompass all the models contributing to the ensemble.

biases appearing in the spatial fields of Figure 2. As per the projected signals of change, the three ensembles point out to a (quasi) linear increase along the century and across all regions, with warming values of about 4-6°C for the far-future in most of cases.

This indicates that DeepESD is able to accurately reproduce the historical climate —even the discrete-continuous nature of precipitation,— and beyond the regional differences showed in Figure 5, there is a synchrony in the temporal evolution of the signals among ensembles. These results also indicate that DeepESD results in a smaller spread of the ensemble due to the adjustment of the models towards the observed climatology.

4 Conclusions

Deep learning topologies are increasingly being tested for downscaling purposes, achieving promising results in present climate due to their ability to infer complex non-linear patterns from climate data. Nevertheless, the ability of these models to generalize to out-of-sample climate change conditions is still to be analyzed with many questions open. Here, we present DeepESD, an ensemble of regional precipitation and temperature projections (up to 2100) over Europe produced by applying convolutional neural networks to downscale a set of eight GCMs over the EUR-44 CORDEX domain. This multi-model perspective permits to analyze unexplored aspects of CNN-based downscaling such as the inter-model uncertainty of the climate change signals or the similarities/differences of the downscaling across GCMs. We build on existing CNNs models (Baño-Medina et al., 2020) and focus on their performance in the climate model space, using GCM projections. In this sense, we follow previous literature (Vrac et al., 2007; San-Martín et al., 2017; Quesada-Chacón et al., 2021) and compare the DeepESD future fields with a set of state-of-the-art CORDEX RCMs, which are used as “pseudo-observations". To our knowledge, this is the first time that CNNs ~~are~~ have been used to produce downscaled multi-model ensembles ~~of downscaled projections~~ based on the perfect-prognosis approach and are compared against an ensemble of RCMs.

We find that CNN-based downscaling is able to reproduce the observed climate over the historical period for both precipitation and temperature fields at a distributional level, reducing the systematic biases exhibited by the global and regional physical models. When analyzing the future climate change signals we find that DeepESD presents spatial patterns and magnitudes which are broadly similar to the ones from the RCMs. Nevertheless, there are ~~considerable~~ regional differences —at a climatological, ~~inter-annual and seasonal~~ and inter-annual scales,— in the projected climate change signals among DeepESD and the physical-based models. For the case of precipitation, these differences are driven towards a decrease of the multi-model uncertainty with respect to the one of their driving GCMs. As per temperature, the CNNs ~~learns regional downscaled pattern which scales differently for each GCM for climate change conditions introducing uncertainty in the future climate change signal. This problem~~ project similar signals of change than the GCMs being able to capture the particularities of each one resulting into a similar ensemble spread. This property was not perceived in previous studies (Baño-Medina et al., 2021) where a single GCM (i.e., EC-Earth) was considered.

Despite the analysis presented herein, the plausibility of the projections has to be further analyzed prior to the integration of DL topologies into climate change applications. For instance, this can be done by developing specific studies dealing with the domain adaptation of the statistical models learned in “perfect" conditions to climate model spaces; by conducting synthetic case-studies permitting to analyze their extrapolation capabilities to climate change conditions; and by comparing the CNN-based fields against other machine learning techniques. To this aim, following the FAIR principles we make publicly available

DeepESD from the ESGF portal, which will allow the scientific community to continue exploring the benefits and shortcomings of these new techniques for the downscaling of climate. Precisely, DeepESD contributes to CORDEX EUR-44 being the first statistical-based dataset to ever participate in this international initiative, entailing a breakthrough of this type of techniques on the study of regional climate.

Code and data availability. To promote transparency and reproducibility of our results, we provide the data (DOI: 10.5281/zenodo.6823421) and the companion Jupyter notebook (DOI: 10.5281/zenodo.6828303), explaining how DeepESD has been produced. This notebook is based on the `R` software and builds on the `climate4R` framework, a set of libraries specifically designed for climate data access and post-processing (Iturbide et al., 2019). To build the CNNs used, we rely on `downscaleR.keras` ((Baño-Medina et al., 2020)), which integrates `Keras`, a state-of-the-art DL library, within `climate4R`. Furthermore, most of the results shown in this manuscript can be replicated by following the indications given in the notebook, providing thus the basis for practitioners to perform their own experiments.

DeepESD downscaled results have been published at the ESGF data node at Cantabria University https://data.meteo.unican.es/thredds/catalog/esgcet/catalog.html.

Author contributions. J.B., R.M. and J.M.G. conceived the experiment, J.B. produced all the results, E.C. and A.S.C. prepared the data for publication. All authors contributed to the analysis of results and to the writing of the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors would like to acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). We also acknowledge the support from the Spanish Government through project PID2020-116595RB-I00: Contribución a la nueva generación de proyecciones climáticas regionales de CORDEX mediante técnicas dinámicas y estadísticas (CORDyS) funded by MCIN/AEI /10.13039/501100011033. Also, J.B-M acknowledges support from Universidad de Cantabria and Consejería de Universidades, Igualdad, Cultura y Deporte del Gobierno de Cantabria via the “instrumentación y ciencia de datos para sondear la naturaleza del universo” project.

References

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Climate Dynamics, pp. 1–11, 2021.

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O. E., and Frei, C.: 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, 727–747, 2022.

Baño-Medina, J.: Understanding Deep Learning Decisions in Statistical Downscaling Models, in: Proceedings of the 10th International Conference on Climate Informatics, pp. 79–85, 2020.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, Geoscientific Model Development, 13, 2109–2124, 2020.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, , Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, 2013.

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kållberg, P., Kobayashi, S., Uppala, S., and Simmons, A.: The ERA-Interim archive Version 2.0, Shinfield Park, Reading, 1, 23, 2011.

Boé, J., Somot, S., Corre, L., and Nabat, P.: Large discrepancies in summer climate change over Europe as projected by global and regional climate models: causes and consequences, Climate Dynamics, 54, 2981–3002, 2020.

Brands, S., Herrera, S., San-Martín, D., and Gutiérrez, J. M.: Validation of the ENSEMBLES global climate models over southwestern Europe using probability density functions, from a downscaling perspective, Climate Research, 48, 145–161, 2011.

Brands, S., Herrera, S., Fernández, J., and Gutiérrez, J. M.: How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa?, Climate dynamics, 41, 803–817, 2013.

Cannon, A. J.: Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network, Journal of Hydrometeorology, 9, 1284–1300, 2008.

Casanueva, A., Herrera, S., Fernández, J., and Gutiérrez, J. M.: Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative, Climatic Change, 137, 411–426, 2016.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic change, 81, 7–30, 2007.

Christian, J., Arora, V., Boer, G., Curry, C., Zahariev, K., Denman, K., Flato, G., Lee, W., Merryfield, W., Roulet, N., and Scinocca, J.: The global carbon cycle in the Canadian Earth system model (CanESM1): Preindustrial control simulation, Journal of Geophysical Research (Biogeosciences), 115, 2010.

Cornes, R. C., Schrier, G. v. d., Besselaar, E. J. M. v. d., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, 2018.

Dai, A.: Precipitation characteristics in eighteen coupled climate models, Journal of climate, 19, 4605–4630, 2006.

Doblas Reyes, F., Acosta Navarro, J. C., Acosta Cobos, M. C., Bellprat, O., Bilbao, R., Castrillo Melguizo, M., Fuckar, N., Guemas, V., Lledó Ponsati, L., Menegoz, M., et al.: Using EC-Earth for climate prediction research, ECMWF Newsletter, pp. 35–40, 2018.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, 2013.

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics, Journal of Climate, 26, 2247–2267, 2013.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Fernández, J., Frías, M., Cabos, W., Cofiño, A., Domínguez, M., Fita, L., Gaertner, M., García-Díez, M., Gutiérrez, J. M., Jiménez-Guerrero, P., et al.: Consistency of climate change projections from multiple global and regional model intercomparison projects, Climate dynamics, 52, 1139–1156, 2019.

François, B., Thao, S., and Vrac, M.: Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks, Climate Dynamics, 57, 3323–3353, 2021.

Giorgi, F.: Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?, Journal of Geophysical Research: Atmospheres, 124, 5696–5723, 2019.

Gutiérrez, C., Somot, S., Nabat, P., Mallet, M., Corre, L., van Meijgaard, E., Perpiñán, O., and Gaertner, M. Á.: Future evolution of surface solar radiation and photovoltaic potential in Europe: investigating the role of aerosols, Environmental Research Letters, 15, 034 035, 2020.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., Martín, D. S., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M. M., Czernecki, B., and Pagé, C.: 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, 3750–3785, 2019.

Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., Arritt, R. W., Bukovsky, M. S., Calvin, K. V., Feng, Z., Jones, A. D., Kooperman, G. J., Monier, E., Pritchard, M. S., Pryor, S. C., Qian, Y., Rhoades, A. M., Roberts, A. F., Sakaguchi, K., Urban, N., and Zarzycki, C.: The Ongoing Need for High-Resolution Regional Climate Models: Process Understanding and Stakeholder Information, Bulletin of the American Meteorological Society, 101, E664 – E683, 2020.

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M. D., Manzanas, R., San-Martín, D., Cimadevilla, E., Cofiño, A. S., and Gutiérrez, J. M.: The R-based climate4R open framework for reproducible climate data access and post-processing, Environmental Modelling & Software, 111, 42–54, 2019.

Klok, E. and Klein Tank, A.: Updated and extended European dataset of daily climate observations, International Journal of Climatology: A Journal of the Royal Meteorological Society, 29, 1182–1191, 2009.

Le Roux, R., Katurji, M., Zawar-Reza, P., Quénol, H., and Sturman, A.: Comparison of statistical and dynamical downscaling results from the WRF model, Environmental modelling & software, 100, 67–73, 2018.

LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks, 3361, 1995, 1995.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press, 2018.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1–14, 2015.

Müller, W., Jungclaus, J., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., and Marotzke, J.: A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM 1.2-HR), Journal of Advances in Modeling Earth Systems, 10, 2018.

Olmo, M. E., Balmaceda-Huarte, R., and Bettolli, M. L.: Multi-model ensemble of statistically downscaled GCMs over southeastern South America: historical evaluation and future projections of daily precipitation with focus on extremes, Climate Dynamics, pp. 1–18, 2022.

Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S.: Improving Precipitation Estimation Using Convolutional Neural Network, Water Resources Research, 55, 2301–2321, 2019.

Quesada-Chacón, D., Barfus, K., and Bernhofer, C.: Climate change projections and extremes for Costa Rica using tailored predictors from CORDEX model output through statistical downscaling with artificial neural networks, International Journal of Climatology, 41, 211–232, 2021.

San-Martín, D., Manzanas, R., Brands, S., Herrera, S., and Gutiérrez, J. M.: Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods, Journal of Climate, 30, 203–223, 2017.

Schoof, J. T. and Pryor, S. C.: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, International Journal of Climatology, 21, 773–790, 2001.

Sun, L. and Lan, Y.: Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods, International Journal of Climatology, 41, 1128–1147, 2021.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American meteorological Society, 93, 485–498, 2012.

Vautard, R., Kadygrov, N., Iles, C., Boberg, F., Buonomo, E., Bülow, K., Coppola, E., Corre, L., van Meijgaard, E., Nogherotto, R., et al.: Evaluation of the large EURO-CORDEX regional climate model ensemble, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 344, 2021.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091–2121, 2013.

Vrac, M. and Ayar, P.: Influence of Bias Correcting Predictors on Statistical Downscaling Models, Journal of Applied Meteorology and Climatology, 56, 2016.

Vrac, M., Stein, M., Hayhoe, K., and Liang, X.-Z.: A general method for validating statistical downscaling methods under future climate change, Geophysical Research Letters, 34, 2007.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific data, 3, 1–9, 2016.

Williams, P. M.: Modelling Seasonality and Trends in Daily Rainfall Data, in: Advances in Neural Information Processing Systems 10, edited by Jordan, M. I., Kearns, M. J., and Solla, S. A., pp. 985–991, 1998.