



# Downscaling multi-model climate projection ensembles with deep learning (DeepESD): contribution to CORDEX EUR-44

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

<sup>1</sup>Instituto de Física de Cantabria (IFCA), CSIC–Universidad de Cantabria, Santander, Spain

<sup>2</sup>Departamento de Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, Santander, Spain

<sup>3</sup>Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain

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

Received: 28 February 2022 – Discussion started: 22 March 2022

Revised: 13 July 2022 – Accepted: 6 August 2022 – Published: 6 September 2022

**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 (CNNs) 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 global circulation models (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 based on the perfect-prognosis approach, allowing us 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 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 Fed-

eration (ESGF), as the first continental-wide PP dataset contributing to CORDEX (EUR-44).

## 1 Introduction

The Coupled Model Intercomparison Project (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 100 km, 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 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>, last access: 26 August 2022) 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 10 km, 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 in a historical period 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 predictor values (from present climate or future scenarios) to obtain the regional/local downscaled results. 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, 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 results at continental scales, e.g., 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 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 ( $0.5^\circ$  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 Zenodo (see the section on code and data availability).

## 2 Data and methods

Following the PP approach, the CNN models have been trained over the period 1979–2005 using daily predictors from the ERA-Interim reanalysis (Dee et al., 2011), upscaled from its original  $0.75^\circ$  resolution to a reference  $2^\circ$  regular grid, and predictands from E-OBS v20 (Cornes et al., 2018), originally at  $0.25^\circ$  but upscaled to  $0.5^\circ$  for consistency with previous works (Baño-Medina et al., 2020, 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, 2021). 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, and geopotential, meridional and zonal wind velocity at 500, 700 and 850 hPa (i.e., a total of 15 variables per grid box) have been used as predictors covering the domain  $34\text{--}76^\circ\text{N}$ ,  $8^\circ\text{W}\text{--}34^\circ\text{E}$ , resulting in a  $22 \times 22 \times 15$  (longitude  $\times$  latitude  $\times$  variable) high-dimensional input grid. To avoid potential artifacts derived from the different scale of the distinct variables, ERA-Interim predictors are standardized at the grid box level (Baño-Medina et al., 2021).

For downscaling we used an ensemble formed by the eight CMIP5 GCMs described in Table 1, whose ability to repro-

duce the large-scale dynamics has already been assessed for PP studies (Brands et al., 2013), which 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 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., 2022) 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 have been interpolated to the reference  $2^\circ$  grid (considering the nearest grid box) to match the predictor space used for ERA-Interim. 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. 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; 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 grid box level for their use in the CNN (the same standardization parameters used for the reanalysis data are applied here).

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

For the CNN models used in this work, we deploy the best-performing topologies developed in Baño-Medina et al. (2020), a recent study which intercompares different CNNs over Europe to downscale temperature (precipitation). They consist of three convolutional layers (LeCun and Bengio, 1995) with 50, 25 and 10 (1) spatial kernels ( $3 \times 3$  grid boxes) followed by a dense connection linking the last hidden layer to the output neurons (corresponding to the land grid points in E-OBS). 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 grid box using two (three) output neurons corresponding 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 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 at 2.60 GHz 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 6 h, 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 using nine nodes with 144 cores). 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, we interpolate these RCMs from their original spatial resolution ( $0.44^\circ$ ) to the predictand  $0.5^\circ$  regular grid.

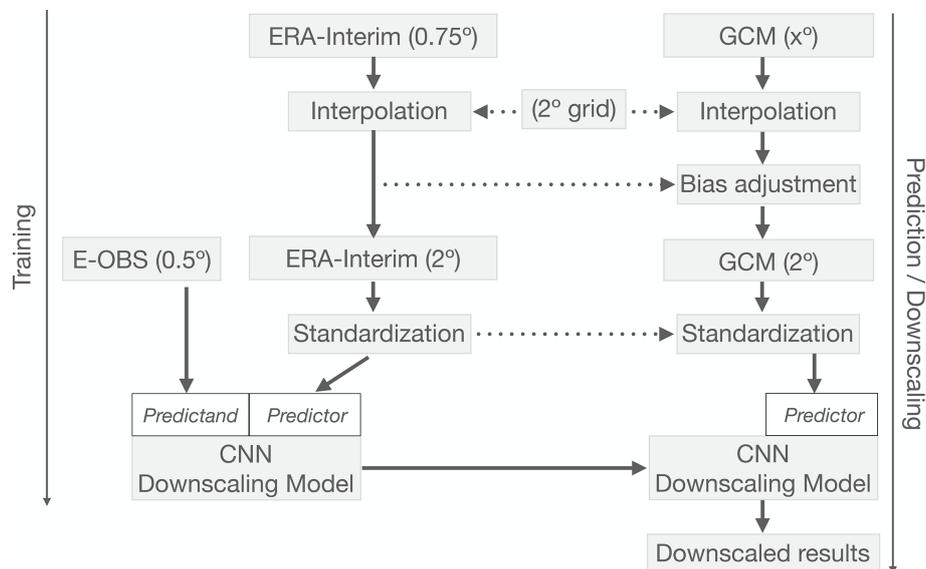
Finally, we test the sensitivity of CNN training on the results by repeating the downscaling experiment 10 times and evaluate historical biases and future climate change signals as shown below without finding appreciable variations.

### 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. For precipitation, the raw

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

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°)

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

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 as a consequence of the training process (Casanueva et al., 2016). Besides these results for the mean, Fig. 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, the Iberian Peninsula and Eastern Europe, as defined in the PRUDENCE regions, Christensen and Christensen, 2007). The reduction of biases for DeepESD is noticeable along the entire distribution (in-

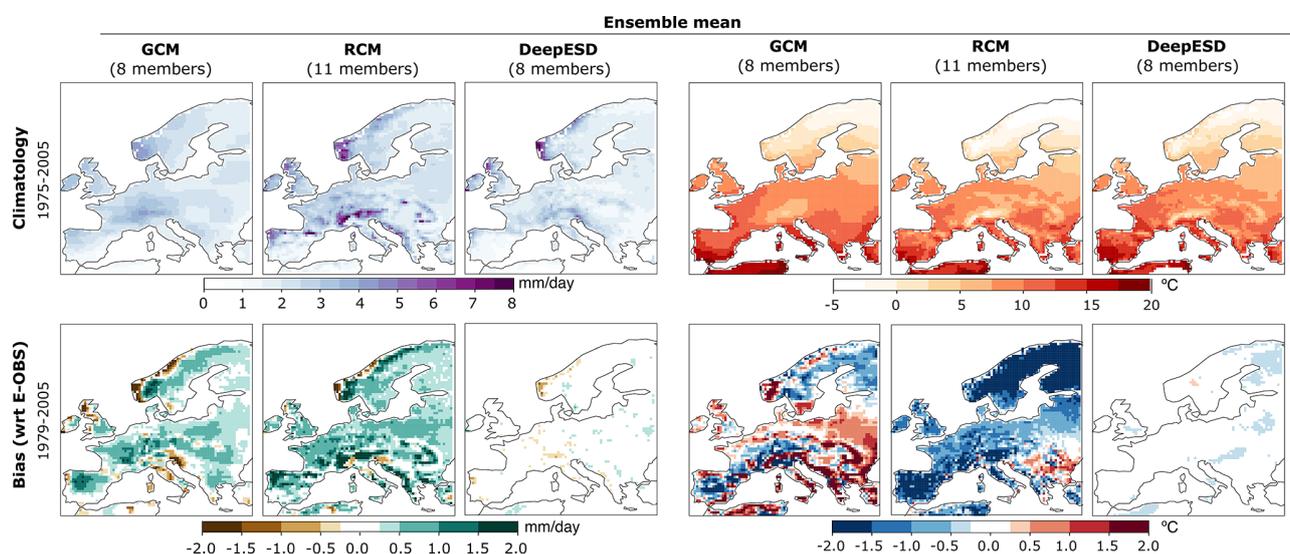
cluding the extremes) for both precipitation and temperature. Note that for precipitation these results are due to the use of the stochastic nature of the method, sampling from the inferred conditional distributions.

Figure 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

**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.

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



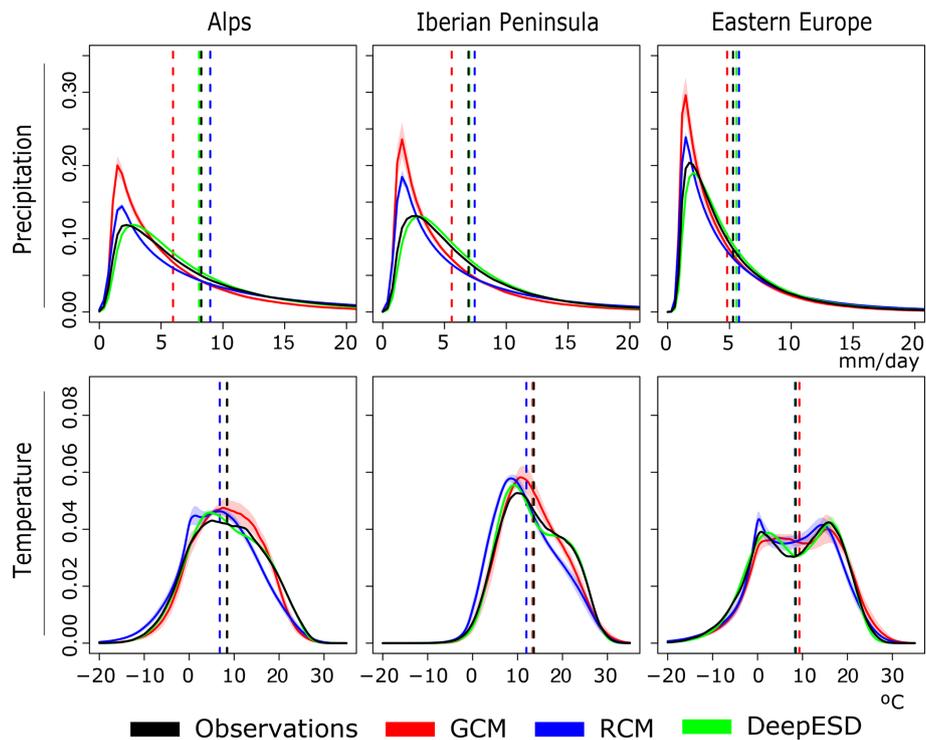
**Figure 2.** Annual daily precipitation (left block) and temperature (right) for the historical period 1975–2005, as obtained from the ensembles of GCMs, RCMs and DeepESD GCM-downscaled results (left, middle and right columns, respectively). The first row shows the ensemble mean climatological values,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

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 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 Europe. As in the case of precipitation, some regional differences exist among ensembles, especially over central and Eastern Europe where both RCMs and DeepESD project lower signals of change than the GCMs, reducing the warming signal by about 0.5–

1 °C by the end of the century. Finally, the GCMs’ ensemble spread ranges 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).

Further research is needed to assess whether 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 them to the lack of time-varying anthropogenic aerosols in the RCM formulation (Boé et al., 2020; Gutiérrez et al., 2020). To further analyze the results for DeepESD, Fig. 5 shows (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



**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, as well as 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 encompasses 2 standard deviations. The dashed line indicates the distributional mean of each PDF.

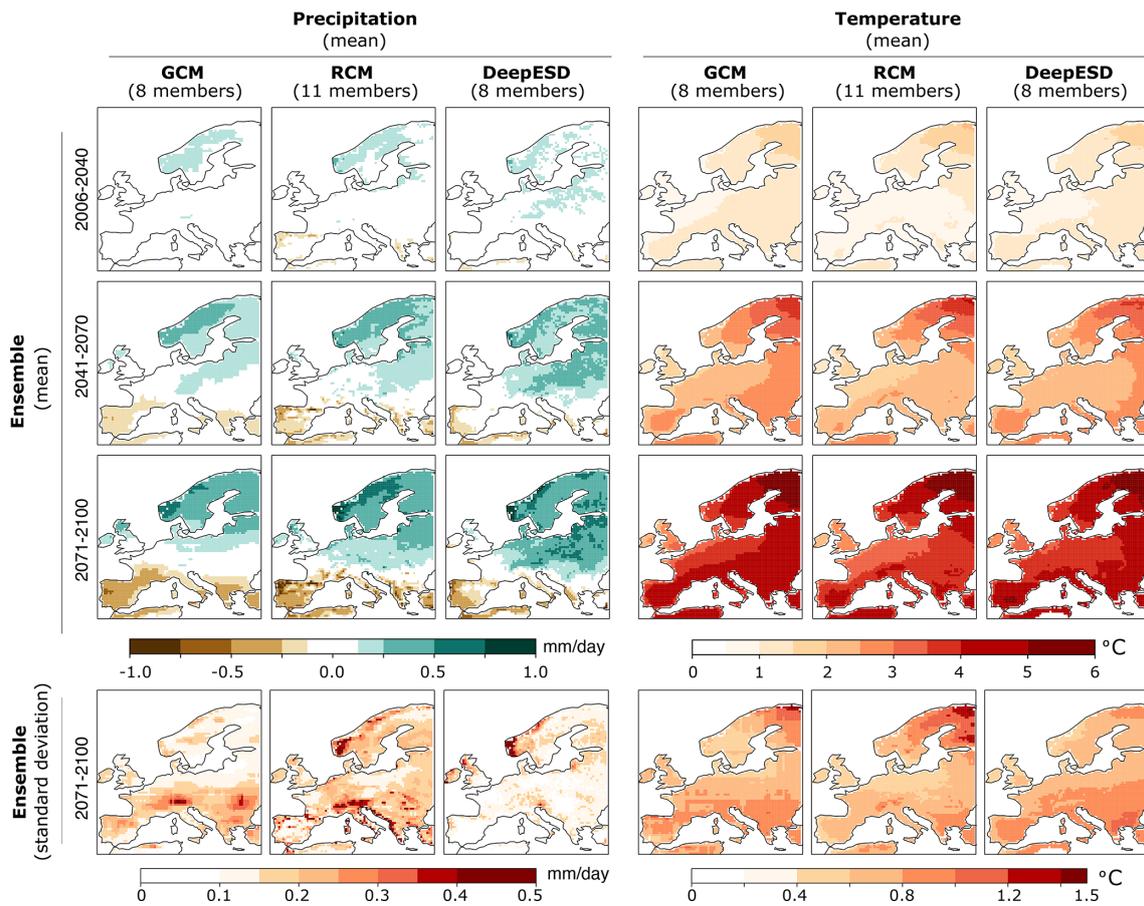
fields for precipitation (rows 1–2) and temperature (rows 4–5). Rows 3 and 6 show the differences between the DeepESD downscaled and raw climate change signals for the different GCMs. As per the climate change signal of precipitation, we observe a south-to-north gradient with different intensities depending on the GCM (i.e., MPI, GFDL and IPSL present the lowest values over the Mediterranean Basin). Differently to the GCMs signal, DeepESD provides a more homogeneous spatial pattern explaining the low inter-model spread of Fig. 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 explains the similar inter-model spread of Fig. 4. Also, we observe that the CNRM-CM5 is the one model responsible for the reduced warming signal over Eastern Europe described in Fig. 4.

To examine the behavior of CNNs beyond climatological fields, Fig. 6 shows the yearly time series for precipitation and temperature averaged over 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 1 mm 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 (green) for the total period 1975–2100 and 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 6a shows that both GCMs and RCMs overestimate the frequency of wet days with respect to the observational reference – a 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 the SDII across all regions, as well as a decrease (increase) of the number of wet days in the southern (northern) regions consistent with the results of Fig. 2.

For temperature, Fig. 6c 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 his-



**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.

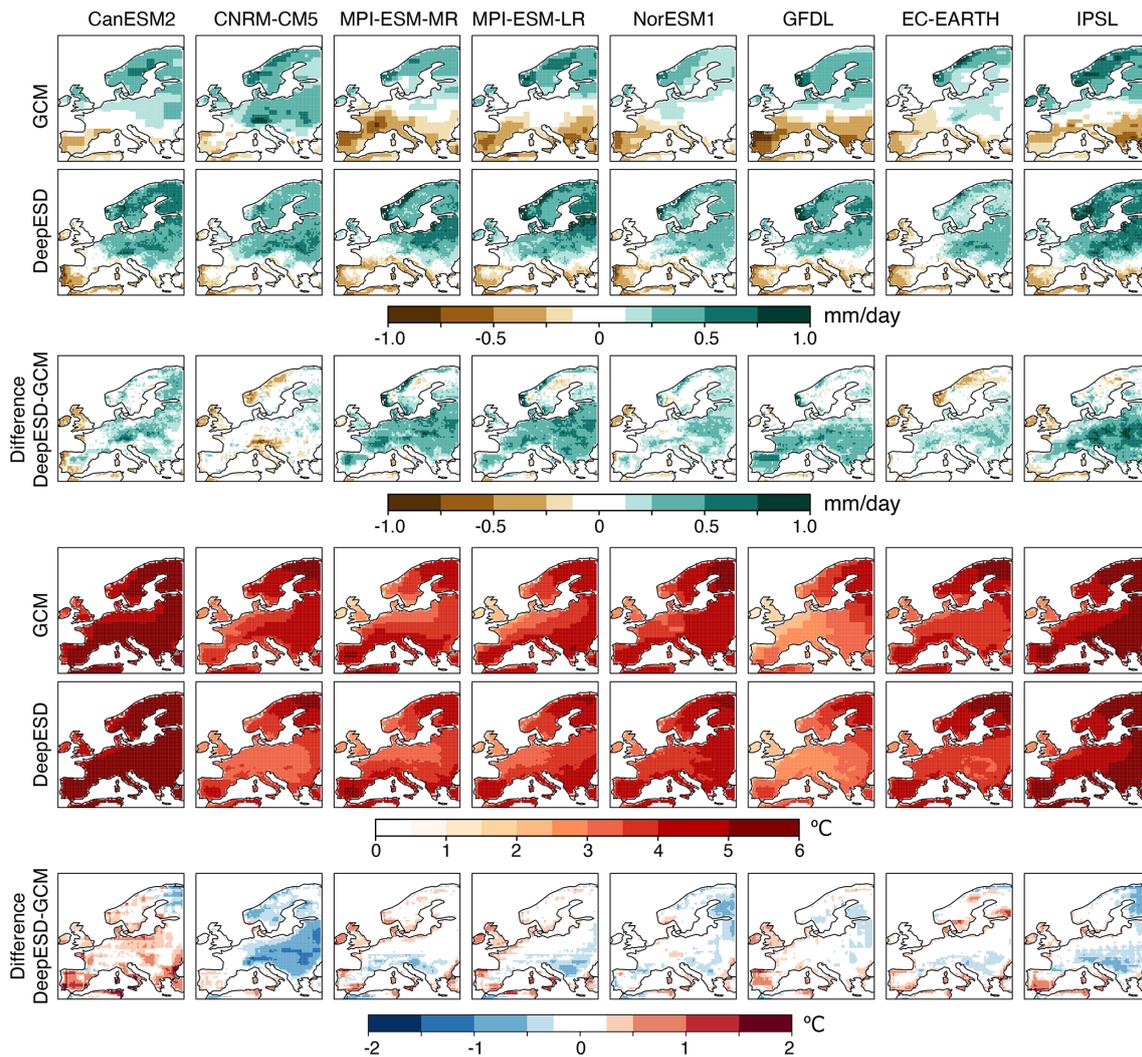
torical scenario. Note that the GCMs time series are mostly unbiased, which is the result of averaging out the positive and negative biases appearing in the spatial fields of Fig. 2. As per the projected signals of change, the three ensembles point out to a (quasi-)linear increase along the century and 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 shown in Fig. 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 pat-

terns 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 us 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 have been used to produce downscaled multi-model ensembles based on the perfect-prognosis approach and are compared against an ensemble of RCMs.

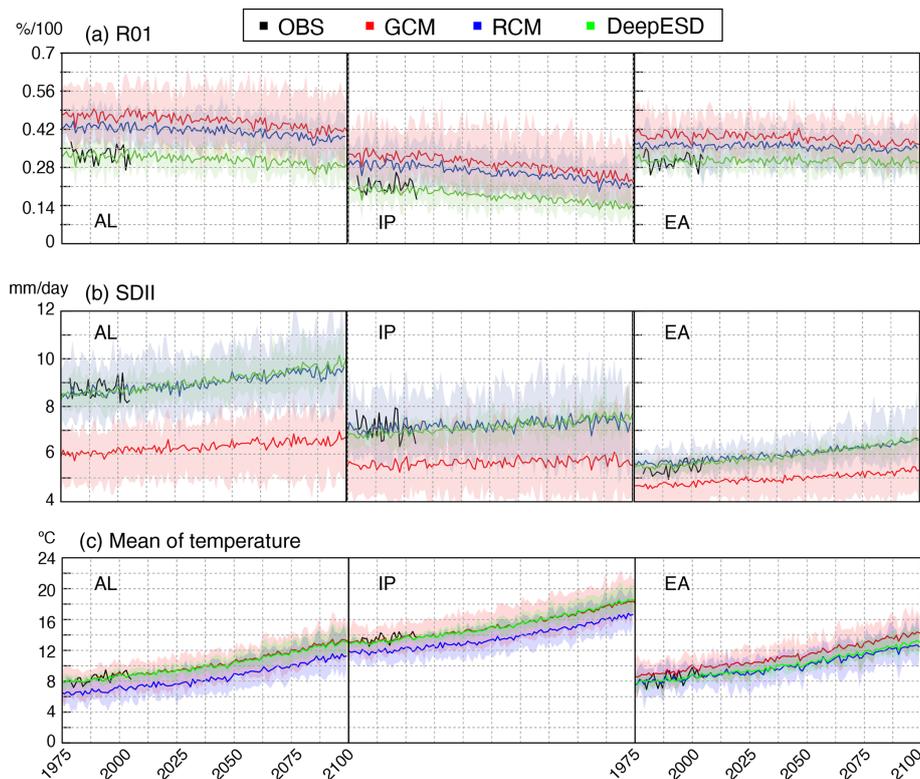


**Figure 5.** The climate change signals (2071–2100 with respect to 1975–2005) of the eight CMIP5 climate models considered and the equivalent DeepESD downscaled fields for precipitation (rows 1–2) and temperature (rows 4–5). Rows 3 and 6 show the difference between the DeepESD downscaled and raw climate change signals for the different GCMs.

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 regional differences – at a climatological scale 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 in the multi-model uncertainty with respect to the one of their driving GCMs. As per temperature, the CNNs project similar signals of change as the GCMs, being able to capture the particularities of each one and resulting in 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 us 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 DeepESD publicly available 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,



**Figure 6.** Annual time series for (a) R01, (b) SDII and (c) the mean of temperature, averaged over the Alps (AL), the Iberian Peninsula (IP) and Eastern Europe (EA) PRUDENCE regions. For every indicator, the ensemble of GCMs (red), RCMs (blue), and DeepESD (green) for the total period 1975–2100 and 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.

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: <https://doi.org/10.5281/zenodo.6823421>, Baño-Medina et al., 2022a) and the companion Jupyter notebook (DOI: <https://doi.org/10.5281/zenodo.6828303>, Baño-Medina et al., 2022b), 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 paper 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 the University of Cantabria (<https://data.meteo.unican.es/thredds/catalog/esgcat/catalog.html>, last access: 26 August 2022).

*Author contributions.* JB, RM and JMG conceived the experiment; JB produced all the results; and EC and ASC prepared the data for publication. All authors contributed to the analysis of results and to the writing of the manuscript.

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Acknowledgements.* The authors would like to acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<http://www.uerra.eu>, last access: 26 August 2022) and the Copernicus Climate Change Service, as well as the data providers in the ECA&D project (<https://www.ecad.eu>, last access: 26 August 2022). 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. In addition, Jorge Baño-

Medina 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.

*Financial support.* This research has been supported by the Spanish Government (MCIN/AEI /10.13039/501100011033) through project CORDyS (grant no. PID2020-116595RB-I00).

*Review statement.* This paper was edited by Charles Onyutha and reviewed by Alexander J. Winkler and four anonymous referees.

## References

- 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, *Int. J. Climatol.*, 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, 79–85, 2020.
- Baño-Medina, J., Manzanar, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, *Geosci. Model Dev.*, 13, 2109–2124, <https://doi.org/10.5194/gmd-13-2109-2020>, 2020.
- Baño-Medina, J., Manzanar, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, *Clim. Dynam.*, 57, 1–11, 2021.
- Baño-Medina, J., Manzanar, R., Cimadevilla, E., Fernández, J., González-Abad, J., Cofiño, A. S., and Gutiérrez, J. M.: 2022\_Bano\_DeepESD\_GMD\_data (1.0.0), Zenodo [data set], <https://doi.org/10.5281/zenodo.6823422>, 2022a.
- Baño-Medina, J., Manzanar, R., Cimadevilla, E., Fernández, J., González-Abad, J., Cofiño, A. S., and Gutiérrez, J. M.: Repository supporting the results presented in the manuscript on Downscaling Multi-Model Climate Projection Ensembles with Deep Learning (DeepESD): Contribution to CORDEX EUR-44 (v1.0.0), Zenodo [data set], <https://doi.org/10.5281/zenodo.6828304>, 2022b.
- Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., 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, *Geosci. Model Dev.*, 6, 687–720, <https://doi.org/10.5194/gmd-6-687-2013>, 2013.
- 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, *Clim. Dynam.*, 54, 2981–3002, 2020.
- 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?, *Clim. Dynam.*, 41, 803–817, 2013.
- Cannon, A. J.: Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network, *J. Hydrometeorol.*, 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, *J. Geophys. Res.-Biogeo.*, 115, G03014, <https://doi.org/10.1029/2008JG000920>, 2010.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, *J. Geophys. Res.-Atmos.*, 123, 9391–9409, 2018.
- Dai, A.: Precipitation characteristics in eighteen coupled climate models, *J. Climate*, 19, 4605–4630, 2006.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, D. P., and Bechtold, P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- 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., Prodhomme, C., Serradell Maronda, K., Tintó Prims, O., Batté, L., Volpi, D., Ceglár, A., Haarsma, R., and Massonnet, F.: Using EC-Earth for climate prediction research, *ECMWF Newsletter*, 154, 35–40, <https://doi.org/10.21957/fd9kz3>, 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., Jous-saume, 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, *Clim. Dynam.*, 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, *J. 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, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Fernández, J., Frías, M., Cabos, W., Cofiño, A., Domínguez, M., Fita, L., Gaertner, M., García-Díez, M., Gutiérrez, J. M., Jiménez-Guerrero, P., Liguori, G., Montávez, J. P., Romera, R., and Sánchez, E.: Consistency of climate change projections from multiple global and regional model intercomparison projects, *Clim. Dynam.*, 52, 1139–1156, 2019.
- François, B., Thao, S., and Vrac, M.: Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks, *Clim. Dynam.*, 57, 3323–3353, 2021.
- Giorgi, F.: Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?, *J. Geophys. Res.-Atmos.*, 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, *Environ. Res. Lett.*, 15, 034035, <https://doi.org/10.1088/1748-9326/ab6666>, 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., Manzananas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Rätty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerener, 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, *Int. J. Climatol.*, 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, *B. Am. Meteorol. Soc.*, 101, E664–E683, 2020.
- Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M. D., Manzananas, 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, *Environ. Modell. Softw.*, 111, 42–54, 2019.
- Klok, E. and Klein Tank, A.: Updated and extended European dataset of daily climate observations, *Int. J. Climatol.*, 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, *Environ. Modell. Softw.*, 100, 67–73, 2018.
- LeCun, Y. and Bengio, Y.: Convolutional networks for images, speech, and time series, *The handbook of brain theory and neural networks*, MIT Press, 255–258, 1995.
- Maraun, D. and Widmann, M.: *Statistical Downscaling and Bias Correction for Climate Research*, Cambridge University Press, Online ISBN 9781107588783, <https://doi.org/10.1017/9781107588783>, 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, *Earths 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), *J. Adv. Model. Earth Sy.*, 10, 1383–1413, 2018.
- Olmo, M. E., Balmaceda-Huarte, R., and Bettolli, M. L.: Multi-model ensemble of statistically downscaled GCMs over south-eastern South America: historical evaluation and future projections of daily precipitation with focus on extremes, *Clim. Dynam.*, online first, <https://doi.org/10.1007/s00382-022-06236-x>, 2022.
- Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S.: Improving Precipitation Estimation Using Convolutional Neural Network, *Water Resour. Res.*, 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, *Int. J. Climatol.*, 41, 211–232, 2021.
- San-Martín, D., Manzananas, R., Brands, S., Herrera, S., and Gutiérrez, J. M.: Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods, *J. 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, *Int. J. Climatol.*, 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, *Int. J. Climatol.*, 41, 1128–1147, 2021.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *B. Am. Meteorol. Soc.*, 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., Sandstad, M., Schwingshackl, C., Somot, S., Aalbers, E., Christensen, O. B., Ciarlo, J. M., Demory, N.-E., Giorgi, F., Jacob, D., Jones, R. G., Keuler, K., Kjellström, E., Lenderink, G., Levvasseur, G., Nikulin, G., Sillmann, J., Solidoro, C., Lund Sørland, S., Steger, C., Teichmann, C., Warrach-Sagi, K., and Wulfmeyer, V.: Evaluation of the large EURO-CORDEX regional climate model ensemble, *J. Geophys. Res.-Atmos.*, 126, e2019JD032344, <https://doi.org/10.1029/2019JD032344>, 2021.
- Voldoire, A., Sanchez-Gomez, E., Salas y Méliá, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevalier, 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, *Clim. Dynam.*, 40, 2091–2121, 2013.

- Vrac, M. and Ayar, P.: Influence of Bias Correcting Predictors on Statistical Downscaling Models, *J. Appl. Meteorol. Clim.*, 56, 5–26, <https://doi.org/10.1175/JAMC-D-16-0079.1>, 2016.
- Vrac, M., Stein, M., Hayhoe, K., and Liang, X.-Z.: A general method for validating statistical downscaling methods under future climate change, *Geophys. Res. Lett.*, 34, L18701, <https://doi.org/10.1029/2007GL030295>, 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., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C., Hoof, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3, 1–9, <https://doi.org/10.1038/sdata.2016.18>, 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., MIT Press, Proceedings of the Neural Information Processing Systems (NIPS), 985–991, ISBN 0-262-10076-2, 1998.