



# Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch

Ana Casanueva<sup>1</sup>  | Sixto Herrera<sup>1</sup>  | Maialen Iturbide<sup>2</sup>  | Stefan Lange<sup>3</sup>  |  
Martin Jury<sup>4,5</sup>  | Alessandro Dosio<sup>6</sup>  | Douglas Maraun<sup>4</sup> | José M. Gutiérrez<sup>2</sup> 

<sup>1</sup>Meteorology Group. Dpto. de Matemática Aplicada y Ciencias de la Computación, University of Cantabria, Santander, Spain

<sup>2</sup>Meteorology Group. Instituto de Física de Cantabria, CSIC-University of Cantabria, Santander, Spain

<sup>3</sup>Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

<sup>4</sup>Wegener Center for Climate and Global Change, University of Graz, Graz, Austria

<sup>5</sup>Earth Sciences Department, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain

<sup>6</sup>European Commission Joint Research Centre (JRC), Ispra, Italy

## Correspondence

José M. Gutiérrez, Meteorology Group. Instituto de Física de Cantabria, CSIC-University of Cantabria, Santander, Spain. Email: gutierjm@ifca.unican.es

## Funding information

Horizon 2020 Framework Programme, Grant/Award Numbers: 641816, 774652

## Abstract

Systematic biases in climate models hamper their direct use in impact studies and, as a consequence, many statistical *bias adjustment* methods have been developed to calibrate model outputs against observations. The application of these methods in a climate change context is problematic since there is no clear understanding on how these methods may affect key magnitudes, for example, the climate change signal or trend, under different sources of uncertainty. Two relevant sources of uncertainty, often overlooked, are the sensitivity to the observational reference used to calibrate the method and the effect of the resolution mismatch between model and observations (*downscaling* effect). In the present work, we assess the impact of these factors on the climate change signal of temperature and precipitation considering marginal, temporal and extreme aspects. We use eight standard and state-of-the-art bias adjustment methods (spanning a variety of methods regarding their nature—empirical or parametric—, fitted parameters and trend-preservation) for a case study in the Iberian Peninsula. The quantile trend-preserving methods (namely quantile delta mapping (QDM), scaled distribution mapping (SDM) and the method from the third phase of ISIMIP-ISIMIP3) preserve better the raw signals for the different indices and variables considered (not all preserved by construction). However, they rely largely on the reference dataset used for calibration, thus presenting a larger sensitivity to the observations, especially for precipitation intensity, spells and extreme indices. Thus, high-quality observational datasets are essential for comprehensive analyses in larger (continental) domains. Similar conclusions hold for experiments carried out at high (approximately 20 km) and low (approximately 120 km) spatial resolutions.

## KEYWORDS

bias adjustment, climate change signal, downscaling, observational uncertainty

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Atmospheric Science Letters* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

## 1 | INTRODUCTION

Bias adjustment (BA) techniques are routinely applied in sectoral impact studies to calibrate the required (biased) global and regional model outputs to regional or local scale, using a particular gridded or point-scale observational reference (Rojas *et al.*, 2012, Barredo *et al.*, 2016, Ruiz-Ramos *et al.*, 2016, Casanueva *et al.*, 2018, Reder *et al.*, 2018, Galmarini *et al.*, 2019). For this purpose, a number of methods have been developed (see e.g., Lafon *et al.*, 2013; Rätty *et al.*, 2014; Sunyer *et al.*, 2015; Maraun and Widmann, 2018), from simple methods calibrating the mean to methods adjusting all quantiles, either parametric or empirical, trend-preserving or not, and with different treatment of new extremes and the wet-day frequency for precipitation.

A number of fundamental limitations and uncertainties of BA methods have been already described in the literature in recent years (see, e.g., Dosio, 2016, Maraun *et al.*, 2017). Two main sources of uncertainty which may largely influence the results of these methods are (a) *observational uncertainty* (the sensitivity to the observational reference used for calibration) and (b) *resolution mismatch* (the mismatch of the horizontal resolution between model outputs and observations). Resolution mismatch typically requires the application of more general statistical downscaling techniques, such as regression or analog based (Gutiérrez *et al.*, 2019) suitable to transfer coarse model outputs to a higher resolution (Maraun and Widmann, 2018). The application of BA methods in this context is subject to several shortcomings (IPCC, 2015), since the observed higher-resolution signal is simply imposed on modelled data without any predictive ability and may produce statistical artefacts and a misrepresentation of the spatiotemporal structures in an attempt to explain unexplained smaller-scale variability (Maraun, 2013, Maraun *et al.*, 2017, Figure 6).

In this paper, we focus on these two aspects and perform an intercomparison study of a number of standard and state-of-the-art BA techniques, including standard empirical and parametric quantile mapping, and also more conservative trend-preserving (only for the mean or for the full distribution—quantiles) methods. The analysis is carried out over the Iberian Peninsula, a region with a large variety of climatic conditions where high-resolution observational datasets are available (in particular E-OBS, Haylock *et al.*, 2008, and Iberia01, Herrera *et al.*, 2019a). The aim is to assess (a) how the different methods may alter the raw climate change signal (of both global and regional model outputs), and (b) the influence that the observational uncertainty and the resolution mismatch may have on the results. Note that these techniques adjust the (biased) model values towards the corresponding observed ones and this

may indirectly affect the trends and the resulting climate change signal (Maraun, 2013). This could be justified in some cases, for instance, for highly biased climate indices such as those defined using absolute thresholds, where the raw signal is not reliable (e.g., Dosio, 2016). However, preserving the trends of the basic distributional statistics (mean and quantiles) is desirable in general if there are no physical mechanisms justifying a modification. Here, we present an intercomparison study of standard and trend-preserving methods focusing on a number of validation indices encompassing marginal, temporal and extreme aspects. The goal here is to assess these differences, and the influence of observational uncertainty and resolution mismatch, in order to facilitate an informed choice of methods based on the required behaviour.

The paper is structured as follows: First, Section 2 presents the data and methods considered in this study, including the gridded observational datasets, the global and regional climate models, the BA methods, experimental framework and the validation indices used. Section 3 presents the main results. Finally, the main conclusions and discussions are detailed in Section 4.

## 2 | DATA AND METHODS

The Iberian Peninsula is located in southwestern Europe, in the transition zone between extratropics and subtropics, spanning a region with complex orography influenced by both the Atlantic and Mediterranean climates. The resulting local climate variability ranges from temperate climates with regular precipitation spread over the whole year with more than 1,000 mm·year<sup>-1</sup> in the north, to dry (semiarid) climates with areas with less than 100 mm·year<sup>-1</sup> in the southeast. This large variability of climatic conditions makes the Iberian peninsula a good candidate to test the performance of different BA techniques.

### 2.1 | Observational gridded datasets for the Iberian Peninsula

Two regional observational gridded datasets have been used in this work, the pan-European E-OBS v19e (Haylock *et al.*, 2008, Cornes *et al.*, 2018, herein only the ensemble mean values<sup>1</sup> are considered) and the recently developed Iberia01 (Herrera *et al.*, 2019a) covering only the Iberian Peninsula. Both provide daily precipitation and temperature (mean, minimum and maximum values) on a 0.1° regular grid for the common period 1971–2015. Note that 0.1° is a nominal resolution and that the representation of the true climate depends on

station density, which varies largely between and within countries. E-OBS builds upon 208 stations in continental Spain and 17 (8) for precipitation (temperatures) in Portugal. Iberia01 is based on a quality controlled observational network of 3487 and 276 stations for precipitation and temperature, respectively, from the Spanish Agency of Meteorology (AEMET), the Portuguese Institute for Sea and Atmosphere (IPMA) and the Portuguese Environmental Agency (APA). Iberia01 was produced following previous efforts in Spain (Herrera *et al.*, 2016) and Portugal (for precipitation only, Belo-Pereira *et al.*, 2011). In this work the two observational datasets are upscaled from  $0.1^\circ$  to a  $0.2^\circ$  grid which matches the original grid at the grid cell boundaries (i.e., each  $0.2^\circ$  grid cell contains exactly four  $0.1^\circ$  grid cells). The main reason for this upscaling is that E-OBS effective resolution is coarser than  $0.1^\circ$  due to the stations' density (especially for precipitation) and any comparison at such a high resolution could rely on statistical artefacts;  $0.2^\circ$  could be considered a good candidate for a fairer comparison.

Herrera *et al.* (2019a) characterised the systematic differences between these two datasets, which are particularly relevant for extreme indices (see also Figure S1 for the indices considered in the present work).

## 2.2 | Climate model simulations

From the different Global Climate Model (GCM) simulations available within the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor *et al.*, 2012), in this work we consider a GCM, EC-EARTH (r12i1p1), that has been shown to consistently reproduce the key large-scale processes influencing the European climate, in particular storm tracks (Lee, 2014; Zappa *et al.*, 2015). This allows us to test BA methods in optimum conditions, that is, with no large (incorrigible) systematic biases for key processes. Temperature and precipitation raw outputs of EC-EARTH are available at the original  $1.125^\circ$  horizontal resolution. In addition, we also use results from the RACMO22E Regional Climate Model (RCM) from the EURO-CORDEX ensemble (see Kotlarski *et al.*, 2014), driven by the above-mentioned EC-EARTH simulation. RACMO22E outputs are available at a  $0.11^\circ$  resolution.

## 2.3 | Experimental framework

In this paper we build on the EURO-CORDEX intercomparison framework for (statistical) downscaling methods, which is a follow-up of the VALUE (Validating and Integrating Downscaling Methods for Climate Change Research) initiative (Maraun *et al.*, 2015). VALUE

conducted a first intercomparison experiment for assessing the relative merits and limitations of the different downscaling approaches (including BA) with perfect (reanalysis) predictors (see Gutiérrez *et al.*, 2019). A second follow on experiment has been proposed ([http://www.value-cost.eu/validation#Experiment\\_3a](http://www.value-cost.eu/validation#Experiment_3a)) to analyse the extrapolation capability of these methods using (global and regional) climate model projections from historical and future scenarios, in particular using the datasets described in Section 2.2. The methods are first trained over the historical period (1981–2010) using GCM/RCM outputs and observational datasets and, then, applied to the future GCM/RCM outputs for the 2071–2100 period under the RCP8.5 scenario.

In order to test the influence of observational uncertainty and resolution gaps between models and observations, we perform two experiments using two different resolutions for the target observational datasets (Iberia01 and E-OBS): (a)  $0.2^\circ$  high resolution and (b)  $1.125^\circ$  coarse resolution of the GCM. For the latter experiment, in which both model and observed data have the same horizontal resolution, observations (and RCM outputs for the case of RCM bias adjustment) are upscaled from their original resolution to the coarse (GCM) counterpart using conservative remapping (using the Climate Data Operators, CDO, Schulzweida, 2019).

## 2.4 | Validation indices

We have selected a number of indices proposed in the VALUE initiative to validate different aspects of downscaling methods (Maraun *et al.*, 2015). Table 1 describes the indices considered and the particular aspects assessed (M = marginal, T = temporal, E = extremes, see their observed climatology as given by the observational datasets in Figure S1). In this work we are interested in assessing the effect of the BA methods on the raw (from both global and regional models) climate change signal (defined as the difference, or delta, between future projections and historical climate) intercomparing methods with and without trend preservation. Note that, by construction, some indices (mean temperature, percentiles) are directly adjusted by some methods, which should be taken into account in the frame of a fair comparison (Casanueva *et al.*, 2016). For instance, for some indices (e.g., the mean climate change signal) some BA methods (e.g., ISIMIP1) are expected to preserve the raw trend (to avoid statistical artefacts introduced by the adjustment method with no physical justification). However, there are other indices largely affected by model biases, for example, threshold-based indices, here specifically FA20 (tasmin), where the raw model signal may be unreliable and, therefore, adjusted trends may not necessarily indicate a bad performance of the method (see

**TABLE 1** Diagnostic indices used in this work

Index name	Pr	Temp	Description	M	T	E
<b>AC1 (tas)</b>		X	Lag-1 autocorrelation			X
<b>Mean (tas)</b>		X	Mean value			X
<b>WarmAnnualMaxSpell (tasmax)</b>		X	Median of the annual warm spell maxima (maximum number of consecutive days with temperature >90th percentile)			X X
FA20 (tasmin)		X	Tropical nights (days with $t_{\min} > 20^{\circ}\text{C}$ )			X
R01	X		Wet-day frequency (number of days with precipitation $\geq 1$ mm)			X
<b>SDII</b>	X		Mean wet-day precipitation			X
<b>DryAnnualMaxSpell</b>	X		Median of the annual dry spell maxima (maximum number of consecutive days with precipitation <1 mm)			X X
<b>DWProb</b>	X		Dry-wet transition probability			X
P98Wet	X		98th percentile of precipitation amount on wet days			X

Note: Indices names refer to the package <https://github.com/SantanderMetGroup/VALUE>, where the code for their calculation can be found. Second and third columns indicate whether the indices are suitable for validating precipitation (Pr) and temperature (Temp), respectively. The last three columns indicate if the indices analyse marginal (M, Gutiérrez *et al.*, 2019), temporal (T, Maraun *et al.*, 2019) or extreme (E, Hertig *et al.*, 2019) aspects, respectively. Indices highlighted in bold are those used in the synthesis VALUE paper (Maraun *et al.*, 2019). Note that a warm or dry spell has been defined as at least two consecutive days fulfilling a particular condition.

**TABLE 2** Bias adjustment methods used in the intercomparison study, indicating the code, name, parametric/empirical (P/E) character, the software package and a reference describing the method

Code	Name	P/E	Package	Reference
<b>EQM</b>	<b>Empirical Quantile mapping</b>	<b>E</b>	<b>climate4R</b>	Déqué (2007)
<b>PQM</b>	<b>Parametric Quantile mapping</b>	<b>P</b>	<b>climate4R</b>	Piani <i>et al.</i> (2010)
<b>GPQM</b>	<b>Generalized Pareto parametric quantile mapping</b>	<b>P</b>	<b>climate4R</b>	Vrac and Naveau (2007)
DQM	Detrended quantile mapping	E	ClimDown	Cannon <i>et al.</i> (2015)
QDM	Quantile delta mapping	E	ClimDown	Cannon <i>et al.</i> (2015)
SDM	Scaled distribution mapping	P	pyCAT	Switanek <i>et al.</i> (2017)
<b>ISIMIP1</b>	<b>Method from the ISIMIP fast track</b>	<b>P</b>	<b>climate4R</b>	Hempel <i>et al.</i> (2013)
ISIMIP3	Method from the third phase of ISIMIP	P	ISIMIP3BASD	Lange (2019)

Note: Methods highlighted in bold were included in the VALUE initiative (Gutiérrez *et al.*, 2019). The package climate4R (which includes the package downscaleR for downscaling; Bedia *et al.*, 2020) is available at <https://github.com/SantanderMetGroup/climate4R>; the package ClimDown is available at <https://github.com/pacificclimate/ClimDown>; the code for ISIMIP3 method is available at <https://doi.org/10.5281/zenodo.3648654>; the code for SDM is available at <https://github.com/wegener-center/pyCAT>.

discussion in Dosio, 2016). The goal here is to assess these differences in order to facilitate an informed choice of methods based on the required behaviour.

## 2.5 | Bias adjustment methods

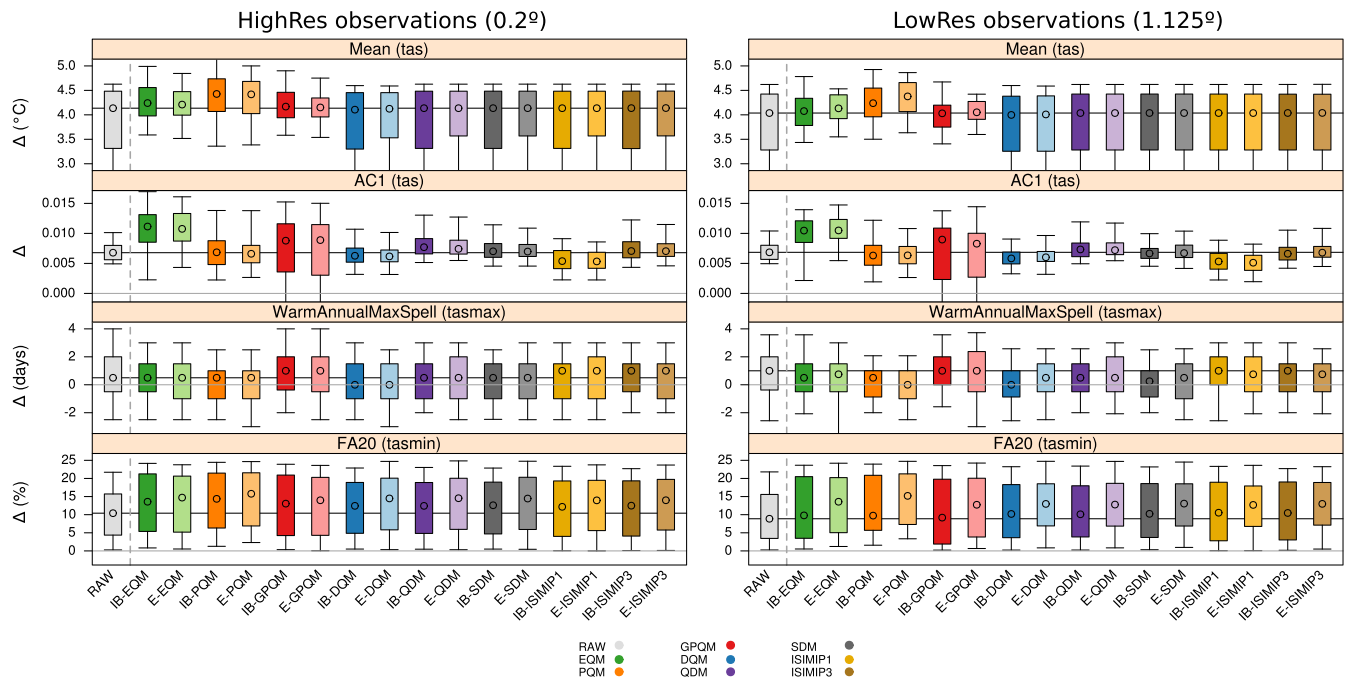
In this work, we compare a selection of BA methods, including some contributing to the VALUE intercomparison experiment (Maraun *et al.*, 2015; Gutiérrez *et al.*, 2019) and four additional trend-preserving methods, as shown in Table 2.

The main characteristics of these methods are described below:

- **Empirical quantile mapping (EQM):** Empirical method where a transfer function calibrated over the control period is used to map quantiles from the empirical cumulative distribution function of the model output onto the corresponding observed distribution. The formulation considered in this work is implemented in the climate4R tools (Iturbide *et al.*, 2019; Bedia *et al.*, 2020). Unlike the method used in the Cost Action VALUE (Gutiérrez *et al.*, 2019), which fits 99 empirical percentiles, it calibrates all experimental quantiles. It uses constant extrapolation (first and last corrections for values below and above the calibration range, respectively) for out-of-sample values, as the implementation used in VALUE. It adjusts wet-day

occurrences using a revised threshold which matches the observed and model simulated wet/dry day frequency. Moreover, frequency adaptation (Themeßl *et al.*, 2012), but sampling from the observed Gamma distribution instead of using linear interpolation, is applied in order to simulate rain in case of an excess of model dry days. It does not consider any specific corrections for trends, but allows for their modification if biases are intensity-dependent.

- Parametric quantile mapping (PQM): Parametric method where a transfer function calibrated over the control period is used which adjusts the theoretical empirical cumulative distribution function (Gamma for precipitation and Gaussian for temperature) of the model output onto the corresponding observed distribution. The formulation considered in this work is the one used in the Cost Action VALUE and implemented in the climate4R tools. The same wet-day frequency correction and frequency adaptation as EQM are applied. It does not consider any specific corrections for trends, and extremes are adjusted by the fitted parametric distribution.
- Generalized Pareto parametric quantile mapping (GPQM): Parametric method that fits Gamma (or Gaussian for temperature) and Pareto distributions below/above the 95th percentile. The same wet-day frequency correction and frequency adaptation as EQM are applied. Distributions are then fitted for wet days only. The method does not consider any specific corrections for trends while extremes are adjusted based on the statistical extreme distribution.
- Detrended quantile mapping (DQM): Empirical method whose application consists of three steps (i) removing the long-term mean (linear) trend; (ii) applying empirical quantile mapping (using all quantiles) to the detrended series; (iii) adding the mean trend to the bias-adjusted series. Zeros in the observed and modelled data are replaced with nonzero uniform random values below the trace threshold prior to bias correction. By doing so, the transfer function can be calibrated using all days from the model and observed series. After the correction, the days with precipitation lower than a predefined wet-day threshold are recorded as zero. This method does not consider any specific corrections for extremes, while the mean trend is preserved by construction.
- Quantile delta mapping (QDM): Empirical method divided in three steps (i) future model outputs are detrended by quantile; (ii) quantile mapping is applied to all empirical detrended quantiles of the detrended series; (iii) the projected trends are reapplied to the bias-adjusted quantiles. Same correction for the wet-day frequency as for DQM is used. The method does not consider any specific corrections for extremes, while the trend is preserved for all quantiles.
- Scaled Distribution Mapping (SDM): Trend-preserving parametric method that scales monthly observed distributions by changes in the model's past and future distributions (multiplicative assuming a Gamma distribution for precipitation and absolute assuming a Normal distribution for temperature) and likelihood of events (Switanek *et al.*, 2017). Prior to the scaling, days with less precipitation than 1 mm are set to zero, temperature values are detrended. After the scaling, bias-corrected values are reordered to their original position in time and the temperature trend is added again. In case a model overestimates the number of wet days, the least wet days are treated as dry days, in case of underestimation raw modelled rain-day frequency remains unchanged. SDM does not consider any specific corrections for extremes.
- ISIMIP1: This is a trend-preserving parametric method developed in the ISIMIP Fast Track (Hempel *et al.*, 2013; Warszawski *et al.*, 2014). Here we use the implementation provided in climate4R (note that the original implementation of this method is available at <https://github.com/ISI-MIP/BC>). The method consists of a correction of the monthly mean value (using correction offsets for temperature and correction factors for precipitation) followed by the correction of daily variability around the monthly mean value (using linear transfer functions for temperature and nonlinear transfer functions for precipitation). The method preserves additive/multiplicative trends in the monthly mean value because the same correction offset/factor is used in all application periods. It adjusts the wet-day frequency only if it is biased high.
- ISIMIP3: This is a parametric bias correction method developed for the third phase of ISIMIP (Lange, 2019). It generates pseudo future observations by transferring, for each quantile, the simulated climate change signal to the historical observations. It then uses these pseudo future observations as “reference” for correcting future model simulations with parametric quantile mapping. Any trend in daily mean temperature is removed before and restored after these two steps. Daily minimum and maximum temperature are not corrected directly. Instead, amplitude and skewness of the diurnal temperature cycle are corrected. Corrected daily minimum and maximum temperature are then derived from those and the corrected daily mean temperature. The method adjusts the wet-day frequency and generates wet days using random values drawn from the distribution. It does not consider any specific corrections for extremes and preserves the trends in all percentiles instead of only the mean trend.



**FIGURE 1** GCM climate change signal (deltas,  $\Delta$ ) of temperature indices for the 2071–2100 (RCP8.5) period with respect to the baseline 1981–2010 for the raw model output (first boxplot in each panel) together with bias adjusted results (rest of boxplots, see Table 2). Results are shown for two similar BA experiments with highRes ( $0.2^\circ$ , left column) and coarse ( $1.125^\circ$ , right column) observational reference data from two different datasets: Iberia01 (IB, dark-coloured boxes) and E-OBS (E, light-coloured boxes). Each coloured box represents the interquartile range, whiskers expand from the fifth to 95th percentiles of the signals' range and outliers are not shown. Note that in the high-resolution experiment the GCM outputs are “downscaled” to the target  $0.2^\circ$  resolution (using the same closest model gridbox for all insider observation gridboxes), whereas in the coarse-resolution experiment both model and observations have the same resolution (no downscaling effect). Black horizontal lines depict the median of the raw delta for reference

For the sake of better comparability among methods, all of them were calibrated on a monthly basis since some methods explicitly work month by month (SDM, ISIMIP1, ISIMIP3). As shown, all methods deal with the wet-day frequency differently but all use a wet-day threshold of 1 mm.

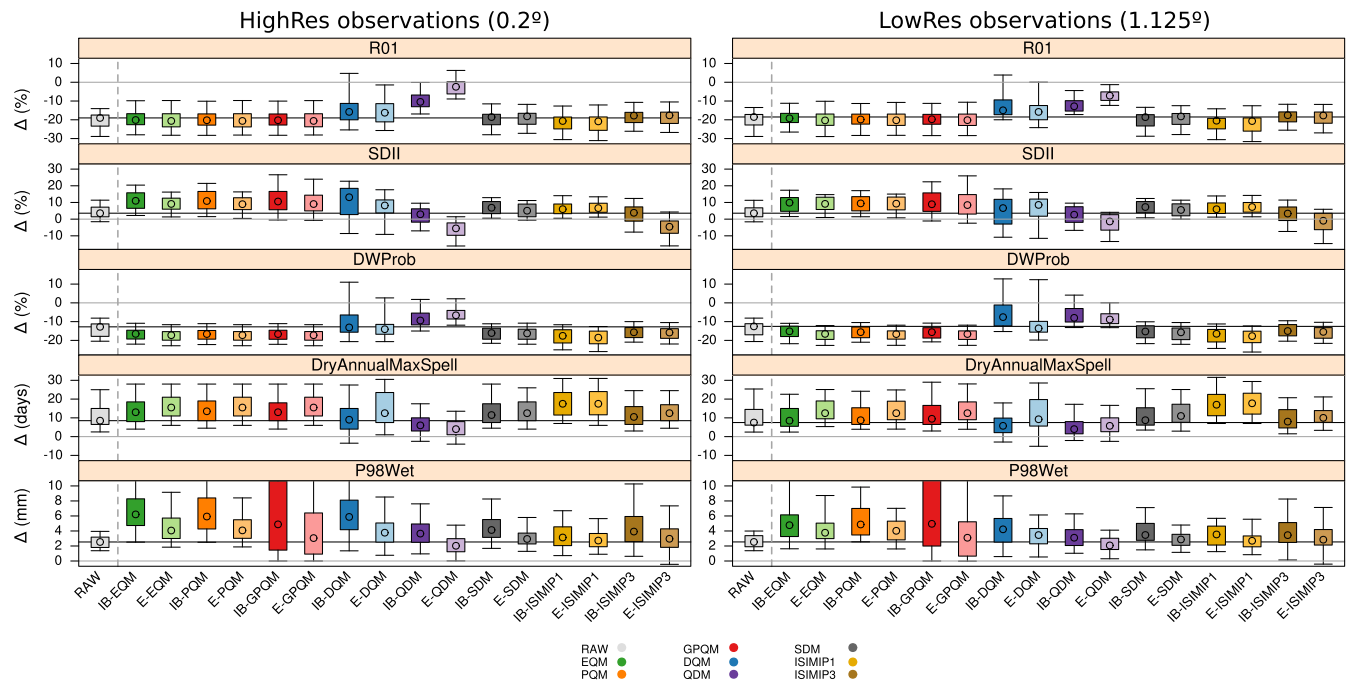
### 3 | RESULTS

Figures 1 and 2 show the GCM climate change signals (absolute or relative deltas for the 2071–2100 RCP8.5 period with respect to the baseline 1981–2010) for the marginal, temporal and extreme indices shown in Table 1, for temperature and precipitation respectively. As mentioned before, results are adjusted with respect to two different observational reference datasets (Iberia01 and E-OBS); consequently, the corresponding results for each BA method are shown as pairs in these figures with darker and lighter colours for Iberia01 and E-OBS, respectively. This allows easy analysis of the influence of the observational reference in the adjusted results, thus estimating the effect of observational uncertainty.

Moreover, both high- and coarse-resolution observations are considered in the process to adjust the coarse-resolution model outputs (left and right columns, respectively), to assess the influence of the resolution gap (downscaling effect).

#### 3.1 | Representation of the climate change signals

Overall, the largest differences arise between the standard and trend-preserving methods, showing the latter a greater preservation of the original raw GCM signal for most of the indices considered (not all of them directly preserved by construction in the BA process). This is especially noticeable for temperature where the three standard methods (EQM, PQM and GPQM) yield warmer future conditions, but with no physical mechanisms justifying the signal increase (see Figure S2 for spatial details). Among the trend-preserving methods, DQM and QDM exhibit large departures from the raw signal for rainfall frequency (R01), which in turn affects other indices. The reason for this could be the particular treatment of wet-



**FIGURE 2** As Figure 1 for different precipitation indices (see Table 1). GCM climate change signal (deltas,  $\Delta$ ) for the 2071–2100 (RCP8.5) period with respect to the baseline 1981–2010 for the raw model output (first boxplot in each panel) together with bias adjusted results (rest of boxplots, see Table 2). Results are shown for two similar BA experiments with highRes ( $0.2^\circ$ , left column) and coarse ( $1.125^\circ$ , right column) observational reference data from two different datasets: Iberia01 (IB, dark-coloured boxes) and E-OBS (E, light-coloured boxes). Each coloured box represents the interquartile range, whiskers expand from the fifth to 95th percentiles of the signals' range and outliers are not shown. Black horizontal lines depict the median of the raw delta for reference

day frequency, which is not explicitly adjusted by these two methods (see Section 2.5 and Figure S3).

With regard to the temporal indices, although lag-1 autocorrelation in mean temperature (AC1) is not explicitly adjusted by any method, the trend-preserving ones show results closer to those of the original data with respect to the rest of the methods considered. In this sense, the empirical quantile mapping (EQM) clearly overestimates the signal while the parametric approach (PQM) provides more consistent results with the raw signals. This is expected since fitting the Gaussian distribution implies a linear transformation of the time series. For the dry and warm spells the signal is mostly preserved by most of the methods (especially for warm spells) as could be expected considering that the thresholds used to define the spells (wet-day frequency and 90th temperature percentile, respectively) are calibrated and, as a consequence, the relative temporal structure is preserved by most BA methods.

In the case of extreme indices, such as the 98th percentile of wet-days, there are significant differences between the different trend-preserving approaches and the remaining methods (Figure 2 and Figure S4). The method presenting the largest deviations to the raw signals is the generalised extreme parametric method (GPQM), which very likely extrapolates to out-of-sample results, thus magnifying the

signals. All methods tend to increase the raw model signals, and to a larger extent when Iberia01 is used as observational reference, especially for the high resolution experiment. This could be due to the higher p98Wet values recorded by Iberia01 compared to E-OBS (Figure S1 and Herrera *et al.*, 2016) which yield higher projected bias adjusted values (see Figure S4) thus amplifying the climate change signal. The signal of the frequency of tropical nights (FA20, see Figure 1) is modified similarly by all the BA methods yielding slightly higher changes, which is expected since threshold-based indices are largely affected by model biases.

Note that similar conclusions are obtained for experiments performed at high ( $0.2^\circ$ ) and low ( $1.125^\circ$ ) resolutions, depicting smoother spatial patterns for the experiment at the coarse resolution (Figures S2–S4). Also similar conclusions regarding the performance of the BA methods hold for the RCM (Figures S6 and S7).

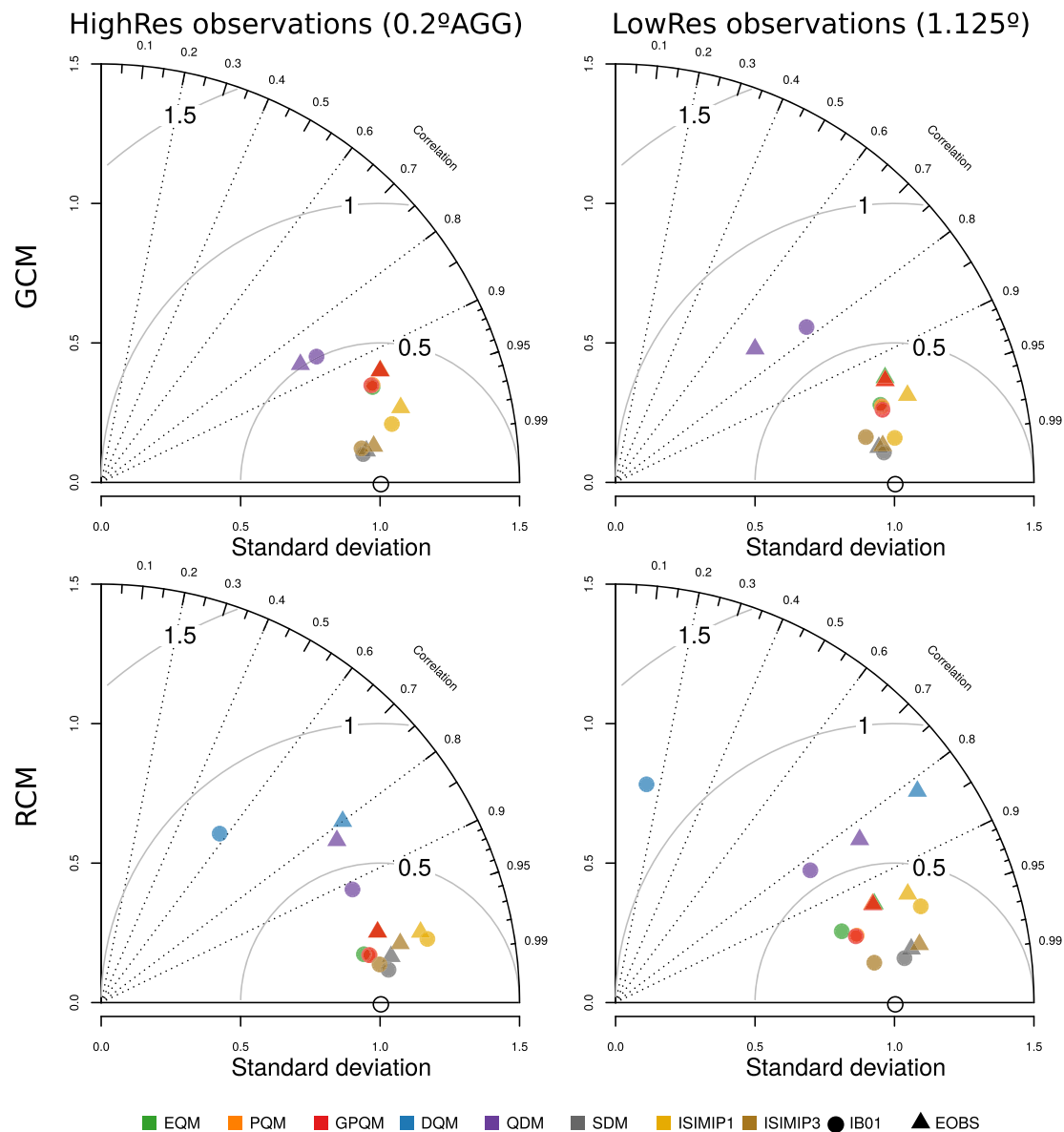
### 3.2 | Observational uncertainty

The effect of the observational reference on the change signal is larger for precipitation indices, in particular, extreme indices (p98Wet) for the standard BA methods and

marginal indices (SDII) for the trend-preserving methods (especially QDM and ISIMIP3). The different sensitivity of the methods to the observational datasets for extreme indices may be related to the special treatment of new extremes. Whereas EQM applies a constant extrapolation based on the correction of the last quantile and the parametric methods use the fitted distributions, extrapolation is avoided by DQM and QDM, since removing the modelled trend prior to quantile mapping shifts the future distribution so that it tends to lie within the historical distribution (Cannon *et al.*, 2015). Note the higher quality and accuracy of Iberia01 than E-OBS representing extremes, especially

for precipitation, regardless of the spatial scale considered (Herrera *et al.*, 2016), partly due to a denser station network (Herrera *et al.*, 2019b). QDM and ISIMIP3 rely on the whole observed distribution to a larger extent than other methods since the simulated signal is transferred to the observations to generate pseudo future observations, to which the quantile mapping is applied.

A greater impact of the observational dataset is observed on the actual projected values (Figure S5). This is evident for indices such as R01, SDII and p98Wet for all methods. Thus, the observational reference may play a more important role than the BA method in the projected values, whereas the



**FIGURE 3** Taylor diagrams for the climate change signal of R01, for the GCM (upper row) and the RCM (lower panel), for the two BA experiments with highRes (0.2°AGG, see text, left column) and coarse (1.125°, right column) observational reference data from two different datasets: Iberia01 (circles) and E-OBS (triangles). Within each panel, the reference value represents the raw signals and the coloured markers depict the signal of the BA methods (see Table 2). Note that DQM lies outside the range in the upper panels; EQM, PQM and GPQM group together in all panels



method poses the largest source of uncertainty when assessing the climate change signals, except for p98Wet which is highly sensitive to the observational reference.

### 3.3 | Resolution effect

In order to assess the resolution effect more specifically, we analyse the spatial patterns of the climate change signals (Figure 3 and Figures S8–S12). These Taylor diagrams (Taylor, 2001) show the degree of agreement between the spatial patterns of raw and bias-adjusted results, considering the high and low resolution experiments. For the sake of comparability of the spatial patterns, the signals at the high resolution have been conservatively remapped onto the  $1.125^\circ$  (hereafter  $0.2^\circ$ AGG). The resolution effect can be then assessed by comparing results at  $0.2^\circ$ AGG and  $1.125^\circ$ , whereas the effect of including the RCM in the downscaling process can be observed by comparing the top panels with the corresponding bottom ones.

The BA methods performance is rather consistent for a given index regardless of the experiment resolution and, in fact, the uncertainty due to the chosen BA method is larger than the uncertainty due to the experiment resolution. Overall, there are small differences when observations at high (i.e., original) or coarse (upscaled to the GCM grid) resolution are used for an assessment performed at the coarse resolution. There is some indication of a better agreement with the raw signal for the RCM for most indices (Figures S8–S12) except for FA20 (the GCM at  $0.2^\circ$ AGG slightly improves upon the other experiments) and WarmAnnualMaxSpell (similar results for RCM and GCM). This could be an indication of the better preservation of the raw signals when the resolution mismatch between the original model and observation is small.

Observational uncertainty is also smaller than the uncertainty due to the chosen BA method. Using Iberia01 as reference generally contributes to a better preservation of the raw signals, especially for precipitation indices (Figure 3 and Figures S10 and S11), except p98Wet (Figure S12). Results are not very conclusive for temperature indices. For FA20 (Figure S9) all methods show very similar signals among them (cf. Figure 1) and increase the raw signal due to the systematic underestimation by the raw RCM and GCM in the calibration period.

## 4 | SUMMARY AND CONCLUSIONS

This work presents an intercomparison of eight standard and state-of-the-art BA methods assessing the role of

observational uncertainty and resolution mismatch in the frame of climate projections. The analyses are performed using a set of climate indices, representing marginal, temporal and extreme aspects of temperature and precipitation.

The use of different BA methods is confirmed to be a large source of uncertainty in this work; most methods produce some modifications of the raw signals. It is well known that BA can potentially modify the climate change signal from the raw model output. Such changes might be advantageous in some cases, for instance for stationary, intensity-dependent biases (Gobiet *et al.*, 2015; Ivanov *et al.*, 2018), but fundamental climate model errors (e.g., unrealistically represented processes) cannot be improved by statistical postprocessing (Maraun *et al.*, 2017). If a climate model simulates a credible climate change signal, and no clear case-specific physical argument exists why the statistically modified signal should be more plausible, trend-preserving methods are a preferable choice (Maraun, 2016).

As expected, trend-preserving methods preserve better the signal of the raw models, while modifying other indices where a change is expected (e.g., FA20), similarly to standard, unconstrained methods. Overall, the methods which largely preserve the raw signals across the different variables and indices are the quantile trend-preserving methods QDM, SDM and ISIMIP3, although the former exhibits some problems with the correction of wet-day frequency. However, there is an indication of higher sensitivity to the choice of the observational reference for these methods than for the standard ones for precipitation indices representing marginal aspects (SDII) whereas standard BA methods are more sensitive to the observational dataset for extreme indices (P98Wet). Thus, a high-quality reference is desirable. Note that the observational reference has a larger impact on the projected indices than in the change signals, and also larger for extreme than for marginal and temporal indices, posing a larger source of uncertainty than the chosen BA method.

Regarding the resolution effect, we found some indication of better preservation of the raw signals when the resolution mismatch between the original model and observation is small (i.e., for the RCM). For a coarser model, similar results are obtained when observations at high (original) and low (upscaled to the GCM grid) resolution are used. Note that the observational datasets used here are gridded products which, despite the scale mismatch, are more robust than point-based observations, for which bias adjustment may introduce undesired statistical artefacts (Maraun, 2013). The choice of the BA method, however, remains as a major source of uncertainty compared to the resolution effect due to the experiment resolution and observational dataset when analysing the change signals.

We conclude that the choice of trend-preserving methods is recommended in general applications of BA to postprocess model outputs since they are conservative methods well suited to alleviate biases while maintaining the raw original climate change signal at the same time. The present work paves the way for further comparisons using recently developed and promising BA methods, which could be further extended to larger domains and used in new international and collaborative initiatives, such as the next VALUE experiment.

## ACKNOWLEDGEMENTS

We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<https://www.uerra.eu>) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (<https://eca.knmi.nl>). The authors are grateful to the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modelling, former coordinating body of CORDEX and responsible panel for CMIP5. We also thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation infrastructure an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modelling and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP). This study contributes to the EURO-CORDEX pillar on statistical downscaling, which is a follow-up of the EU COST Action ES1102 VALUE (Validating and Integrating Downscaling Methods for Climate Change Research). Participation of S. Herrera and J.M. Gutiérrez was partially supported by the project AfriCultuReS (European Union's Horizon 2020 program, grant agreement no. 774652). S. Lange acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 641816 (CRESCENDO). The authors are also grateful to three anonymous reviewers who helped to improve the original manuscript.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ORCID


Ana Casanueva  <https://orcid.org/0000-0002-7568-0229>


Sixto Herrera  <https://orcid.org/0000-0002-5384-179X>

Maialen Iturbide  <https://orcid.org/0000-0002-5048-0941>

Stefan Lange  <https://orcid.org/0000-0003-2102-8873>

Martin Jury  <https://orcid.org/0000-0003-0590-7843>

Alessandro Dosio  <https://orcid.org/0000-0002-6365-9473>

José M. Gutiérrez  <https://orcid.org/0000-0002-2766-6297>

## ENDNOTE

<sup>1</sup>More details about the ensemble versions of E-OBS in [https://surfobs.climate.copernicus.eu/dataaccess/access\\_eobs.php](https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php).

## REFERENCES

- Barredo, J.I., Caudullo, G. and Dosio, A. (2016) Mediterranean habitat loss under future climate conditions: assessing impacts on the Natura 2000 protected area network. *Applied Geography*, 75, 83–92. <https://doi.org/10.1016/j.apgeog.2016.08.003>.
- Bedia, J., Baño-Medina, J., Legasa, M.N., Iturbide, M., Manzanas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A.S. and Gutiérrez, J.M. (2020) Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, 13, 1711–1735. <https://doi.org/10.5194/gmd-13-1711-2020>.
- Belo-Pereira, M., Dutra, E. and Viterbo, P. (2011) Evaluation of global precipitation data sets over the Iberian Peninsula. *Journal of Geophysical Research: Atmospheres*, 116, D20101. <https://doi.org/10.1029/2010JD015481>.
- Cannon, A.J., Sobie, S.R. and Murdock, T.Q. (2015) Bias correction of GCM precipitation by Quantile mapping: how well do methods preserve changes in quantiles and extremes? *J. Climate*, 28, 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>.
- Casanueva, A., Herrera, S., Fernández, J. and Gutiérrez, J.M. (2016) Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative. *Climatic Change*, 137, 411–426. <https://doi.org/10.1007/s10584-016-1683-4>.
- Casanueva, A., Bedia, J., Herrera, S., Fernández, J. and Gutiérrez, J.M. (2018) Direct and component-wise bias correction of multivariate climate indices: the percentile adjustment function diagnostic tool. *Climatic Change*, 147, 411–425. <https://doi.org/10.1007/s10584-018-2167-5>.
- Cornes, R., van der Schrier, G., van den Besselaar, E.J.M. and Jones, P.D. (2018) An ensemble version of the E-OBS temperature and precipitation datasets. *Journal of Geophysical Research – Atmospheres*, 123, 9391–9409. <https://doi.org/10.1029/2017JD028200>.
- Déqué, M. (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global and Planetary Change*, 57, 16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>.
- Dosio, A. (2016) Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models. *Journal of Geophysical Research—Atmospheres*, 121, 5488–5511. <https://doi.org/10.1002/2015JD024411>.
- Galmarini, S., Cannon, A.J., Ceglar, A., Christensen, O.B., de Noblet-Ducoudré, N., Dentener, F., Doblas-Reyes, F.J., Dosio, A., Gutiérrez, J.M., Iturbide, M., Jury, M., Lange, S., Loukos, H., Maiorano, A., Maraun, D., McGinnis, S., Nikulin, G., Riccio, A., Sanchez, E., Solazzo, E., Toret, A., Vrac, M. and Zampieri, M. (2019) Adjusting climate model bias for agricultural impact assessment: how to cut the mustard. *Climate Services*, 13, 65–69. <https://doi.org/10.1016/j.cliser.2019.01.004>.

- Gobiet, A., Suklitsch, M. and Heinrich, G. (2015) The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal. *Hydrology and Earth System Sciences*, 19, 4055–4066. <https://doi.org/10.5194/hess-19-4055-2015>.
- Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzananas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtolos, J., Rätty, 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. (2019) An Intercomparison of a large Ensemble of Statistical Downscaling Methods over Europe: results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*, 39(9), 3750–3785. <https://doi.org/10.1002/joc.5462>.
- Haylock, M., Hofstra, N., Klein-Tank, A., Klok, E.J., Jones, P. and New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113, D20119. <https://doi.org/10.1029/2008JD010201>.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J. and Piontek, F. (2013) A trend-preserving bias correction - the ISI-MIP approach. *Earth System Dynamics*, 4, 219–236. <https://doi.org/10.5194/esd-4-219-2013>.
- Herrera, S., Fernández, J. and Gutiérrez, J.M. (2016) Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology. *International Journal of Climatology*, 36(2), 900–908. <https://doi.org/10.1002/joc.4391>.
- Herrera, S., Cardoso, R.M., Soares, P.M., Espírito-Santo, F., Viterbo, P. and Gutiérrez, J.M. (2019a) Iberia01: a new gridded dataset of daily precipitation and temperatures over Iberia. *Earth System Science Data*, 11, 1947–1956. <https://doi.org/10.5194/essd-11-1947-2019>.
- Herrera, S., Kotlarski, S., Soares, P.M.M., Cardoso, R.M., Jaczewski, A., Gutiérrez, J.M. and Maraun, D. (2019b) Uncertainty in gridded precipitation products: influence of station density, interpolation method and grid resolution. *International Journal of Climatology*, 39, 3717–3729. <https://doi.org/10.1002/joc.5878>.
- Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J.M., Wibig, J., Casanueva, A. and Soares, P.M.M. (2019) Comparison of statistical downscaling methods with respect to extreme events over Europe: validation results from the perfect predictor experiment of the COST action VALUE. *International Journal of Climatology*, 39, 3846–3867. <https://doi.org/10.1002/joc.5469>.
- IPCC. (2015) Workshop report of the intergovernmental panel on climate change workshop on regional climate projections and their use in impacts and risk analysis studies. In: Stocker, T.F., D. Qin, G.-K. Plattner, and M. Tignor (Eds.) IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, p. 171.
- 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. (2019) The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software*, 111, 42–54. <https://doi.org/10.1016/j.envsoft.2018.09.009>.
- Ivanov, M.A., Luterbacher, J. and Kotlarski, S. (2018) Climate model biases and modification of the climate change signal by intensity-dependent bias correction. *Journal of Climate*, 31, 6591–6610. <https://doi.org/10.1175/JCLI-D-17-0765.1>.
- Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V. (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, 7, 1297–1333. <https://doi.org/10.5194/gmd-7-1297-2014>.
- Lafon, T., Dadson, S., Buys, G. and Prudhomme, C. (2013) Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, 33, 1367–1381. <https://doi.org/10.1002/joc.3518>.
- Lange, S. (2019) Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, 12, 3055–3070. <https://doi.org/10.5194/gmd-12-3055-2019>.
- Lee, R.W. (2014) Storm track biases and changes in a warming climate from an extratropical cyclone perspective using CMIP5. PhD Thesis. Atmosphere, Oceans and Climate. Department of Meteorology. <http://www.met.reading.ac.uk/~xy744611/stormtracks/>.
- Maraun, D. (2013) Bias correction, Quantile mapping, and downscaling: revisiting the inflation issue. *Journal of Climate*, 26, 2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>.
- Maraun, D. (2016) Bias correcting climate change simulations—a critical review. *Current Climate Change Reports*, 2(4), 211–220. <https://doi.org/10.1007/s40641-016-0050-x>.
- Maraun, D. and Widmann, M. (2018) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107588783>.
- Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. I. (2015) VALUE: a framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1), 2014EF000259. <https://doi.org/10.1002/2014EF000259>.
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A. and Mearns, L.O. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), 764–773. <https://doi.org/10.1038/nclimate3418>.
- Maraun, D., Widmann, M. and Gutiérrez, J.M. (2019) Statistical downscaling skill under present climate conditions: a synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology*, 39(9), 3692–3703. <https://doi.org/10.1002/joc.5877>.
- Piani, C., Haerter, J.O. and Coppola, E. (2010) Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99, 187–192. <https://doi.org/10.1007/s00704-009-0134-9>.
- Rätty, O., Räisänen, J. and Ylhäisi, J.S. (2014) Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. *Climate Dynamics*, 42, 2287–2303. <https://doi.org/10.1007/s00382-014-2130-8>.

- Reder, A., Iturbide, M., Herrera, S., Rianna, G., Mercogliano, P. and Gutiérrez, J.M. (2018) Assessing variations of extreme indices inducing weather-hazards on critical infrastructures over Europe—the INTACT framework. *Climatic Change*, 148(1), 123–138. <https://doi.org/10.1007/s10584-018-2184-4>.
- Rojas, R., Feyen, L., Bianchi, A. and Dosio, A. (2012) Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. *Journal of Geophysical Research—Atmospheres*, 117, D17109. <https://doi.org/10.1029/2012JD017461>.
- Ruiz-Ramos, M., Rodríguez, A., Dosio, A., Goodess, C.M., Harpham, C., Mínguez, M.I. and Sánchez, E. (2016) Comparing correction methods of RCM outputs for improving crop impact projections in the Iberian Peninsula for 21st century. *Climatic Change*, 134, 283–297. <https://doi.org/10.1007/s10584-015-1518-8>.
- Schulweida, U. (2019) *CDO User Guide* (Version 1.9.8), <https://doi.org/10.5281/zenodo.3539275>.
- Sunyer, M.A., Hundsdoerfer, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., Vormoor, K., Bürger, G., Hanel, M., Kriaučiūnienė, J., Loukas, A., Osuch, M. and Yücel, I. (2015) Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrology and Earth System Sciences*, 19, 1827–1847. <https://doi.org/10.5194/hess-19-1827-2015>.
- Switanek, M.B., Troch, P.A., Castro, C.L., Leuprecht, A., Chang, H.-I., Mukherjee, R. and Demaria, E.M.C. (2017) Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences*, 21, 2649–2666. <https://doi.org/10.5194/hess-21-2649-2017>.
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7), 7183–7192.
- Taylor, K., Stouffer, R. and Meehl, G. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498. <https://doi.org/10.1175/BAMS-D-11-00094>.
- Thiemeßl, M.J., Gobiet, A. and Heinrich, G. (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112, 449–468. <https://doi.org/10.1007/s10584-011-0224-4>.
- Vrac, M. and Naveau, P. (2007) Stochastic downscaling of precipitation: from dry events to heavy rainfalls. *Water Resources Research*, 43, W07402. <https://doi.org/10.1029/2006WR005308>.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J. (2014) The inter-Sectoral impact model Intercomparison project (ISI-MIP): project framework. *Proceedings of the National Academy of Sciences*, 111, 3228–3232. <https://doi.org/10.1073/pnas.1312330110>.
- Zappa, G., Hawcroft, M.K., Shaffrey, L., Black, E. and Brayshaw, D.J. (2015) Extratropical cyclones and the projected decline of winter Mediterranean precipitation in the CMIP5 models. *Climate Dynamics*, 45, 1727–1738. <https://doi.org/10.1007/s00382-014-2426-8>.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Casanueva A, Herrera S, Iturbide M, *et al.* Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos Sci Lett*. 2020;21:e978. <https://doi.org/10.1002/asl.978>