



# Harmonized evaluation of daily precipitation downscaled using SDSM and WRF+WRFDA models over the Iberian Peninsula

Santos J. González-Rojí<sup>1,2,3</sup> · Robert L. Wilby<sup>4</sup> · Jon Sáenz<sup>1,5</sup> · Gabriel Ibarra-Berastegi<sup>5,6</sup>

Received: 26 February 2018 / Accepted: 6 February 2019 / Published online: 12 February 2019  
© The Author(s) 2019

## Abstract

There have been numerous statistical and dynamical downscaling model comparisons. However, differences in model skill can be distorted by inconsistencies in experimental set-up, inputs and output format. This paper harmonizes such factors when evaluating daily precipitation downscaled over the Iberian Peninsula by the Statistical DownScaling Model (SDSM) and two configurations of the dynamical Weather Research and Forecasting Model (WRF) (one with data assimilation (D) and one without (N)). The ERA-Interim reanalysis at 0.75° resolution provides common inputs for spinning-up and driving the WRF model and calibrating SDSM. WRF runs and SDSM output were evaluated against ECA&D stations, TRMM, GPCP and EOBS gridded precipitation for 2010–2014 using the same suite of diagnostics. Differences between WRF and SDSM are comparable to observational uncertainty, but the relative skill of the downscaling techniques varies with diagnostic. The SDSM ensemble mean, WRF-D and ERAI have similar correlation scores ( $r = 0.45\text{--}0.7$ ), but there were large variations amongst SDSM ensemble members ( $r = 0.3\text{--}0.6$ ). The best Linear Error in Probability Space (LEPS = 0.001–0.007) and simulations of precipitation amount were achieved by individual members of the SDSM ensemble. However, the Brier Skill Score shows these members do not improve the prediction by ERA-Interim, whereas precipitation occurrence is reproduced best by WRF-D. Similar skill was achieved by SDSM when applied to station or gridded precipitation data. Given the greater computational demands of WRF compared with SDSM, clear statements of expected value-added are needed when applying the former to climate impacts and adaptation research.

**Keywords** Regional climate modelling · Dynamical downscaling · Statistical downscaling · Precipitation · WRF · SDSM · Iberian Peninsula · Data assimilation

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00382-019-04673-9>) contains supplementary material, which is available to authorized users.

✉ Santos J. González-Rojí  
santosjose.gonzalez@ehu.es

- <sup>1</sup> Applied Physics II Department, University of the Basque Country UPV/EHU, Leioa, Bizkaia, Spain
- <sup>2</sup> Present Address: Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
- <sup>3</sup> Present Address: Climate and Environmental Physics, University of Bern, Bern, Switzerland
- <sup>4</sup> Department of Geography, Loughborough University, Leicestershire, United Kingdom
- <sup>5</sup> Joint Research Unit BEGIK, Spanish Institute of Oceanography (IEO), University of the Basque Country, Plentziako Itsas Estazioa, University of the Basque Country UPV/EHU, Plentzia, Bizkaia, Spain
- <sup>6</sup> Fluid Mechanics and N.I. Department, University of the Basque Country UPV/EHU, Bilbao, Bizkaia, Spain

## 1 Introduction

Downscaling methods were developed in the 1990s to improve the spatial resolution of simulations by Global Climate Models (GCMs). The spatial resolution of GCMs is now typically 100–200 km, but some run at 16–40 km resolution thanks to advances in computing power (Davini et al. 2017). Nonetheless, most GCMs are still unable to simulate correctly the leading modes of climate variability (Delworth et al. 2012) or sub-grid scale weather phenomena (Zappa et al. 2013). In order to obtain more accurate local simulations, statistical and dynamical downscaling techniques were developed. The former applies empirical relationships between large-scale atmospheric circulation and local weather variables (von Storch et al. 1993). The latter is based on the numerical integration of equations describing mass, momentum and energy balances in the atmosphere at finer scales than the host GCM (Giorgi 2006; von Storch 2006).

Numerous statistical downscaling techniques have been developed, but most fall into one of three main categories as defined by Maraun et al. (2010). Perfect Prognosis methods apply statistical relationships between the large and local scales to adjust directly the outputs of climate models. Model Output Statistics use direct transfer functions between model and observed data to post-process model output. Weather Generators are stochastic models that modulate the spatial distribution and temporal dependence of local variables depending on large scale atmospheric conditions. Early comparisons of statistical downscaling methods for Europe (Goodess et al. 2007) and North America (Wilby et al. 1998), showed greater skill for mean temperature and precipitation than for their extremes. Those statistical methods applied to precipitation were more skilful at reproducing persistence of wet-spells than the distribution or frequency of rainfall. Additionally, performance tends to be better in winter (Timbal and Jones 2008; Yang et al. 2010) and for mid latitudes (Cavazos and Hewitson 2005), with the correlation coefficients between predictions and observations reaching values around 0.7 for monthly averages and about 0.5 for daily precipitation. For the Iberian Peninsula (henceforth IP), a comparison of different techniques for precipitation downscaling in the Ebro Valley, including machine learning algorithm Random Forest (RF), a Multiple Linear Regression (MLR) model and analogues applied according to a Perfect Prognosis approach, concluded that RF and MLR did not represent any significant improvement over simpler analogues (Ibarra-Berastegi et al. 2011). Improved results were also found by Fernández and Sáenz (2003) by the search of analogues in the space of the canonical correlation coefficients.

Statistical downscaling methods are known to have limitations. Long and accurate meteorological records are needed to calibrate these models and the relationship between large and local scale must be assumed stationary. In addition, downscaled results are ultimately dependent on the quality of inputs from GCMs (Vrac and Vaittinada Ayar 2017). Previous studies show that the Statistical DownScaling Model (SDSM) (Wilby et al. 2014) provides good estimates of daily temperature (Liu et al. 2007), total precipitation (Wetterhall et al. 2006, 2007) and areal rainfall (Hashmi et al. 2011). However, estimates of extreme precipitation may be less reliable for arid climate regimes and/or during dry seasons (Wilby and Dawson 2013).

Dynamical downscaling models nested within a GCM can provide more accurate climate simulations than the coarse global model itself (Jones et al. 1995; Foley 2010; Rummukainen 2010; Feser et al. 2011; Önlol 2012). In these cases, the GCM provides the boundary conditions that drive the regional climate model (RCM). Importantly, even if only boundary conditions are provided to the model, RCMs are able to reproduce small-scale features dependent on the

surface boundary of the domain such as orographic precipitation or large water bodies (Rockel et al. 2008; Leung and Qian 2009). That is mainly achieved by the finer spatial resolution of these models, which is typically less than 25 km.

Nevertheless, RCMs are also known to have limitations, foremost of which is the rising computational cost as the resolution and/or area of the domain increases. This can translate into simulations running for months or even years (depending on processor speeds and storage). The spatial resolution of the RCM domain also determines the weather features that can be simulated by the models, and those that cannot must be parameterized. Such outputs contain significant biases in temperature and precipitation (Varis et al. 2004; Ines and Hansen 2006; Christensen et al. 2008; Teutschbein and Seibert 2012; Turco et al. 2013). For example the Weather Research and Forecasting Model (WRF, ARW) (Skamarock et al. 2008), is known to have a cold bias in summer temperature over the IP (Fernández et al. 2007; Argüeso et al. 2011; Jerez et al. 2012) which can be reduced by the 3DVAR data assimilation scheme (González-Rojí et al. 2018). The spatial variability of precipitation over the IP is well captured by WRF (Cardoso et al. 2013), but there are recognized limitations in some monthly totals (Argüeso et al. 2012).

Comparisons of statistical and dynamical techniques are not straightforward to perform because of their different inputs, spatial (and temporal) resolution, and methods of calibration. Moreover, the outputs of the former are typically point-scale and the latter grid-scale which may conceal true differences in model performance (Tustison et al. 2001). In the past, nearest neighbour or bilinear interpolation techniques have been used to improve comparability. In order to compare highly anisotropic fields such as precipitation, the nearest neighbour technique produces more accurate results than the bilinear interpolation as it does not smooth the fields (Accadia et al. 2003; Casati et al. 2008; Moseley 2011). Ideally, configuration of both downscaling techniques would be as similar as possible taking into account the restrictions associated with each model, including inputs from the same GCM or Reanalysis.

There have been many comparisons between dynamical and statistical downscaling focusing on hydrological applications (Fowler and Wilby 2007). These include assessments of downscaled future projections (Mearns et al. 1999; Hunderca et al. 2016; Onyutha et al. 2016) or current weather conditions (Wilby et al. 2000; Haylock et al. 2006; Huth et al. 2015; Casanueva et al. 2016; Vaittinada Ayar et al. 2016; Roux et al. 2018). These studies show that statistical and dynamical methods have comparable skill at simulating the present climate and should be regarded as complementary tools (Wilby et al. 2000; Haylock et al. 2006; Osma et al. 2015). However, there remains considerable scope for method refinement around the many decisions involved in downscaling model set-up and

simulation. These choices include the source and spatial resolution of the driving boundary conditions used in the dynamical downscaling model; the optimum suite of large-scale variables for statistically downscaling (on a site-by-site basis); the period of record used for model calibration or spin-up; whether or not to use data assimilation (for dynamical downscaling of past climate); and the choice of diagnostics for assessing downscaling model skill and value-added to coarser resolution GCM inputs.

Our objective is to minimize the factors that can distort model-dependent differences in skill between statistical and dynamical downscaling methods. We do this by analysing daily precipitation downscaled over the IP via the Statistical DownScaling Model (SDSM) (<https://www.sdsml.org.uk/>) and two configurations of the dynamical WRF model (one with and one without 3DVAR data assimilation). In order to fairly compare these models, we harmonize the experimental inputs, outputs and diagnostics. In all our experiments, the large-scale variables fed into the downscaling originate from the ERA-Interim reanalysis (Dee et al. 2011), and the nearest neighbour technique is used to identify grid-cells closest to station data. There have been other comparisons between WRF and various statistical downscaling methods (Schmidli et al. 2007; Gutmann et al. 2012; Casanueva et al. 2016) but, as far as the authors are aware, there has been only one previous direct comparison between SDSM and WRF (for China) and, in this case, WRF did not include data assimilation (Tang et al. 2016). The present study assesses the strengths and weaknesses of both downscaling techniques using a range of daily precipitation diagnostics for the IP. Furthermore, we determine whether differences between downscaling techniques exceed differences between observational datasets. Our downscaled precipitation series are either point measurements or have a 15 km resolution, which are finer scales than the reanalysis ( $0.75^\circ$ ) and are comparable to other gridded precipitation datasets such as Spain02 (Herrera et al. 2012, 2016) or SPREAD (Serrano-Notivol et al. 2017), with around 10 km and 5 km of spatial resolution respectively.

This paper is organised in five main parts. In Sect. 2, we introduce the study area then describe the techniques for comparing and validating our downscaling models. Section 3 presents our evaluation of precipitation downscaled by SDSM and WRF for 21 stations using a common validation period (2010–2014). In Sect. 4 we discuss the key insights emerging from the analysis, and in Sect. 5 we conclude with some remarks about the wider implications of the research.

## 2 Study area, data and methodology

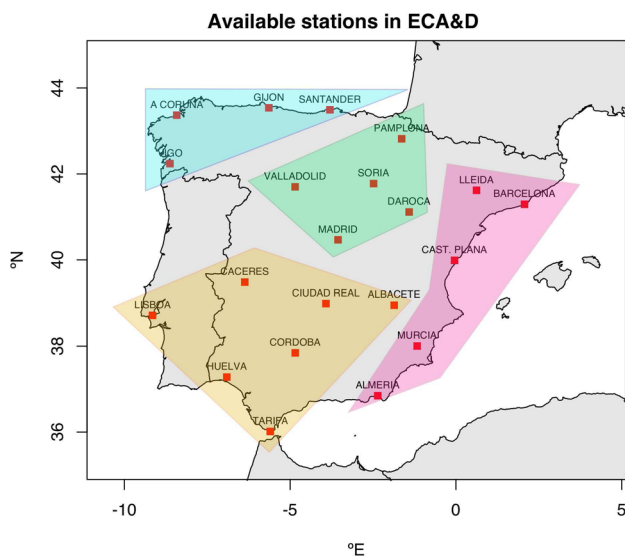
### 2.1 Study area

The precipitation regimes of the IP are influenced by sea level pressure patterns over the Atlantic Ocean and by convective storms that mostly develop in the south-eastern part of Spain (Zorita et al. 1992; Rodríguez-Puebla et al. 1998; Fernández et al. 2003). The IP is a challenging area to undertake regional climate downscaling experiments because of these contrasting rainfall mechanisms and marked topographic gradients which broadly delimit four regimes (Kottek et al. 2006; Peel et al. 2007; Lionello et al. 2012; Rubel et al. 2017): (1) (Semi)Arid climate, in some locations of the Ebro basin and southeastern IP; (2) Mediterranean, in the southwestern IP; (3) Oceanic, located mainly in the northern IP; and (4) Alpine climate, in the mountain ranges. However, inter-annual variability of precipitation cannot be explained entirely by changes in atmospheric circulation; other factors need to be taken into account, including air temperature and humidity (Goodess and Jones 2002). Particularly, European precipitation extremes are influenced by the North Atlantic Oscillation (Haylock and Goodess 2004; Zveryaev et al. 2008) and East Atlantic oscillation (Rodríguez-Puebla et al. 1998; Sáenz et al. 2001; Zveryaev et al. 2008) which modulate the location of Atlantic storm tracks.

### 2.2 Data

Several datasets were used to compare the downscaling techniques:

- ERA-Interim (ERA-Interim) reanalysis atmospheric data at  $0.75^\circ$  horizontal resolution ( $\sim 80$  km) were obtained from the Meteorological Archival and Retrieval System (MARS) repository at ECMWF. These data were used as the boundary conditions for WRF, and for creating the large-scale input variables for SDSM. These data were also used to drive the validation experiments of both downscaling techniques.
- The European Climate Assessment and Dataset project (ECA&D) provides daily precipitation data from land stations (Klein Tank et al. 2002). Daily precipitation occurrence and amount at these sites were used as the local-scale data for calibrating SDSM, as well as for validating model experiments. Twenty-one stations were chosen (red dots in Fig. 1), evenly spaced over the IP and without oversampling any region. Fourteen stations were available for Portugal, but only Lisbon had data during our validation period. Further informa-



**Fig. 1** Location of the 21 stations used in the study highlighted with red squares. The climate regions are Northern, Central, Mediterranean and SouthWestern (shown in blue, green, pink and orange respectively)

tion about the chosen stations can be found in Online Resources 1. Four regions were also defined according to their climate regime: Northern, Central, Mediterranean and SouthWestern regions (in blue, green, pink and orange respectively in Fig. 1). These regions are similar to those defined by Serrano et al. (1999) and reflect the spatial patterns of long-term annual precipitation obtained by Rodríguez-Puebla et al. (1998). The Northern region encompasses the Oceanic climate, the Mediterranean region is related to the (Semi)Arid climate and the SouthWestern region encloses the Mediterranean climate. The Central region is a mix of the Oceanic and Mediterranean climates.

- The ensembles observations (EOBS) Version 12.0 dataset (Haylock et al. 2008; van den Besselaar et al. 2011) was included as a validation dataset. This provides daily precipitation at  $0.25^\circ$  longitude and latitude grid-resolution ( $\sim 20$  km in mid-latitudes). Note that EOBS was built using ECA&D thus, the two datasets are not independent. Additionally, some studies suggest that EOBS is biased toward lower values of precipitation even if the correlations overall are high (Hofstra et al. 2009; Kjellström et al. 2010). For the IP, there are known deficiencies in the south because of data scarcity (Herrera et al. 2012) and near mountainous regions (González-Rojí et al. 2018).
- Tropical Rainfall Measuring Mission (TRMM) data (Wang et al. 2014) were used to evaluate both downscaling techniques. These data are available every 3-h at  $0.25^\circ$  longitude and latitude ( $\sim 20$  km) resolution. In

order to compare with other observational datasets and experiments, TRMM data were aggregated to daily values. Previous studies suggest that TRMM estimates differ from ground-based observations (Nesbitt and Anders 2009; Condom et al. 2011; Hunink et al. 2014), with a precipitation rate-dependent low bias (Huffman et al. 2007; Hashemi et al. 2017).

- Version 2.2 of the Global Precipitation Climatology Project (GPCP) (Huffman et al. 2001) dataset was also included for downscaling model evaluation. Daily precipitation data are available at  $1^\circ$  longitude and latitude grid resolution ( $\sim 100$  km). GPCP reproduces large-scale precipitation patterns, but some errors in the estimates are observed at regional scales (Janowiak et al. 1998), particularly in areas with sparse gauges (Beck et al. 2017).

### 2.3 SDSM set-up

SDSM (version 5.2) (Wilby et al. 2014) is defined as a conditional weather generator, as large-scale atmospheric variables (called predictors) are used to simulate time-varying parameters describing local variables (predictands) specified by the user. Downscaling relationships are hence established between regional predictors from reanalyses or GCMs and local predictands such as daily precipitation (used here) or temperature.

Several steps must be followed to run SDSM. Model calibration begins by selecting locations with observational data, in this case, the 21 stations (in the first experiment) or nearest EOBS grid cells (in the second experiment). Station latitude and longitude determines the closest reanalysis grid cell (herein ERAI), for which a predictor set is extracted. Available predictors include: downward short-wave radiation flux, mean sea level pressure, precipitation, near-surface specific humidity, mean temperature at 2 m and both wind components, wind strength, geopotential height, vorticity, divergence and relative humidity (all at 500 and 850 hPa) (Cavazos 2000; Wilby and Wigley 2000; Schoof and Pryor 2001). Others also use moisture flux (Yang et al. 2010) or total precipitable water (Timbal and Jones 2008) as predictors.

Once the predictand and predictor sets are assembled, relationships between them are explored within the software. A three-step calibration strategy was implemented to minimize subjectivity in the choice of predictor variables. First, any candidate predictor variable with explained variance (R-squared) greater than 0.1 (for each month of the monthly analysis) was short-listed. Second, those with statistically insignificant partial correlation ( $p > 0.05$ ) were eliminated. Finally, the predictor with weakest partial correlation was removed until the point at which only significant variables remain. This implementation of SDSM follows the

procedures customarily applied before (Wilby et al. 1998, 2002; Gulacha and Mulungu 2017). Examples of more specific applications of SDSM can be found elsewhere (Hansen-Bauer et al. 2005; Huth 2005; Crawford et al. 2007; Mahmood and Babel 2013, 2014; Wilby and Dawson 2013).

Having selected the predictors, SDSM is calibrated by ordinary least squares with a fourth-root transformation applied to daily rainfall amounts (to approximate a normal distribution). During simulations, white noise is added to replicate missing variance and thereby generate an ensemble that represents model uncertainty. Here, SDSM was set up to generate a 20-member ensemble for each station, having calibrated the model using data for the period 1979–2009 and validated against 2010–2014.

A second experiment was performed with SDSM to check whether model skill changes when area-average (gridded) precipitation is used for calibration instead of point (station) precipitation. This experiment (henceforth described as SDSM-EOBS) uses data from the closest EOBS cell to the station. Under this arrangement, SDSM-EOBS and WRF now have the same source of downscaling inputs (ERA-Interim), validation period (2010–2014) and similar horizontal scale (gridded data).

## 2.4 WRF model set-up

Two simulations were run using version 3.6.1 of WRF, forced by ERA-Interim. Both begin on 1st of January 2009, but the whole year 2009 was used to spin-up the model and establish correct land-atmosphere fluxes. Hence, only data covering the period 2010–2014 are analysed below.

In the first experiment (WRF-N), boundary conditions drive the model after the initialization. The second experiment (WRF-D) is configured as WRF-N from the point of view of the physical parameterizations. These are: WRF five-class microphysics; RRTMG scheme for both short-wave and long-wave radiation; MYNN2 Planetary Boundary Layer scheme; the Tiedtke scheme for cumulus convection; and the Noah Land Surface Model. However, 3DVAR data assimilation (Barker et al. 2012) is run every six hours (00, 06, 12 and 18 UTC) using observations from NCEP ADP Global Upper Air and Surface Weather Observations (PREPBUFR dataset) inside a 120-minute assimilation window centered on those times. Both experiments use high-resolution sea surface temperature (SST) field NOAA OI SST v2 (Reynolds et al. 2007), which is updated daily.

The background error covariance matrices used in the 3DVAR step of WRF-D were prepared such that they change from month to month. Ninety days computed from 13-months (from January 2007 until February 2008) of WRF 12-h runs (initialized at 00 and 12 UTC) are used to prepare background error covariances for every month. These 12-h runs are centred around the month that the background error

covariance is prepared for, such that 90 days in December, January and February are used to produce the background error covariance for January and so on. The background error covariances are adjusted to the domain and the physical parameterizations used in this study following Parrish and Derber (1992), namely, the cv5 method of WRFDA (Barker et al. 2012).

The WRF-Domain was centred on the IP but covers much of Western Europe and north-west Africa (20°–60°N, 25°W–15°E) (see Fig. 2). The horizontal resolution is 15 × 15 km<sup>2</sup> with 51 vertical levels. Due to the distance between the margins of the domain and the IP, boundary effects from nudging (magenta region in Fig. 2) are discounted (Rummukainen 2010). The mountainous regions of the IP (shown by the GLOBE dataset in Fig. 2) are recognizable in a domain with spatial resolution 15 km, even though topographic effects are still underestimated.

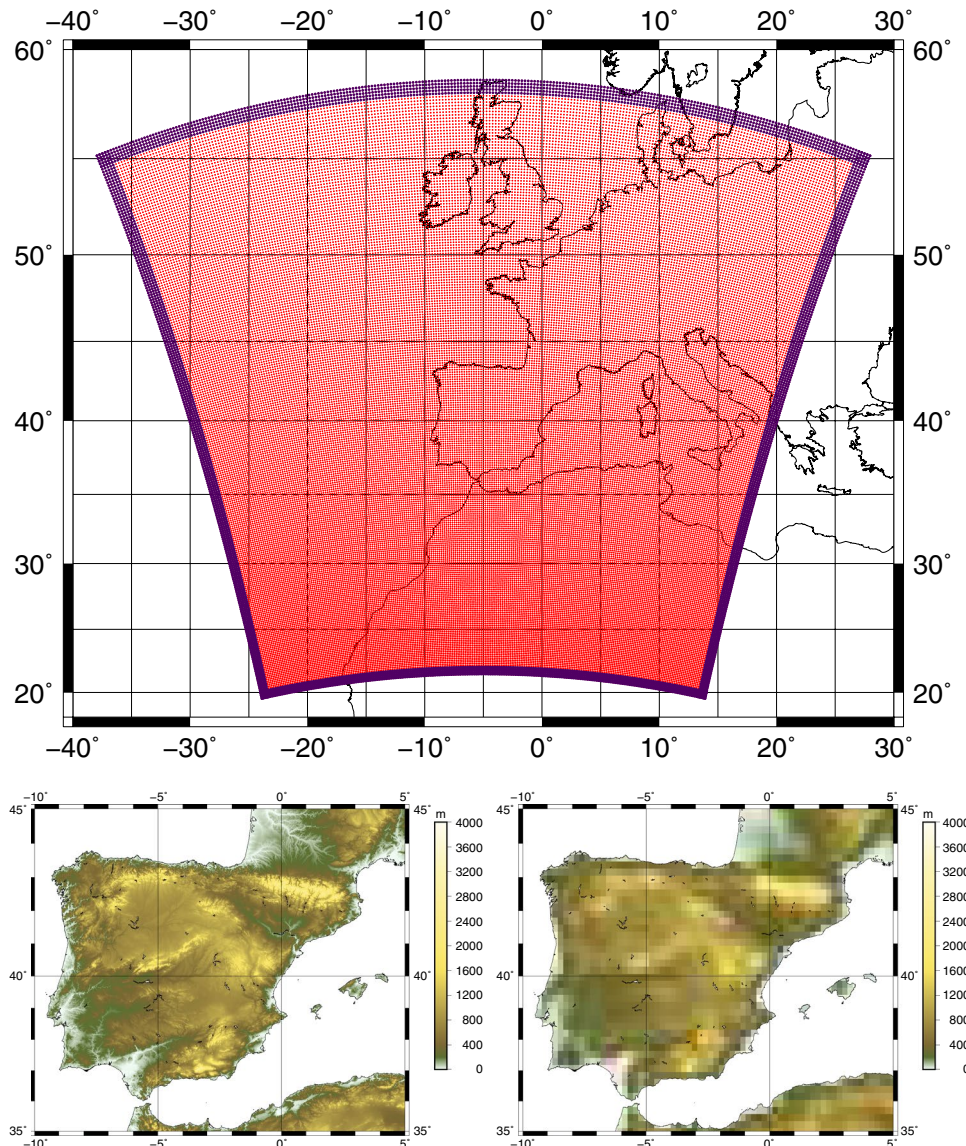
## 2.5 Analytical methods and diagnostics of model skill

The analyses are organised in three parts. First, we examine the calibration of SDSM by evaluating the predictors selected for each station. We assess whether any distinct regional patterns emerge in, for example, the spatial distribution of explained variance in daily precipitation totals. That way, we will be able to check if some stations with the same characteristics can be calibrated in the same way.

Second, precipitation downscaled from ERA-Interim by both SDSM and WRF was compared with the observational datasets described above (i.e. EOBS, TRMM and GPCP). Different sources of precipitation were used to explore downscaling model uncertainty relative to observational uncertainty. This will help us to check how well the downscaling techniques perform in the IP. As these validation datasets are gridded, the nearest cell to the station was selected. Evaluation of time-series against station data was performed using Taylor diagrams (2001), which enable simultaneous, graphical comparison of downscaling model and observational root mean squared error (RSME), Pearson's correlation ( $r$ ) and standard deviation (SD). The statistical reliability of the differences shown by the Taylor diagrams was assessed via bootstrap estimation (not shown). For brevity, Taylor diagrams are shown for representative stations in each climatic region, but all others can be found in the supporting Online Resources. The statistical confidence of the correlations was examined using 1000 bootstrap time series created with replacement and compared with observational data. Distributions of correlations between each bootstrap and observed series are presented as box and whisker plots, again organized by climatic region.

Third, in order to see which downscaling experiments are the best at reproducing the observed precipitation, the

**Fig. 2** Upper panel: domain of the experiments created with WRF at  $15 \times 15$  km resolution (in red). The margins of the domain with nudging are shown in magenta. Lower panel: topography of the IP represented by the GLOBE dataset at 1 km resolution (left) and within the WRF model at 15 km resolution (right)



outputs were compared from the three of them: WRF-D, SDSM-stations and SDSM-EOBS (to assess sensitivity of results to the resolution of the predictand). Only WRF-D was included in this comparison as results from Sect. 3.2 will show that it consistently outperforms the experiment without data assimilation (WRF-N) at reproducing the observed weather. The same result was found by González-Rojí et al. (2018). Diagnostics are derived for each SDSM ensemble member and the ensemble mean in order to compare the deterministic output of WRF (single realization) with the probabilistic output of SDSM (20-member ensemble). The following statistical tests were applied in each case:

- The Linear Error in Probability Space (LEPS) (Ward and Folland 1991) measures how well WRF and SDSM predict the distribution of observed precipitation at

each station. The LEPS score varies between 0 and 1, with 0 representing a perfect model.

- The Brier Skill Score (BSS) shows the value-added by the various downscaling models to precipitation derived from ERAI (our reference model) (Winterfeldt et al. 2011; García-Díez et al. 2015). The BSS was calculated following von Storch and Zwiers (1999) using the error variances of the experiments (i.e. WRF-D simulation and SDSM, the forecast  $F$  on each case) relative to ERAI (the reference  $R$ ):

$$BSS = \begin{cases} 1 - \frac{\sigma_F^2 \sigma_R^{-2}}{\sigma_R^2 \sigma_F^{-2}} & \text{for } \sigma_F^2 \leq \sigma_R^2 \\ \frac{\sigma_F^2}{\sigma_R^2} - 1 & \text{for } \sigma_F^2 > \sigma_R^2 \end{cases} \quad (1)$$

where the error variance is defined as  $\sigma_x^2 = (1/N) \sum_{i=1}^N (x_i - O_i)^2$ ,  $x_i$  are forecast data and  $O_i$  are observed data. The BSS varies between  $-1$  and  $1$  where

positive values indicate that the experiment improves the reference forecast; conversely, negative values signal that the reference performs better than the experiment (i.e. the downscaling does not add value to the forcing model).

- Several diagnostics were derived noting that, because there are only five years of simulations, extreme rainfall can not be assessed. Instead, we have characterized how well WRF and SDSM simulate the following widely used indices (Haylock et al. 2006; Wilby and Yu 2013; Nicholls and Murray 1999):

- Absolute mean daily precipitation (pav)—the average precipitation of all days;
- Wet-day intensity (pint)—the average precipitation on days with more than 1 mm;
- 90th percentile wet-day total (pq90)—the 90th percentile of precipitation on days with more than 1 mm;
- Maximum consecutive dry days (pxcdd)—the number of consecutive days with precipitation less than 1 mm;

- Wet-day probability (pwet)—the number of days with precipitation exceeding 1 mm divided by the number of days of the analysed period.
- Maximum 5-day precipitation total (px5d)—the maximum precipitation total in any five consecutive days.

### 3 Results

#### 3.1 SDSM calibration

We begin this section by presenting the predictor suites created for SDSM at each station (Fig. 3). The most frequently selected predictor variables were: precipitation (PREC, 21 sites), downward shortwave radiation flux (DSWR, 20 sites), 850 and 500 hPa geopotential height (H850, 16 and H500, 10 sites respectively), mean sea level pressure (MSLP) and relative humidity at 500 hPa (R500, 9 sites). Other predictors were selected less frequently: zonal wind at 850 hPa (U850, 5 sites), meridional wind at 500 hPa (V500, 5 sites), zonal wind at 500 hPa and meridional wind at the surface (U500

**Fig. 3** SDSM predictor suite calibrated using inputs from ERAI at 0.75°. The last column is the explained variance in % (rounded to the nearest whole percent) achieved for the calibration period (1979–2009). Acronyms of the predictors follow Wilby and Dawson (2013), and are defined in Online Resource 2

	DSWR	MSLP	USUR	VSUR	DSUR	F500	U500	V500	Z500	H500	D500	F850	U850	V850	Z850	H850	D850	PREC	R500	R850	RSUR	R2
Vigo	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1	39
ACoruna	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	27
Gijon	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	26
Santander	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	22
Pamplona	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	21
Soria	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	18
Valladolid	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	23
Madrid	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	21
Daroca	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	17
Lleida	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	20
Barcelona	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	16
Cast. Plana	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	16
Murcia	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	18
Almeria	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	18
Caceres	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	20
Ciudad Real	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	22
Albacete	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	20
Tarifa	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	24
Huelva	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	26
Cordoba	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	32
Lisbon	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	19

and VSUR, 4 sites), and zonal wind at the surface (USUR, 3 sites). Hence, SDSM uses information about the state of the atmosphere as well as the incident radiation during calibration; variables related to atmospheric moisture are also important.

The most frequently selected predictors vary by region. According to Table 1, downward shortwave radiation flux, precipitation, 850 hPa geopotential height and zonal wind, relative humidity and 500 hPa geopotential height are favoured for the northern region. Downward shortwave radiation, precipitation and relative humidity at 500 hPa are important predictors in the centre of the IP. Radiation, precipitation, zonal wind at 500 hPa and geopotential height at 500 and 850 hPa emerge for the Mediterranean region. Finally, radiation, geopotential height at 850 hPa, precipitation, meridional wind at surface, geopotential height and relative humidity at 500 hPa, and zonal wind at 850 hPa are prominent in the southwest.

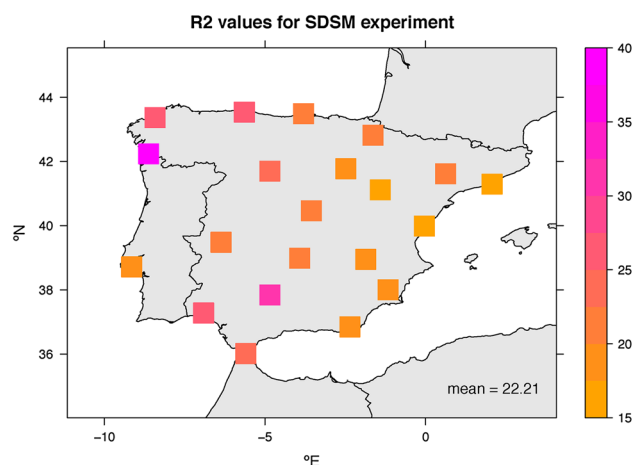
The  $\chi$ -squared test was used to check for dependency of the frequency of predictor variable selection by site latitude, longitude, elevation and annual precipitation during the calibration period 1979–2009. No statistically significant dependencies of predictor suite on site were found. This suggests that the optimum predictor set cannot be inferred from physiographic properties and that each site has to be calibrated on a case-by-case basis. Given the large area of the IP and known spatial variability in precipitation, a single multi-site version of SDSM (Wilby et al. 2003) was also deemed unsuitable.

Having established the predictor suites, the explained variance ( $R^2$ ) during the calibration period is presented in Fig. 4.  $R^2$  varied between 15 and 40 percent with mean value 22%. Vigo and Córdoba achieved 39% and 32% respectively which are comparable to Gulacha and Mulungu (2017) for different periods and regions. Relatively high  $R^2$  were also obtained in the Cantabrian region. Conversely, the lowest  $R^2$  were observed at the Mediterranean coast, particularly near the Ebro basin and Barcelona. Overall, a northwest-southeast gradient in  $R^2$  is evident across the IP which reflects dominance of large scale weather systems near the Atlantic (higher  $R^2$ ) to local convective precipitation near the Mediterranean (lower  $R^2$ ). These results are consistent with those reported by Goodess and Palutikof (1998).

**Table 1** Number of times that each predictor was used for each region

Region	DSWR	MSLP	USUR	VSUR	F500	U500	V500	H500	U850	Z850	H850	PREC	R500	R850	RSUR
North (4)	4	1	1	1	1	2	0	2	2	0	3	4	2	1	1
Centre (5)	5	5	1	0	1	1	3	1	0	1	5	5	3	0	0
Mediterranean (5)	5	1	0	0	0	0	2	3	0	1	2	5	1	0	0
SouthWest (7)	6	2	1	3	0	1	0	4	3	0	6	7	3	0	1

Acronyms are defined in Online Resource 2



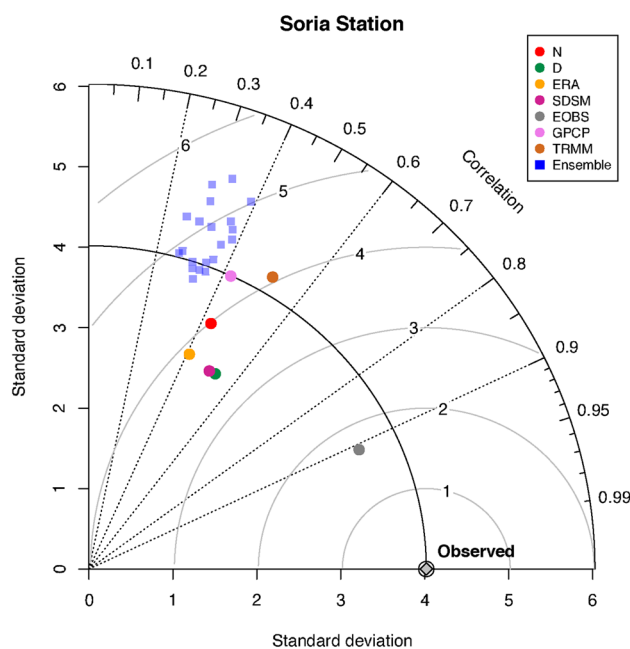
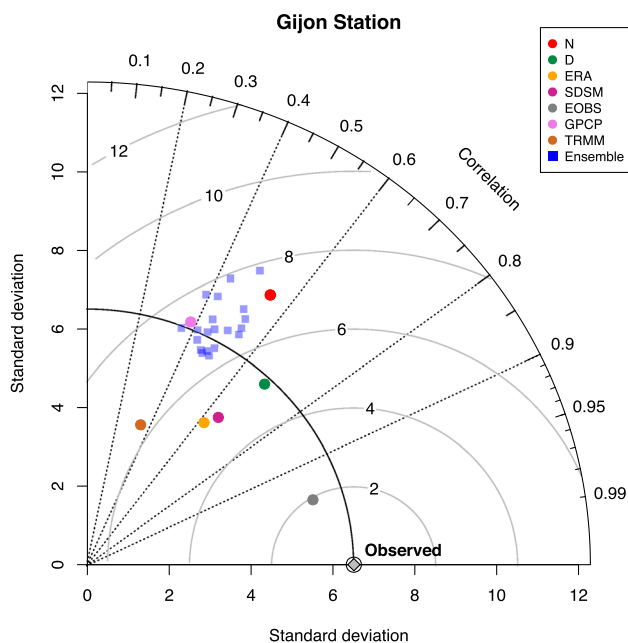
**Fig. 4** Spatial variations in explained variance (in %) achieved by the SDSM for the calibration period (1979–2009)

The predictor suites are similar for SDSM-EOBS except at five stations: Lisbon, Ciudad Real, Lleida, Almería and Vigo. The main differences relate to changes in the height of the variables (near surface, 500 or 850 hPa) or with a reduction in the number of predictors used. Ciudad Real also included the additional predictor VSUR. The  $R^2$  values improve at all stations (except for Vigo) and has mean value 28%. The predictor suites and  $R^2$  values are provided for all sites in Online Resources 3.

### 3.2 Comparison of observational datasets and downscaling experiments

The Northern region has four stations, namely: A Coruna, Vigo, Gijón and Santander. Here, we present only the Taylor diagram for Gijón (Fig. 5); others are available in Online Resource 4. The correlation values for SDSM mean and WRF-D are in the range 0.6 and 0.8. However, even if the SD is quite similar to that observed, WRF-D outperforms the SDSM mean for most sites. Similar results are observed for ERAI and WRF-N for the stations included in this region, but the correlation and the SD are not as good for SDSM and WRF-D. In this case, the correlation ranges between 0.4 and 0.65 and the SD is worse. The EOBS dataset has the best correlation (0.95), compared with TRMM and GPCP (which



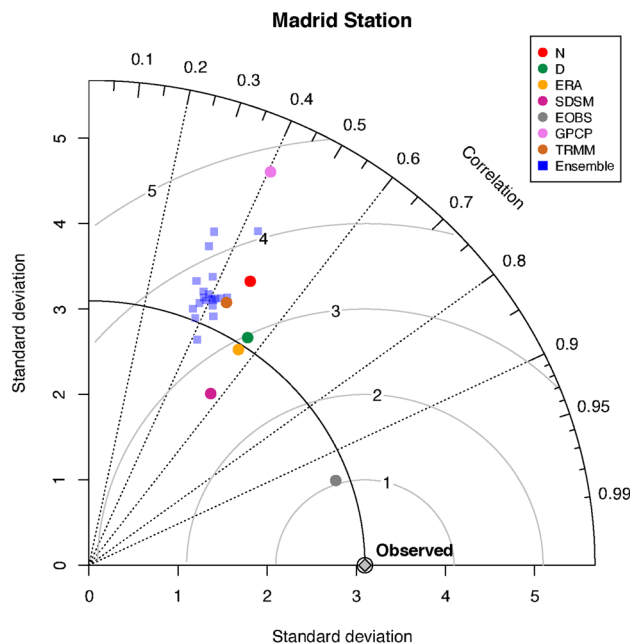


**Fig. 5** Taylor diagram for Gijón Station, Northern Region. The points are as follows: WRF-N (red), WRF-D (green), ERAI (orange), SDSM ensemble mean (magenta), EOBS (grey), GPCP (violet) and TRMM (brown). Observed station data (grey diamond) and individual SDSM ensemble members (blue squares) are also shown. Other Taylor diagrams for this region are available at Online Resource 4

range between 0.3 and 0.4, and with much worse RMSE than the downscaling). SDSM ensemble members overestimate the SD in every station, and have correlation values in the range 0.3–0.5; the SDSM ensemble mean underestimates the SD but presents better correlation and RMSE scores than the ensemble members.

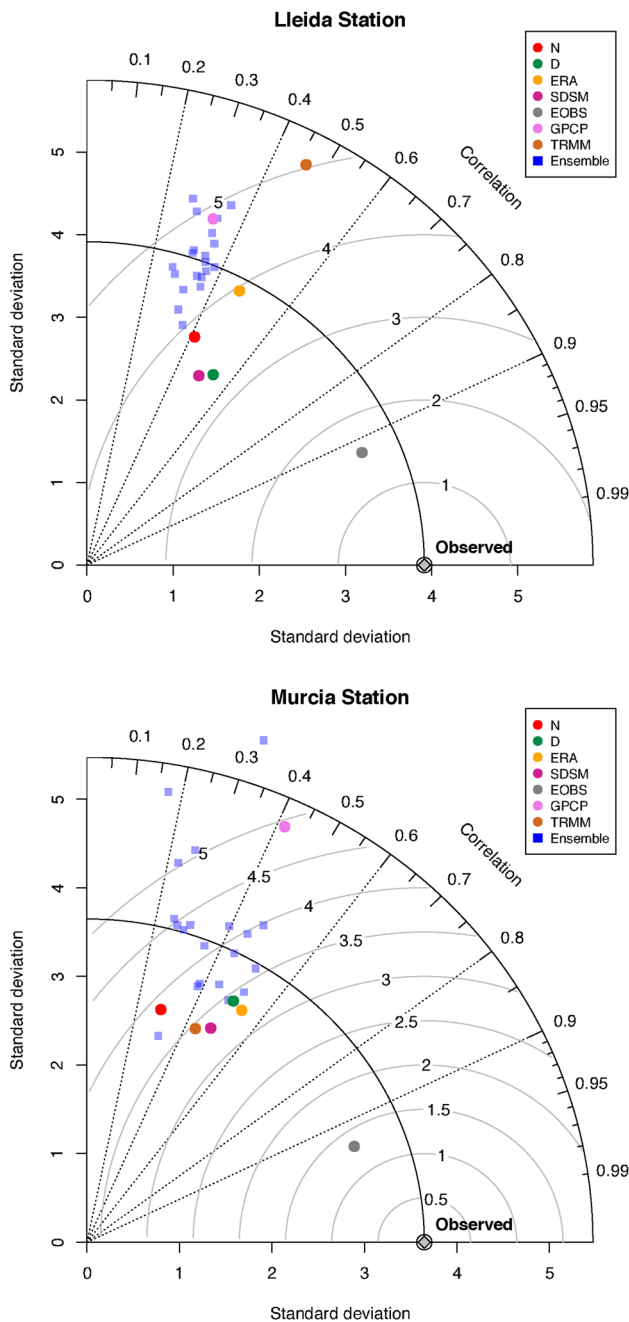
The Central Region has five stations, namely: Pamplona, Soria, Madrid, Valladolid and Daroca. Here we present the results for Soria and Madrid stations (Fig. 6) as representatives of the region; other Taylor diagrams are available via Online Resource 5. Similar correlations emerge for WRF-D, ERAI and SDSM mean data, ranging between 0.5–0.7, 0.4–0.6 and 0.5–0.6, respectively. The SD is underestimated by WRF-D, ERAI and SDSM mean in Soria (and Pamplona), but it is accurately simulated by WRF-D and ERAI in Madrid (and Valladolid and Daroca). The correlation, SD and RMSE obtained by TRMM outperformed GPCP, but those scores are not comparable to those obtained by the downscaling experiments. Again, the EOBS dataset has the highest correlations (above 0.9 for each station). Most members of the SDSM ensemble overestimate the SD and have weaker correlations than the SDSM mean.

The Mediterranean has five stations, namely: Lleida, Barcelona, Murcia, Almería, Castellón de la Plana. Here we present the results for Lleida and Murcia stations (Fig. 7; other Taylor diagrams are available in Online Resource 6).



**Fig. 6** As in Fig. 5 but for Soria (top) and Madrid (bottom) stations, Central Region. Other Taylor diagrams for this region are available at Online Resource 5

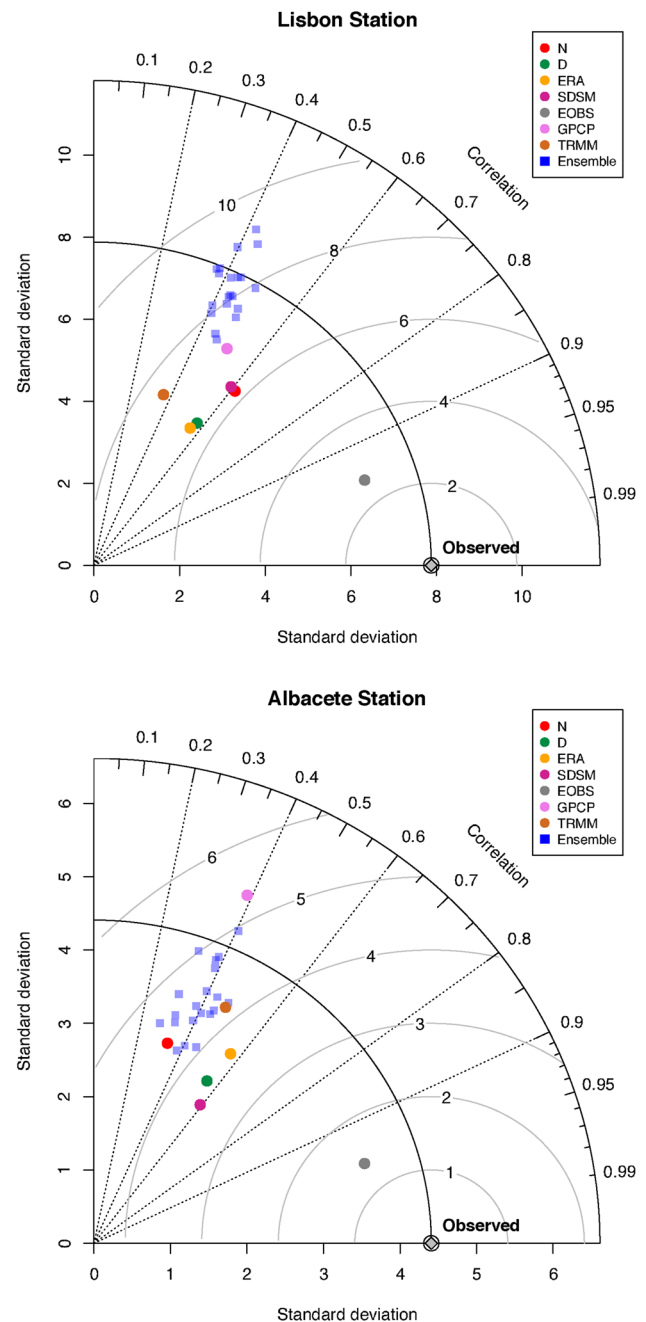
Similar correlations are observed for WRF-D, ERAI and the SDSM mean, with values ranging between 0.4 and 0.6. However, the RMSE is better for WRF-D and the SDSM ensemble mean compared with other datasets. TRMM and GPCP datasets achieved the lowest correlations. The downscaling experiments are always located between these datasets and EOBS in the Taylor diagram. The members of the ensemble reproduced the SD for some stations (Lleida and



**Fig. 7** As in Figs. 5 and 6 but for Lleida (top) and Murcia (bottom) stations, Mediterranean Region. Other Taylor diagrams for this region are available at Online Resource 6

Barcelona), while for others they tend to overestimate it. These correlations range between 0.3 and 0.5.

The Southwestern region has seven stations, namely: Lisbon, Ciudad Real, Cáceres, Albacete, Córdoba, Huelva and Tarifa. Here we present the results for Lisbon and Albacete (Fig. 8); other Taylor diagrams are available from *Online Resource 7*. Correlations for WRF-D, ERAI and the SDSM mean range between 0.5 and 0.6. WRF-D achieves the



**Fig. 8** As in Figs. 5, 6 and 7 but for Lisbon (top) and Albacete (bottom) stations, Southwestern Region. Other Taylor diagrams for this region are available at Online Resource 7

highest correlation of 0.7 in Tarifa (and 0.55 for the others). WRF-N has similar correlation values ( $\sim 0.6$ ) to WRF-D, ERAI and SDSM mean in Lisbon. Most experiments underestimate the observed SD, particularly for Lisbon, Córdoba and Albacete stations. Individual members of the SDSM ensemble have correlations in the range 0.3–0.5, and tend to underestimate the SD. TRMM and GPCP are closer to the results obtained by the rest of the experiments than in

other regions. However, these datasets still obtain the worst results in this region. The EOBS dataset again outperforms the others.

Significance of all the above results was assessed via bootstrap with resampling. Correlation values obtained for the 1000 time-series created for each experiment are shown in Fig. 9. The EOBS dataset achieved the best results for all regions and the other observational datasets (GPCP and TRMM) the worst. Correlations for both WRF experiments, ERAI and the SDSM ensemble mean fall within the range of the observational datasets. The WRF-D experiment achieved higher correlations than WRF-N in each region, but the difference was greatest in the Mediterranean region.

### 3.3 Comparison of downscaling techniques

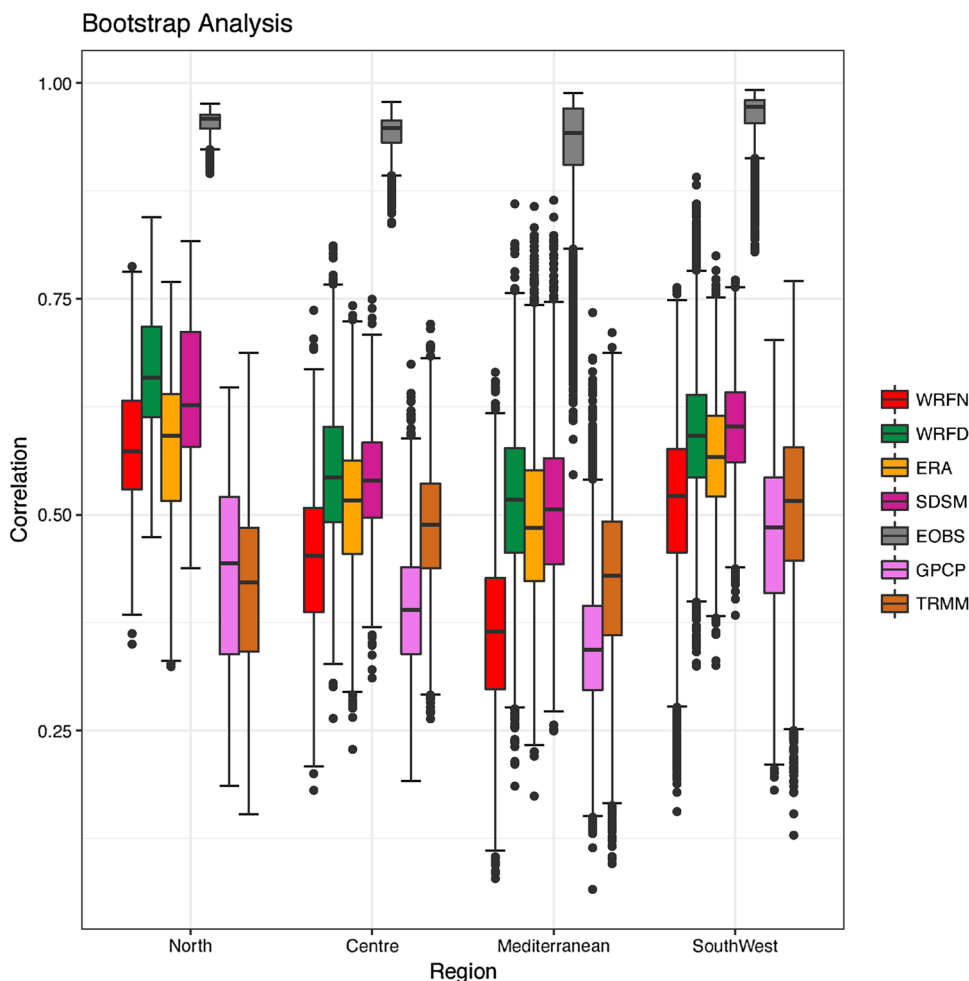
Correlation values for ERAI tend to be lower than those for WRF-D and the SDSM ensemble mean, particularly in the North. However, we cannot differentiate which experiment (WRF-D or SDSM ensemble mean) is best at simulating precipitation over the IP based on these results alone. Thus, other diagnostics are applied to these

experiments, and each member of the SDSM ensemble. Furthermore, we also assess the extent to which the skill of the SDSM mean and ensemble members depends on station or gridded precipitation.

The upper panel of Fig. 10 presents the LEPS for WRF-D, the mean of the SDSM experiments and individual members of both ensembles compared with observed precipitation at each station (D-STAT, SDSM-STAT and EnsSDSM-STAT, SDSM\_EOBS-STAT and EnsSDSM\_EOBS-STAT respectively). Figure 10 shows that WRF-D has superior LEPS scores to the SDSM mean except for Vigo. Conversely, WRF-D is outperformed by individual SDSM members at all stations except Santander, Castellón de la Plana, Tarifa, Córdoba and Lisbon.

The lower panel of Fig. 10 shows similar results for SDSM-EOBS. Again, WRF-D outperforms the SDSM ensemble mean (except for cells overlying Vigo and Ciudad Real). As before, individual SDSM-EOBS members are superior to WRF-D at the majority of sites (with the exception of the same stations before plus A Coruna and Pamplona).

**Fig. 9** Correlations between each experiment/dataset and the corresponding observed precipitation amount. 1000 daily time-series were created by the bootstrap with resampling technique. Results are shown by region. As in previous figures, WRF-N, WRF-D, ERAI, the SDSM ensemble mean, EOBS, GPCP and TRMM are coloured in red, green, orange, magenta, grey, violet and brown respectively



**Fig. 10** Top panel: LEPS scores for WRF-D (green), SDSM mean (magenta) and individual ensemble members (light pink boxes) at each station. Bottom panel: same as the top panel but for the SDSM-EOBS experiment, with the ensemble mean and members in dark and light orange respectively. Station names are color-coded by regions as follows: Northern (cyan), Central (green), Mediterranean (magenta) and Southwestern (orange)

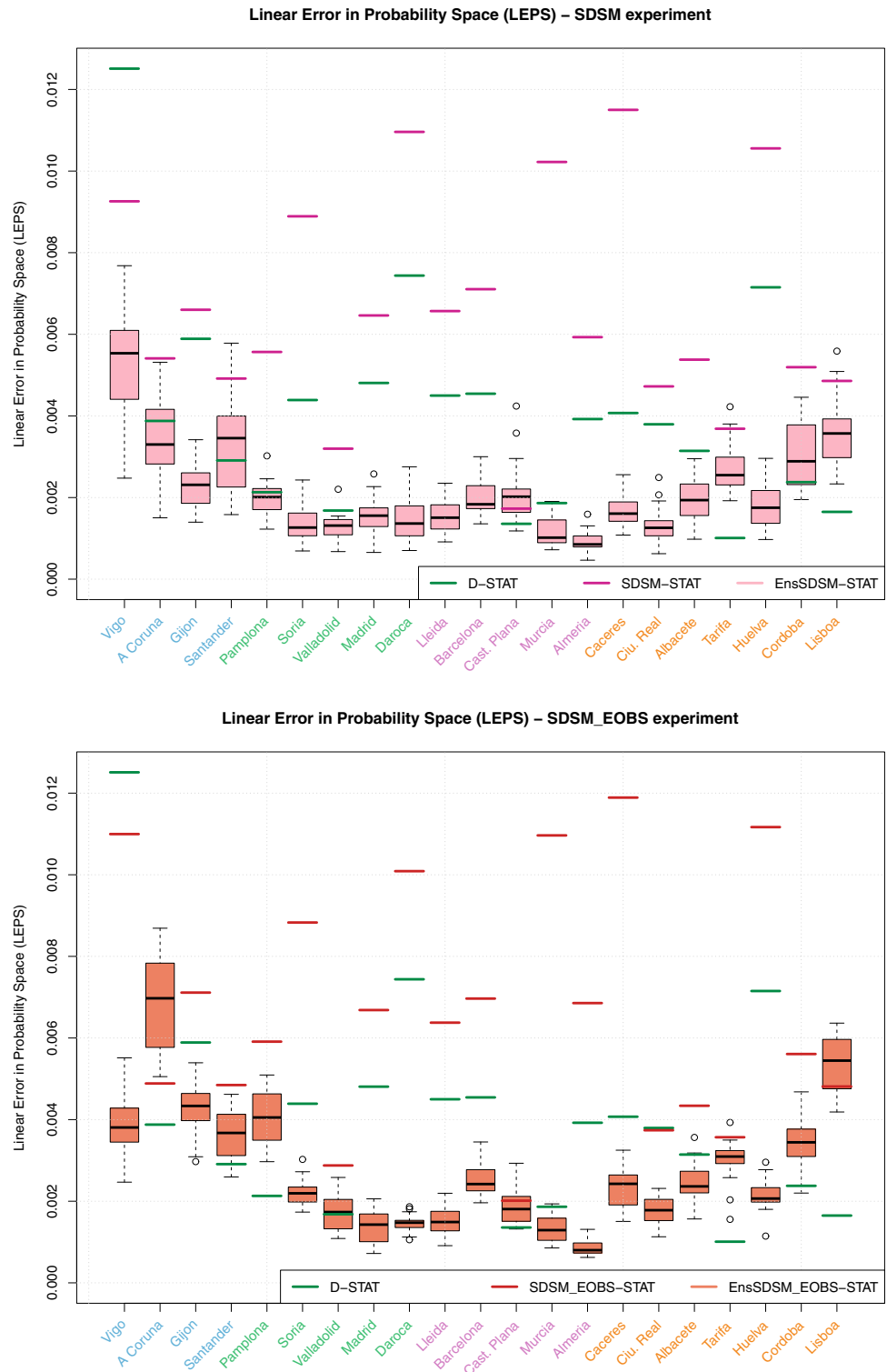
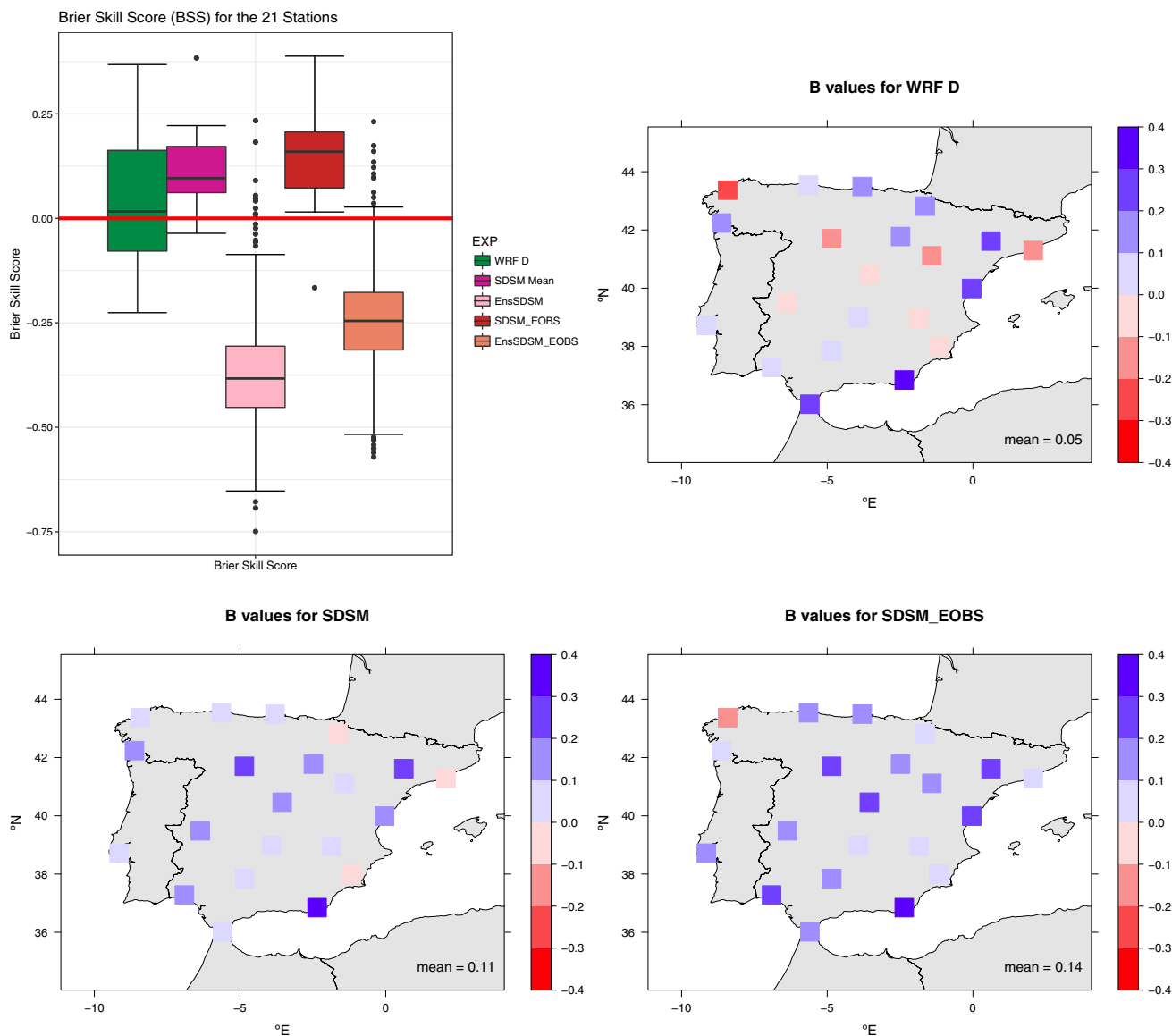


Figure 11 shows the BSS for WRF-D, SDSM means and ensemble members. Both the WRF-D and SDSM means added value compared with ERAI. The average scores were 0.05 for D, 0.11 for the SDSM mean and 0.14 for the mean of SDSM-EOBS. Individual SDSM members do not add value (negative scores).

WRF-D does not add value to ERAI at eight stations: A Coruña (North), Valladolid, Madrid and Daroca (Central), Murcia and Barcelona (Mediterranean), Cáceres and Albacete (Southwest). In comparison, the ensemble mean of SDSM does not add value at three stations: Pamplona (Central), Murcia and Barcelona (Mediterranean). The



**Fig. 11** Top left: BSS of the 21 stations for WRF-D experiment (green), SDSM mean (magenta), SDSM ensemble (pink), SDSM-EOBS mean (red) and SDSM-EOBS ensemble (orange). Spatial vari-

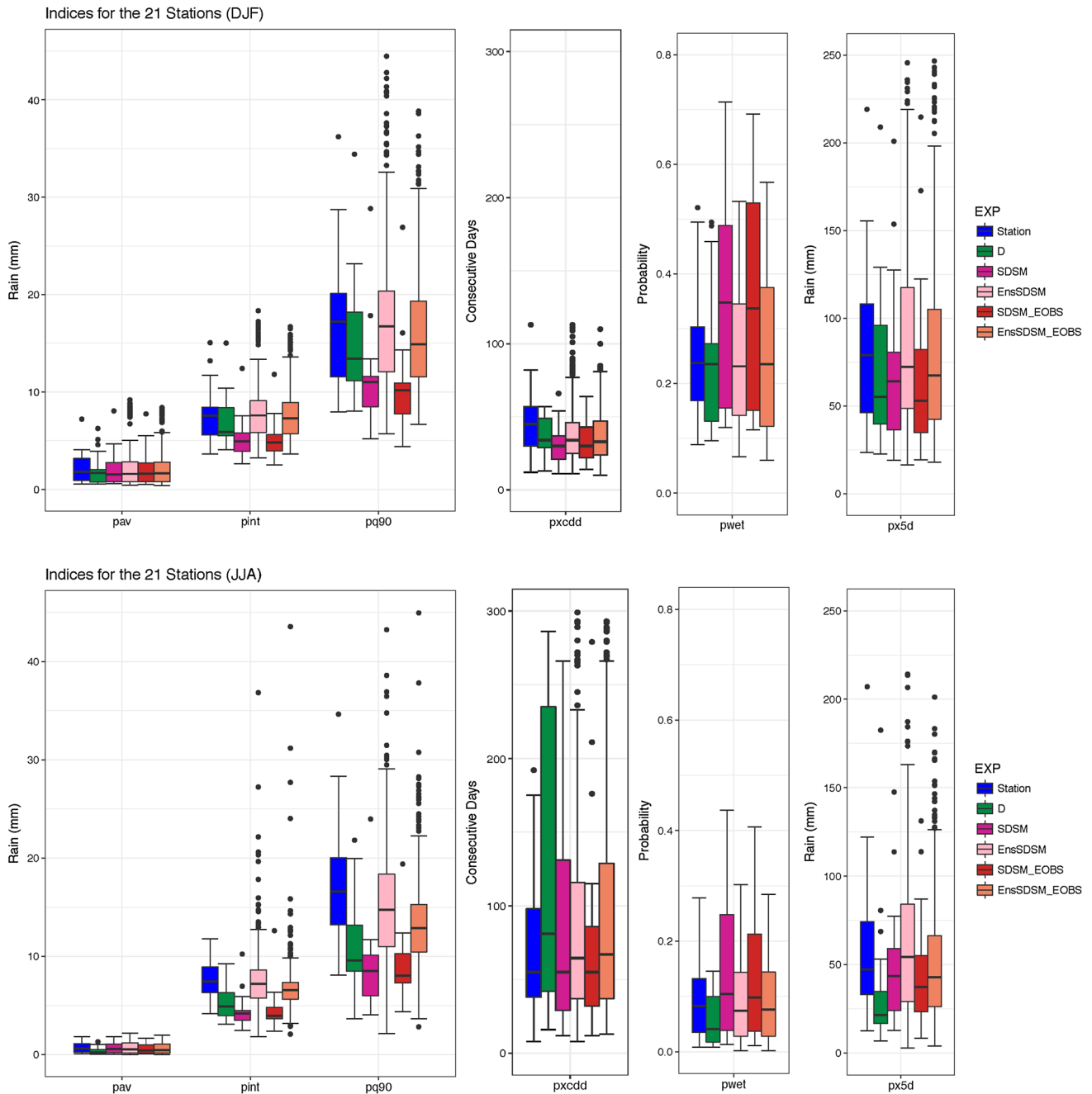
ations in BSS for WRF-D (top right), SDSM mean (lower left) and mean of SDSM-EOBS (bottom right). Corresponding areal means of BSS are displayed in the bottom right corner of each map

SDSM-EOBS mean does not add value for the cell over A Coruna. While WRF-D is able to significantly improve skill for the Mediterranean region (omitting Barcelona and Murcia), the best scores for SDSM are achieved in Almería, Lleida and Valladolid. Remarkable scores are obtained for six stations by SDSM-EOBS. These are those the same as SDSM-stations plus Castellón de la Plana and Madrid. Additionally, the relationship of BSS with the elevation of EOBS and WRF grids was explored, but no significant connection was found between them.

Finally, seasonal precipitation diagnostics were derived for each station, namely: mean precipitation ( $p_{av}$ ), precipitation intensity ( $p_{int}$ ), precipitation 90th quantile ( $p_{q90}$ ),

maximum consecutive dry days ( $pxcdd$ ), wet-day probability ( $p_{wet}$ ) and maximum five-day precipitation total ( $px5d$ ). Figure 12 shows how these compare when based on observed data, WRF-D, SDSM-stations mean, SDSM-EOBS mean and the ensemble members of both SDSM experiments. For brevity, we limit our findings to winter and summer, however, results for spring and autumn are provided in Online Resources 8 and demonstrate similar behaviour to winter and summer respectively.

WRF-D outperforms SDSM when simulating mean precipitation ( $p_{av}$ ) in winter (Fig. 12, top panel). The median value for stations is 1.82 mm, compared with 1.70 mm for WRF-D, 1.52 mm for the SDSM-stations mean, 1.59 mm for



**Fig. 12** Precipitation diagnostics (pav, pint, pq90, pxcdd, pwet and px5d) based on observed data (blue), WRF-D (green), ensemble means and members for SDSM-stations (magenta, light pink), and

SDSM-EOBS (red and orange respectively) during winter (DJF, top) and summer (JJA, bottom). Each box and whisker encloses the results of the chosen 21 stations

the SDSM-stations ensemble, 1.62 for the mean of SDSM-EOBS and 1.64 for the ensemble. In summer, observed pav is lower as expected (0.63 mm for stations) (Fig. 12, lower panel). SDSM still has a dry bias but is now closer (0.44 and 0.59 mm) than WRF-D (0.23 mm).

Precipitation intensity (pint) in winter is simulated well by the members of both SDSM ensembles (with median 7.59 mm for SDSM-stations, 7.27 mm for SDSM-EOBS,

compared with 7.57 mm for stations). WRF-D is too dry (5.88 mm), but less biased than the mean of both SDSM experiments (4.94 mm for SDSM-stations, and 4.83 mm for SDSM-EOBS). In summer, the individual members of both SDSM ensembles obtained similar results to the observations.

The winter 90th percentile of precipitation (pq90) is underestimated by the mean of both SDSM experiments

(with median 11.0 mm for SDSM-stations and 10.2 mm for SDSM-EOBS, compared with 17.2 mm for stations). However, the spread of the members of the SDSM ensembles is similar to observed precipitation (16.7 mm and 14.9 mm respectively). WRF-D also underestimates pq90 (13.4 mm). In summer, pq90 is significantly underestimated by WRF-D and the means of both SDSM experiments. In this case, the SDSM-stations ensemble (14.8 mm) is closer to observations (16.6 mm).

The median maximum consecutive dry days (pxcdd) for the stations is 45 days. All of the experiments underestimate the number of days: WRF-D (34 days), the means of both SDSM experiments (34 and 33 days) and the members of both SDSM ensembles (30 days in both cases). In summer, as expected, pxcdd is larger (55 days) and the behaviour of downscaling models differ from winter. Now the members of both SDSM ensembles overestimate the number of days (65 and 67 days respectively), but not as much as WRF-D (81 days). The SDSM ensemble means match observations.

The mean probability of occurrence of a wet-day (pwet) during winter is 0.24 for stations. WRF-D and the ensemble members of both SDSM experiments agree with observations (0.24 for WRF-D and SDSM-EOBS, 0.23 for SDSM-stations). However, both the ensemble means overestimate pwet (0.35 for SDSM-stations and 0.34 for SDSM-EOBS). In summer, pwet is lower as expected (0.08). Both SDSM ensemble means slightly overestimate the probability (0.10), while WRF-D (0.04) and members of both ensembles (0.07) underestimate pwet.

Finally, the median maximum 5-day precipitation total (px5d) for winter at stations is 79.0 mm. This diagnostic is underestimated by all downscaling models as follows: 76.6 mm (SDSM-stations members) and 68.3 mm (SDSM-EOBS members), 64.2 mm (SDSM-stations mean), 55.1 mm (WRF-D) and 53 mm (SDSM-EOBS mean). In summer, observed px5d is lower (47.3 mm), and the downscaling models exhibit the same pattern of low bias, except for SDSM-stations ensemble members, in which case the bias is positive.

Differences in downscaling skill between winter and summer reflect the contrasting precipitation mechanisms across the IP. Winter precipitation is associated with large frontal systems originating from the Atlantic Ocean (Fernández et al. 2003; Gimeno et al. 2010; Gómez-Hernández et al. 2013), so the probability of precipitation is relatively high (and dry-spells lengths low). Conversely, in summer, there is more convective precipitation (Zorita et al. 1992; Rodríguez-Puebla et al. 1998; Fernández et al. 2003) and the number of consecutive dry days is higher.

The above seasonal variations are reflected in downscaling model skill. WRF-D performs better in winter, when large-scale precipitation is predominant as previously shown by Cardoso et al. (2013). Similarly, SDSM is able

to reproduce the persistence of rain or dry conditions as reported by Goodess et al. (2007). Furthermore, SDSM simulates total precipitation and areal rainfall well as shown by Wetterhall et al. (2007) and Hashmi et al. (2011).

## 4 Discussion

Use of precipitation as a predictor variable for SDSM calibration departs somewhat from convention but is defensible. Dynamical models simulate precipitation based on large-scale and convective processes (such as precipitation from fronts or cumulonimbus clouds). However, neither convective nor microphysical processes are taken into account by other grid-scale predictors, so ERAI precipitation adds important independent information. Additionally, since the predictor suite used in the calibration of SDSM does not suffer from multicollinearity, precipitation is adding explanatory power. Other studies have also shown that precipitation produced by a numerical model can be helpful for statistical downscaling (Schmidli et al. 2006) and that the correlation skill of statistical techniques using precipitation as a predictor yields is improved over conventional downscaling (Widmann et al. 2003). Moreover, Wilks (1992) conditioned the local parameters of a stochastic daily weather generator using monthly precipitation from coarse resolution models, whereas Fealy and Sweeney (2007) designed a statistical technique to predict rainfall occurrence and amount, basing their selection of predictors on strength of correlation with precipitation.

The correlation, SD and RMSE achieved by the statistical and dynamical downscaling models against observed station data lie between those obtained for different observational datasets (EOBS, TRMM and GPCP). This not only happens on a site by site basis, but more generally at the regional level. This shows that, even though dynamically or statistically downscaled precipitation might seem far from perfect according to the correlation or RMSE metrics, they have comparable levels of agreement as between different sources of observational data. This further highlights the uncertainty in precipitation products used to evaluate downscaling methods. Such discrepancies can not simply be attributed to representativeness error because of the horizontal resolution of different datasets. For instance, GPCP and TRMM precipitation estimates involve different satellites, spatial coverage and merging with rain-gauge data, whereas EOBS precipitation estimates are based on gridding of point-based station data.

Dynamical downscaling with data assimilation outperforms the experiment without. Hence, 3DVAR data assimilation improves the quality of the simulations made with WRF, as shown by the Taylor diagrams and the bootstrap analysis (see Figs. 5, 6, 7, 8, 9). This is consistent with Navascués

et al. (2013), Ulazia et al. (2016, 2017) and González-Rojí et al. (2018). However, such simulations are 55% more computationally demanding to perform, so our present set of WRF experiments were limited to five years. Longer simulations would be needed to reliably estimate the skill of WRF at downscaling extreme precipitation events.

The scores of the precipitation diagnostics obtained by both SDSM experiments are similar. Hence, this model is able to simulate realistic estimates of precipitation independently of the resolution of the predictands (i.e. whether point- or grid-based). Our results show that the correlation with observations and RMSE are much improved by the SDSM ensemble mean compared with individual members. Conversely, the SD deteriorates when the ensemble mean is calculated. Average precipitation does not seem to be affected, but there are important differences in the behaviour of the individual members and the mean of both SDSM ensembles for precipitation intensity, 90th percentile, wet-day probability and maximum five-day precipitation.

Since the same source of predictors was used in all our downscaling experiments (ERA-Interim) and gridded output was used to calculate diagnostics, we have undertaken a harmonized evaluation of the dynamical and statistical techniques. According to the results, the most skillful downscaling technique varies according to the choice of diagnostic. Overall, WRF-D produces more consistent results (better scores in the Taylor diagrams and some precipitation diagnostics), but it is clear that SDSM is able to produce comparable (or even better) results to WRF if the ensemble is taken into account. This result agrees with previous research. For example, Schmidli et al. (2007) reported that over flat terrain, both downscaling techniques showed similar results. Gutmann et al. (2012) compared a statistical model with WRF in the mountainous regions of Colorado, and found that statistical downscaling was able to improve the results of the original model. Casanueva et al. (2016) compared eight RCMs with five statistical downscaling methods over continental Spain, and found that statistical methods outperformed RCMs in terms of seasonal mean precipitation.

Data assimilation is not an option (due to the lack of observations) for climate change or seasonal forecast applications of WRF. Rather, only N-type WRF simulations can be performed. Additionally, in these applications, the correlation coefficient is not a fair verification index, and, as shown by Figs. 10 and 12, if the temporal occurrence of precipitation is not taken into account members of the SDSM ensemble perform better than WRF-D.

Another key-aspect to consider when one of the downscaling techniques must be chosen is the time and computational cost. The calibration step of SDSM is the most demanding and time-consuming part of this downscaling technique. Even so, the time expended on this task is modest compared with the investment require to run WRF.

Moreover, the storage required for all the sites downscaled by SDSM was tiny (< 100 Mb) when compared with each run of WRF for the whole domain of the IP. The storage needed by the raw data for WRF-N is 2 Tb and 4.7 Tb for WRF-D. On the other hand, WRF provides the spatial distribution of simulated precipitation (and many other variables) over the whole domain compared with just 21 stations (or EOBS cells) from SDSM.

Finally, it is important to note that the selection of SDSM predictors (including precipitation) and high resolution gridded inputs means that the statistical downscaling is now more closely approximating the ingredients of the dynamical model. We have statistically downscaled ERA-Interim data using local information from EOBS gridded data with SDSM and found that the results are comparable to those obtained with local information from station data. We see a growing opportunity for deploying both downscaling techniques in combination (Díez et al. 2005; Fernández-Ferrero et al. 2009). For example, dynamical modelling could give the spatial coverage and persistence, whereas statistical downscaling could improve the precision of highly local weather metrics (Roux et al. 2018).

## 5 Conclusions

This study shows how more consistent evaluation of statistical and dynamical downscaling techniques can be achieved by minimizing differences in model set-up, inputs and outputs. Moreover, the observational datasets against which downscaling techniques are routinely compared should not be considered as absolute *truth*.

We evaluated daily precipitation simulated by two downscaling techniques: WRF (boundary forcing by ERA-Interim, with and without 3DVAR data assimilation) and SDSM (using predictors also from ERA-Interim fit to station and gridded precipitation). By harmonizing the inputs (ERA-Interim), validation period (2010–2014), output format (EOBS gridded precipitation) and diagnostics, the remaining differences can be more confidently attributed to the downscaling techniques themselves rather than experimental set-up. The complex landscapes and climate regimes of Iberia provide a diverse and challenging set of conditions for testing both models.

We confirm that SDSM must be calibrated on a site by site basis - exploration of the predictor sets in relation to station elevation, latitude or longitude found no coherent regional patterns or gradients. Having followed consistent calibration procedures, the statistically downscaled series were no more different to WRF series, than variations between observational datasets (EOBS, TRMM and GPCP). Our results also showed that WRF-D (with data assimilation) yields more skillful precipitation than WRF-N (without assimilation). The comparison of our experiments showed



that no downscaling technique was superior across all verification metrics. According to our results, comparable correlations are obtained for the SDSM ensemble means, WRF-D and ERAI for the regions studied. However, individual members of the SDSM ensembles, were generally less skillful than WRF-D. Focusing on the precipitation metrics, the skill of SDSM was similar whether station or gridded precipitation was used for calibration. Thus, the different spatial scales involved in the station versus gridded precipitation data problem do not apparently play an important role, at least at the 20 km range checked in this study.

The best LEPS values were achieved by SDSM ensemble members at most stations (16 for SDSM-stations and 13 for SDSM-EOBS). Both WRF-D and the members of the SDSM ensemble outperformed the SDSM mean in all cases. In contrast, the BSS showed that SDSM ensemble means and WRF-D added value to the prediction of precipitation when compared with unadjusted ERAI. This was not the case for the SDSM ensemble members. This is because the BSS index accounts for the temporal occurrence of precipitation, whilst SDSM stochastically generates individual series that are not expected to match observed series.

WRF-D outperformed SDSM when simulating winter daily mean precipitation, whereas SDSM ensemble members were most skillful for precipitation intensity and the 90th percentile. This was also the case in summer. WRF-D and SDSM ensemble members reproduced observed maximum consecutive dry-days and the probability of a wet-day in winter. Maximum 5-days precipitation totals were underestimated by both downscaling techniques. In summer, WRF-D overestimated maximum consecutive dry-days and underestimated the probability of a wet-day. Both SDSM experiments behaved similarly to winter. According to this seasonal analysis, we conclude that SDSM can outperform WRF-D when simulating these indices, but WRF-D presents more consistent results between seasons.

Ultimately, these downscaled products could be useful for analysis of the long-term water balance of western Europe. SDSM generates similar results to WRF at grid scales, but can simulate weather variables for much longer periods, at lower computational costs and in considerably less time. Hence, clear statements about the expected value-added are needed when applying WRF to climate impacts and adaptation research. However, further research is needed to explore the extent to which different types of precipitation mechanism (e.g. intense local convection or widespread frontal system) are reproduced as well by SDSM as by WRF, rather than pooling all rain days as in the present study. There is also scope for detailed analysis of multi-season extremes, or simulations of inter-annual variability.

**Acknowledgements** This work was supported by Grants (CGL2013-45198-C2-1-R and CGL2016-76561-R, MINECO/ERDF, UE).

S.J.G.R. was supported by a FPI Postdoctoral Research Grant (MINECO, BES-2014-069977). Additional funding was received from the University of the Basque Country (UPV/EHU, GIU 17/002). Intensive computational resources used in the project were provided by I2BASQUE. The authors thank the creators of the WRF/ARW and WRFDA systems for making them freely available to the community. The authors acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES and the data providers in the ECA&D project. NOAA\_OI\_SST\_V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. Comments by two anonymous reviewers have significantly improved the manuscript.

**OpenAccess** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Accadia C, Mariani S, Casaioli M, Lavagnini A, Speranza A (2003) Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Weather Forecast* 18(5):918–932. [https://doi.org/10.1175/1520-0434\(2003\)018%3c\\$0918:SOFPS\\$3e\\$2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018%3c$0918:SOFPS$3e$2.0.CO;2)
- Argüeso D, Hidalgo-Muñoz JM, Gámiz-Fortis SR, Esteban-Parra MJ, Dudhia J, Castro-Díez Y (2011) Evaluation of WRF parameterizations for climate studies over southern Spain using a multi-step regionalization. *J Clim* 24(21):5633–5651. <https://doi.org/10.1175/JCLI-D-11-00073.1>
- Argüeso D, Hidalgo-Muñoz JM, Gámiz-Fortis SR, Esteban-Parra MJ, Castro-Díez Y (2012) Evaluation of WRF mean and extreme precipitation over Spain: present climate (1970–99). *J Clim* 25(14):4883–4897. <https://doi.org/10.1175/JCLI-D-11-00276.1>
- Barker D, Huang XY, Liu Z, Auligné T, Zhang X, Rugg S, Ajjaji R, Bourgeois A, Bray J, Chen Y, Demirtas M, Guo YR, Henderson T, Huang W, Lin HC, Michalakes J, Rizvi S, Zhang X (2012) The Weather Research and Forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bull Am Meteorol Soc* 93(6):831–843. <https://doi.org/10.1175/BAMS-D-11-00167.1>
- Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk AIJM, Weedon GP, Brocca L, Pappenberger F, Huffman GJ, Wood EF (2017) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol Earth Syst Sci* 21(12):6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>
- van den Besselaar EJM, Haylock MR, van der Schrier G, Klein Tank AMG (2011) A European daily high-resolution observational gridded data set of sea level pressure. *J Geophys Res Atmos* 116:D11. <https://doi.org/10.1029/2010JD015468>
- Cardoso RM, Soares PMM, Miranda PMA, Belo-Pereira M (2013) WRF high resolution simulation of Iberian mean and extreme precipitation climate. *Int J Climatol* 33(11):2591–2608. <https://doi.org/10.1002/joc.3616>
- Casanueva A, Herrera S, Fernández J, Gutiérrez J (2016) Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative. *Clim Change* 137(3):411–426. <https://doi.org/10.1007/s10584-016-1683-4>
- Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocerlich M, Damrath U, Ebert EE, Brown BG, Mason S (2008) Forecast

- verification: current status and future directions. *Meteorol Appl* 15(1):3–18. <https://doi.org/10.1002/met.52>
- Cavazos T (2000) Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans. *J Clim* 13(10):1718–1732. [https://doi.org/10.1175/1520-0442\(2000\)013\\$%3c\\$1718:USOMTIS%3e\\$2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013$%3c$1718:USOMTIS%3e$2.0.CO;2)
- Cavazos T, Hewitson BC (2005) Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. *Clim Res* 28(2):95–107. <https://doi.org/10.3354/cr028095>
- Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys Res Lett* 35:20. <https://doi.org/10.1029/2008GL035694>
- Condom T, Rau P, Espinoza JC (2011) Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2007. *Hydrol Process* 25(12):1924–1933. <https://doi.org/10.1002/hyp.7949>
- Crawford T, Betts NL, Favis-Mortlock D (2007) GCM grid-box choice and predictor selection associated with statistical downscaling of daily precipitation over Northern Ireland. *Clim Res* 34(2):145–160. <https://doi.org/10.3354/cr034145>
- Davini P, von Hardenberg J, Corti S, Christensen HM, Juricke S, Subramanian A, Watson PAG, Weisheimer A, Palmer TN (2017) Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geosci Mod Dev* 10(3):1383–1402. <https://doi.org/10.5194/gmd-10-1383-2017>
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. <https://doi.org/10.1002/qj.828>
- Delworth TL, Rosati A, Anderson W, Adcroft AJ, Balaji V, Benson R, Dixon K, Griffies SM, Lee HC, Pacanowski RC, Vecchi GA, Wittenberg AT, Zeng F, Zhang R (2012) Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J Clim* 25(8):2755–2781. <https://doi.org/10.1175/JCLI-D-11-00316.1>
- Díez E, Primo C, García-Moya JA, Gutiérrez JM, Orfila B (2005) Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus A* 57(3):409–423. <https://doi.org/10.1111/j.1600-0870.2005.00130.x>
- Fealy R, Sweeney J (2007) Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. *Int J Climatol* 27(15):2083–2094. <https://doi.org/10.1002/joc.1506>
- Fernández J, Sáenz J (2003) Improved field reconstruction with the analog method: searching the CCA space. *Clim Res* 24(3):199–213. <https://doi.org/10.3354/cr024199>
- Fernández J, Sáenz J, Zorita E (2003) Analysis of wintertime atmospheric moisture transport and its variability over Southern Europe in the NCEP reanalyses. *Clim Res* 23(3):195–215. <https://doi.org/10.3354/cr023195>
- Fernández J, Montávez JP, Sáenz J, González-Rouco JF, Zorita E (2007) Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: annual cycle. *J Geophys Res Atmos* 112:D4. <https://doi.org/10.1029/2005JD006649>
- Fernández-Ferrero A, Sáenz J, Ibarra-Berastegi G, Fernández J (2009) Evaluation of statistical downscaling in short range precipitation forecasting. *Atmos Res* 94(3):448–461. <https://doi.org/10.1016/j.atmosres.2009.07.007>
- Feser F, Rockel B, von Storch H, Winterfeldt J, Zahn M (2011) Regional climate models add value to global model data: a review and selected examples. *Bull Am Meteorol Soc* 92(9):1181–1192. <https://doi.org/10.1175/2011BAMS3061.1>
- Foley A (2010) Uncertainty in regional climate modelling: a review. *Progress Phys Geogr* 34(5):647–670. <https://doi.org/10.1177/0309133310375654>
- Fowler HJ, Wilby RL (2007) Beyond the downscaling comparison study. *Int J Climatol* 27(12):1543–1545. <https://doi.org/10.1002/joc.1616>
- García-Díez M, Fernández J, San-Martín D, Herrera S, Gutiérrez JM (2015) Assessing and improving the local added value of WRF for wind downscaling. *J Appl Meteorol Climatol* 54(7):1556–1568. <https://doi.org/10.1175/JAMC-D-14-0150.1>
- Gimeno L, Nieto R, Trigo RM, Vicente-Serrano SM, López-Moreno JJ (2010) Where does the Iberian Peninsula moisture come from? An answer based on a Lagrangian approach. *J Hydrometeorol* 11(2):421–436. <https://doi.org/10.1175/2009JHM1182.1>
- Giorgi F (2006) Regional climate modeling: status and perspectives. *J Phys IV (Proc)* 139:101–118. <https://doi.org/10.1051/jp4:2006139008>
- Gómez-Hernández M, Drumond A, Gimeno L, García-Herrera R (2013) Variability of moisture sources in the Mediterranean region during the period 1980–2000. *Water Resour Res* 49:6781–6794. <https://doi.org/10.1002/wrcr.20538>
- González-Rojí SJ, Sáenz J, Ibarra-Berastegi G, Díaz de Argandoña J (2018) Moisture balance over the Iberian Peninsula according to a regional climate model: the impact of 3DVAR data assimilation. *J Geophys Res Atmos* 123(2):708–729. <https://doi.org/10.1002/2017JD027511>
- Goodess C, Anagnostopoulou C, Bárdossy A, Frei C, Harpham C, Haylock MR, Hurrell Y, Maheras P, Ribalaygua J, Schmidli J, Schmith T, Tolika K, Tomozeiu R, Wilby RL (2007) An inter-comparison of statistical downscaling methods for Europe and European regions—assessing their performance with respect to extreme temperature and precipitation events. *Clim Change*
- Goodess CM, Jones PD (2002) Links between circulation and changes in the characteristics of Iberian rainfall. *Int J Climatol* 22(13):1593–1615. <https://doi.org/10.1002/joc.810>
- Goodess CM, Palutikof JP (1998) Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *Int J Climatol* 18(10):1051–1083. [https://doi.org/10.1002/\(SICI\)1097-0088\(199808\)18:10%3c1051::AID-JOC304%3e3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0088(199808)18:10%3c1051::AID-JOC304%3e3.0.CO;2-1)
- Gulacha MM, Mulungu DM (2017) Generation of climate change scenarios for precipitation and temperature at local scales using SDSM in Wami-Ruvu River Basin Tanzania. *Phys Chem Earth Parts A/B/C* 100:62–72. <https://doi.org/10.1016/j.pce.2016.10.003>
- Gutmann ED, Rasmussen RM, Liu C, Ikeda K, Gochis DJ, Clark MP, Dudhia J, Thompson G (2012) A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J Clim* 25(1):262–281. <https://doi.org/10.1175/2011JCLI4109.1>
- Hanssen-Bauer I, Achberger C, Benestad RE, Chen D, Førland EJ (2005) Statistical downscaling of climate scenarios over Scandinavia. *Clim Res* 29(3):255–268. <https://doi.org/10.3354/cr029255>
- Hashemi H, Nordin M, Lakshmi V, Huffman GJ, Knight R (2017) Bias correction of long-term satellite monthly precipitation product (TRMM 3B43) over the conterminous United States. *J Hydrometeorol* 18(9):2491–2509. <https://doi.org/10.1175/JHM-D-17-0025.1>

- Hashmi MZ, Shamseldin AY, Melville BW (2011) Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. *Stoch Environ Res Risk Assess* 25(4):475–484. <https://doi.org/10.1007/s00477-010-0416-x>
- Haylock MR, Goodess CM (2004) Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. *Int J Climatol* 24(6):759–776. <https://doi.org/10.1002/joc.1033>
- Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM (2006) Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *Int J Climatol* 26(10):1397–1415. <https://doi.org/10.1002/joc.1318>
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J Geophys Res Atmos* 113:D20. <https://doi.org/10.1029/2008JD010201>
- Herrera S, Gutiérrez JM, Ancell R, Pons MR, Frías MD, Fernández J (2012) Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *Int J Climatol* 32(1):74–85. <https://doi.org/10.1002/joc.2256>
- Herrera S, Fernández J, Gutiérrez JM (2016) Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology. *Int J Climatol* 36(2):900–908. <https://doi.org/10.1002/joc.4391>
- Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *J Geophys Res Atmos* 114:D21. <https://doi.org/10.1029/2009JD011799>
- Huffman GJ, Adler RF, Morrissey MM, Bolvin DT, Curtis S, Joyce R, McGavock B, Susskind J (2001) Global precipitation at one-degree daily resolution from multisatellite observations. *J Hydrometeorol* 2(1):36–50. [https://doi.org/10.1175/1525-7541\(2001\)002%3c0036:GPAODD%3e2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002%3c0036:GPAODD%3e2.0.CO;2)
- Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF (2007) The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J Hydrometeorol* 8(1):38–55. <https://doi.org/10.1175/JHM560.1>
- Hundecha Y, Sunyer MA, Lawrence D, Madsen H, Willems P, Bürger G, Kriaučiūniene J, Loukas A, Martinkova M, Osuch M, Vasilias L, von Christierson B, Vormoor K, Yücel I (2016) Inter-comparison of statistical downscaling methods for projection of extreme flow indices across Europe. *J Hydrol* 541:1273–1286. <https://doi.org/10.1016/j.jhydrol.2016.08.033>
- Hunink J, Immerzeel W, Droogers P (2014) A high-resolution precipitation 2-step mapping procedure (HiP2P): development and application to a tropical mountainous area. *Remote Sens Environ* 140:179–188. <https://doi.org/10.1016/j.rse.2013.08.036>
- Huth R (2005) Downscaling of humidity variables: a search for suitable predictors and predictands. *Int J Climatol* 25(2):243–250. <https://doi.org/10.1002/joc.1122>
- Huth R, Mikšovský J, Štěpánek P, Belda M, Farda A, Chládková Z, Pišoft P (2015) Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature. *Theor Appl Climatol* 120(3):533–553. <https://doi.org/10.1007/s00704-014-1190-3>
- Ibarra-Berastegi G, Saénz J, Ezcurra A, Elías A, Diaz Argandoña J, Errasti I (2011) Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression. *Hydrol Earth Syst Sci* 15(6):1895–1907. <https://doi.org/10.5194/hess-15-1895-2011>
- Ines AV, Hansen JW (2006) Bias correction of daily GCM rainfall for crop simulation studies. *Agric Forest Meteorol* 138(1):44–53. <https://doi.org/10.1016/j.agrformet.2006.03.009>
- Janowiak JE, Gruber A, Kondragunta CR, Livezey RE, Huffman GJ (1998) A comparison of the NCEP-NCAR reanalysis precipitation and the GPCP rain gauge-satellite combined dataset with observational error considerations. *J Clim* 11(11):2960–2979. [https://doi.org/10.1175/1520-0442\(1998\)011<2960:ACOTNN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2960:ACOTNN>2.0.CO;2)
- Jerez S, Montávez JP, Gómez-Navarro JJ, Jiménez PA, Jiménez-Guerrero P, Lorente R, González-Rouco JF (2012) The role of the land-surface model for climate change projections over the Iberian Peninsula. *J Geophys Res Atmos* 117:D1. <https://doi.org/10.1029/2011JD016576>
- Jones RG, Murphy JM, Noguer M (1995) Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Q J R Meteorol Soc* 121(526):1413–1449. <https://doi.org/10.1002/qj.49712152610>
- Kjellström E, Boberg F, Castro M, Christensen H, Nikulin G, Sánchez E (2010) Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Clim Res* 44(2–3):135–150. <https://doi.org/10.3354/cr00932>
- Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Antonio López J, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int J Climatol* 22(12):1441–1453. <https://doi.org/10.1002/joc.773>
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen–Geiger climate classification updated. *Meteorol Z* 15(3):259–263. <https://doi.org/10.1127/0941-2948/2006/0130>
- Leung LR, Qian Y (2009) Atmospheric rivers induced heavy precipitation and flooding in the western US simulated by the WRF regional climate model. *Geophys Res Lett* 36:3. <https://doi.org/10.1029/2008GL036445>
- Lionello P, Gacic M, Gomis D, Garcia-Herrera R, Giorgi F, Planton S, Trigo R, Theocharis A, Tsimplis MN, Ulbrich U, Xoplaki E (2012) Program focuses on climate of the Mediterranean region. *Eos Trans Am Geophys Union* 93(10):105–106. <https://doi.org/10.1029/2012EO100001>
- Liu X, Coulibaly P, Evora N (2007) Comparison of data-driven methods for downscaling ensemble weather forecasts. *Hydrol Earth Syst Sci Discuss* 4(1):189–210. <https://doi.org/10.5194/hess-12-615-2008>
- Mahmood R, Babel MS (2013) Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theor Appl Climatol* 113(1):27–44. <https://doi.org/10.1007/s00704-012-0765-0>
- Mahmood R, Babel MS (2014) Future changes in extreme temperature events using the Statistical DownScaling Model (SDSM) in the trans-boundary region of the Jhelum river basin. *Weather Clim Extrem* 5–6:56–66. <https://doi.org/10.1016/j.wace.2014.09.001>
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48:3. <https://doi.org/10.1029/2009RG000314>

- Mearns LO, Bogardi I, Giorgi F, Matyasovszky I, Palecki M (1999) Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *J Geophys Res Atmos* 104(D6):6603–6621. <https://doi.org/10.1029/1998JD200042>
- Moseley S (2011) From observations to forecasts—Part 12: Getting the most out of model data. *Weather* 66(10):272–276. <https://doi.org/10.1002/wea.844>
- Navascués B, Calvo J, Morales G, Santos C, Callado A, Cansado A, Cuxart J, Díez M, del Río P, Escribà P, García-Colombo O, García-Moya J, Geijo C, Gutiérrez E, Hortal M, Martínez I, Orfila B, Parodi J, Rodríguez E, Sánchez-Arriola J, Santos-Atienza I, Simarro J (2013) Long-term verification of HIRLAM and ECMWF forecasts over Southern Europe: history and perspectives of Numerical Weather Prediction at AEMET. *Atmos Res* 125(Supplement C):20–33. <https://doi.org/10.1016/j.atmosres.2013.01.010>
- Nesbitt SW, Anders AM (2009) Very high resolution precipitation climatologies from the Tropical Rainfall Measuring Mission precipitation radar. *Geophys Res Lett* 36:15. <https://doi.org/10.1029/2009GL038026>
- Nicholls N, Murray W (1999) Workshop on indices and indicators for climate extremes: Asheville, NC, USA, 3–6 June 1997 breakout group b: precipitation. *Clim Change* 42(1):23–29. <https://doi.org/10.1023/A:1005495627778>
- Önol B (2012) Effects of coastal topography on climate: high-resolution simulation with a regional climate model. *Clim Res* 52:159–174. <https://doi.org/10.3354/cr01077>
- Onyutha C, Tabari H, Rutkowska A, Nyeko-Ogiramo P, Willems P (2016) Comparison of different statistical downscaling methods for climate change rainfall projections over the Lake Victoria basin considering CMIP3 and CMIP5. *J Hydro-environ Res* 12:31–45. <https://doi.org/10.1016/j.jher.2016.03.001>
- Osma VC, Romá JEC, Martín MAP (2015) Modelling regional impacts of climate change on water resources: the Júcar basin, Spain. *Hydrol Sci J* 60(1):30–49. <https://doi.org/10.1080/02626667.2013.866711>
- Parrish DF, Derber JC (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon Weather Rev* 120:1747–1763. [https://doi.org/10.1175/1520-0493\(1992\)120%3c1747:TNMCS%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120%3c1747:TNMCS%3e2.0.CO;2)
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. *Hydrol Earth Syst Sci Discuss* 4(2):439–473. <https://doi.org/10.5194/hess-11-1633-2007>
- Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution-blended analyses for sea surface temperature. *J Clim* 20(22):5473–5496. <https://doi.org/10.1175/2007JCLI1824.1>
- Rockel B, Castro CL, Pielke RA, von Storch H, Leoncini G (2008) Dynamical downscaling: Assessment of model system dependent retained and added variability for two different regional climate models. *J Geophys Res Atmos* 113:D21. <https://doi.org/10.1029/2007JD009461>
- Rodríguez-Puebla C, Encinas AH, Nieto S, Garmendia J (1998) Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula. *Int J Climatol* 18(3):299–316. [https://doi.org/10.1002/\(SICI\)1097-0088\(19980315\)18:3%3c299::AID-JOC247%3e3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(19980315)18:3%3c299::AID-JOC247%3e3.0.CO;2-L)
- Roux RL, Katurji M, Zawar-Reza P, Quénol H, Sturman A (2018) Comparison of statistical and dynamical downscaling results from the WRF model. *Environ Model Softw* 100:67–73. <https://doi.org/10.1016/j.envsoft.2017.11.002>
- Rubel F, Brugger K, Haslinger K, Auer I (2017) The climate of the European Alps: shift of very high resolution Köppen–Geiger climate zones 1800–2100. *Meteorol Z* 26(2):115–125. <https://doi.org/10.1127/metz/2016/0816>
- Rummukainen M (2010) State-of-the-art with regional climate models. *Wiley Interdiscip Rev Clim Change* 1(1):82–96. <https://doi.org/10.1002/wcc.8>
- Sáenz J, Rodríguez-Puebla C, Fernández J, Zubillaga J (2001) Interpretation of interannual winter temperature variations over Southwestern Europe. *J Geophys Res Atmos* 106(D18):20,641–20,651. <https://doi.org/10.1029/2001JD900247>
- Schmidli J, Frei C, Vidale PL (2006) Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *Int J Climatol* 26(5):679–689. <https://doi.org/10.1002/joc.1287>
- Schmidli J, Goodess CM, Frei C, Haylock MR, Hundsdoerfer Y, Ribalaya J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. *J Geophys Res Atmos*. <https://doi.org/10.1029/2005JD007026>
- Schoof J, Pryor S (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *Int J Climatol* 21(7):773–790. <https://doi.org/10.1002/joc.655>
- Serrano A, García JA, Mateos VL, Cancillo ML, Garrido J (1999) Monthly modes of variation of precipitation over the Iberian Peninsula. *J Clim* 12(9):2894–2919. [https://doi.org/10.1175/1520-0442\(1999\)012%3c2894:MMOVO%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012%3c2894:MMOVO%3e2.0.CO;2)
- Serrano-Notivol R, Beguería S, Saz MA, Longares LA, de Luis M (2017) SPREAD: a high-resolution daily gridded precipitation dataset for Spain—an extreme events frequency and intensity overview. *Earth Syst Sci Data* 9(2):721–738. <https://doi.org/10.5194/essd-9-721-2017>
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, Powers JG (2008) A description of the advanced research WRF version 3. NCAR Technical Note NCAR/TN-475+STR. <https://doi.org/10.5065/D68S4MVH>
- von Storch H (2006) Encyclopedia of hydrological sciences, chap models of global and regional climate. Wiley, Oxford. <https://doi.org/10.1002/0470848944.hsa035>
- von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511612336>
- von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *J Clim* 6(6):1161–1171. [https://doi.org/10.1175/1520-0442\(1993\)006%3c1161:DOGCE%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006%3c1161:DOGCE%3e2.0.CO;2)
- Tang J, Niu X, Wang S, Gao H, Wang X, Wu J (2016) Statistical downscaling and dynamical downscaling of regional climate in China: present climate evaluations and future climate projections. *J Geophys Res Atmos* 121(5):2110–2129. <https://doi.org/10.1002/2015JD023977>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106(D7):7183–7192. <https://doi.org/10.1029/2000JD900719>
- Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J Hydrol* 456–457:12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- Timbal B, Jones DA (2008) Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. *Clim Change* 86(1):165–187. <https://doi.org/10.1007/s10584-007-9279-7>
- Turco M, Sanna A, Herrera S, Llasat MC, Gutiérrez JM (2013) Large biases and inconsistent climate change signals in ENSEMBLES

- regional projections. *Clim Change* 120(4):859–869. <https://doi.org/10.1007/s10584-013-0844-y>
- Tustison B, Harris D, Fofoula-Georgiou E (2001) Scale issues in verification of precipitation forecasts. *J Geophys Res Atmos* 106(D11):11,775–11,784. <https://doi.org/10.1029/2001JD900066>
- Ulazia A, Sáenz J, Ibarra-Berastegui G (2016) Sensitivity to the use of 3DVAR data assimilation in a mesoscale model for estimating offshore wind energy potential. A case study of the Iberian northern coastline. *Appl Energy* 180(Supplement C):617–627. <https://doi.org/10.1016/j.apenergy.2016.08.033>
- Ulazia A, Sáenz J, Ibarra-Berastegui G, González-Rojí SJ, Carreno-Madinabeitia S (2017) Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean. *Appl Energy* 208(Supplement C):1232–1245. <https://doi.org/10.1016/j.apenergy.2017.09.030>
- Vaittinada Ayar P, Vrac M, Bastin S, Carreau J, Déqué M, Gallardo C (2016) Intercomparison of statistical and dynamical downscaling models under the EURO and MEDCORDEX initiative framework: present climate evaluations. *Clim Dyn* 46(3):1301–1329. <https://doi.org/10.1007/s00382-015-2647-5>
- Varis O, Kajander T, Lemmelä R (2004) Climate and water: from climate models to water resources management and vice versa. *Clim Change* 66(3):321–344. <https://doi.org/10.1023/B:CLIM.0000044622.42657.d4>
- Vrac M, Vaittinada Ayar P (2017) Influence of bias correcting predictors on statistical downscaling models. *J Appl Meteorol Climatol* 56(1):5–26. <https://doi.org/10.1175/JAMC-D-16-0079.1>
- Wang JJ, Adler RF, Huffman GJ, Bolvin D (2014) An updated TRMM composite climatology of tropical rainfall and its validation. *J Clim* 27(1):273–284. <https://doi.org/10.1175/JCLI-D-13-00331.1>
- Ward MN, Folland CK (1991) Prediction of seasonal rainfall in the north nordeste of Brazil using eigenvectors of sea-surface temperature. *Int J Climatol* 11(7):711–743. <https://doi.org/10.1002/joc.3370110703>
- Wetterhall F, Bárdossy A, Chen D, Halldin S, Xu CY (2006) Daily precipitation-downscaling techniques in three Chinese regions. *Water Resour Res* 42:11. <https://doi.org/10.1029/2005WR004573>
- Wetterhall F, Halldin S, Xu CY (2007) Seasonality properties of four statistical-downscaling methods in central Sweden. *Theor Appl Climatol* 87(1):123–137. <https://doi.org/10.1007/s00704-005-0223-3>
- Widmann M, Bretherton CS, Salathé EP (2003) Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor. *J Clim* 16(5):799–816. [https://doi.org/10.1175/1520-0442\(2003\)016%3c0799:SPDOTN%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016%3c0799:SPDOTN%3e2.0.CO;2)
- Wilby R, Dawson C, Barrow E (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ Model Softw* 17(2):145–157. [https://doi.org/10.1016/S1364-8152\(01\)00060-3](https://doi.org/10.1016/S1364-8152(01)00060-3)
- Wilby RL, Dawson CW (2013) The statistical downscaling model: insights from one decade of application. *Int J Climatol* 33(7):1707–1719. <https://doi.org/10.1002/joc.3544>
- Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. *Int J Climatol* 20:641–661. [https://doi.org/10.1002/\(SICI\)1097-0088\(200005\)20:6<641::AID-JOC501>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0088(200005)20:6<641::AID-JOC501>3.0.CO;2-1)
- Wilby RL, Yu D (2013) Rainfall and temperature estimation for a data sparse region. *Hydrol Earth Syst Sci* 17(10):3937. <https://doi.org/10.5194/hess-17-3937-2013>
- Wilby RL, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. *J Hydrol* 205(1):1–19. [https://doi.org/10.1016/S0022-1694\(97\)00130-3](https://doi.org/10.1016/S0022-1694(97)00130-3)
- Wilby RL, Hay LE, Gutowski WJ, Arritt RW, Takle ES, Pan Z, Leavesley GH, Clark MP (2000) Hydrological responses to dynamically and statistically downscaled climate model output. *Geophys Res Lett* 27(8):1199–1202. <https://doi.org/10.1029/1999GL006078>
- Wilby RL, Tomlinson OJ, Dawson CW (2003) Multi-site simulation of precipitation by conditional resampling. *Clim Res* 23(3):183–194. <https://doi.org/10.3354/cr023183>
- Wilby RL, Dawson CW, Murphy C, O'Connor P, Hawkins E (2014) The statistical downscaling model-decision centric (SDSM-DC): conceptual basis and applications. *Clim Res* 61(3):259–276. <https://doi.org/10.3354/cr01254>
- Wilks DS (1992) Adapting stochastic weather generation algorithms for climate change studies. *Clim Change* 22(1):67–84. <https://doi.org/10.1007/BF00143344>
- Winterfeldt J, Geyer B, Weisse R (2011) Using QuikSCAT in the added value assessment of dynamically downscaled wind speed. *Int J Climatol* 31(7):1028–1039. <https://doi.org/10.1002/joc.2105>
- Yang W, Bárdossy A, Caspary HJ (2010) Downscaling daily precipitation time series using a combined circulation- and regression-based approach. *Theor Appl Climatol* 102(3):439–454. <https://doi.org/10.1007/s00704-010-0272-0>
- Zappa G, Shaffrey LC, Hodges KI (2013) The ability of CMIP5 models to simulate north atlantic extratropical cyclones. *J Clim* 26(15):5379–5396. <https://doi.org/10.1175/JCLI-D-12-00501.1>
- Zorita E, Kharin V, von Storch H (1992) The atmospheric circulation and sea surface temperature in the North Atlantic area in winter: their interaction and relevance for Iberian precipitation. *J Clim* 5(10):1097–1108. [https://doi.org/10.1175/1520-0442\(1992\)005%3c1097:TACASS%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005%3c1097:TACASS%3e2.0.CO;2)
- Zveryaev II, Wibig J, Allan RP (2008) Contrasting interannual variability of atmospheric moisture over Europe during cold and warm seasons. *Tellus A* 60(1):32–41. <https://doi.org/10.1111/j.1600-0870.2007.00283.x>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.