

Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors

Alfonso Hernanz¹  | Juan Andrés García-Valero²  | Marta Domínguez¹  |
Petra Ramos-Calzado³ | María A. Pastor-Saavedra¹ | Ernesto Rodríguez-Camino¹ 

¹Spanish Meteorological Agency (AEMET), Madrid, Spain

²AEMET, Murcia, Spain

³AEMET, Sevilla, Spain

Correspondence

Alfonso Hernanz, Spanish Meteorological Agency (AEMET), Madrid, 28040, Spain.
Email: ahernanzl@aemet.es

Abstract

The Spanish Meteorological Agency (AEMET) is responsible for the elaboration of downscaled climate projections over Spain to feed the Second National Plan of Adaptation to Climate Change (PNACC-2). The main objective of this article is to establish a comparison among five statistical downscaling methods developed at AEMET: (1) Analog, (2) Regression, (3) Artificial Neural Networks, (4) Support Vector Machines and (5) Kernel Ridge Regression. This comparison has been carried out under present conditions and with perfect predictors, based on the framework established by the VALUE network, in particular, on its perfect predictor experiment. In this experiment, we evaluate the marginal aspects of the distributions of daily maximum/minimum temperatures and daily accumulated precipitation analysed by seasons, on a high resolution observational grid (0.05°) over mainland Spain and the Balearic Islands. This is the first of a set of three experiments aimed to allow us to decide which methods, and under what configuration, is more appropriate for the generation of downscaled climate projections over our region. For maximum/minimum temperatures, all methods display a similar behaviour. They capture very satisfactorily the mean values although slight biases are detected on the extremes. In general, results for maximum temperature appear to be more accurate than for minimum temperature, and the nonlinear methods display certain added value. For precipitation, remarkable differences are found among all methods. Most of the methods are capable of reproducing the total precipitation amount quite satisfactorily, whereas other aspects such as intense precipitations and the precipitation occurrence are captured with more accuracy by the Analog method.

KEYWORDS

climate projections, evaluation, Spain, statistical downscaling

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

© 2021 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

1 | INTRODUCTION

Global climate models (GCMs) are the primary tool to simulate future climate projections. Nevertheless, they have known biases and their resolution is not enough to feed the necessities of the impact and adaptation communities (Charles *et al.*, 2004; Wilby *et al.*, 2004; Schoof, 2013). Despite their physical basis and the ability to represent historical climate, GCMs have two key limitations. First, there are substantial uncertainties in and among GCMs (Peel *et al.*, 2015). These uncertainties can result in large differences between simulations, and are usually addressed by analysing simulations from multiple GCMs or different ensemble members (e.g., Chiew *et al.*, 2009; Lauri *et al.*, 2012). Second, GCMs outputs are of too coarse a scale to be directly used in impact studies. Aspects of change not captured by GCMs, due to resolution or regional factors, may be captured by downscaling. If these aspects are of interest, then downscaling may be useful once its added value has been demonstrated (Ekström *et al.*, 2015).

Numerous downscaling techniques have been developed to derive local climate change information from large-scale GCMs outputs (Maraun *et al.*, 2010). The two primary categories of downscaling techniques are (a) dynamic downscaling and (b) empirical/statistical downscaling (ESD), having been both families widely reviewed (Wilby and Wigley, 1997; Charles *et al.*, 2004; Wilby *et al.*, 2004; Rummukainen, 2010; Trzaska and Schnarr, 2014). Most dynamic downscaling methods obtain regional information by nesting a high-resolution regional climate model within a GCM whereas ESD methods are based on the existence of statistical relationships between large-scale variables (predictors) and local variables (predictands). ESD methods have been more widely applied in adaptation and impact studies due to their computational cheapness compared with dynamic downscaling (Trzaska and Schnarr, 2014), which allows to explore uncertainties via multiple realizations, and to their capability of downscaling to single point scale. Predictors/predictand relationships used by ESD methods are first identified during a historical period (calibration) and then applied to predictors from GCMs to get the local climate projections. The main disadvantage of ESD methods is that they rely on the assumption that predictors/predictand relationships remain unaltered under climate change (Charles *et al.*, 2004; Wilby *et al.*, 2004).

Until recently, no comprehensive intercomparison and evaluation of different ESD approaches existed. The EU COST Action VALUE (Maraun *et al.*, 2015) set out to systematically address this gap as far as possible. Three types of experiments were proposed in that framework, as well as by Vrac *et al.* (2007): ‘Experiment 1: Perfect predictor’, ‘Experiment 2: GCM Predictor’ and ‘Experiment 3: Pseudo reality’. In addition, the relevance of credible

GCMs projections has been assessed in a bias correction context (Maraun *et al.*, 2017). Experiment 1 focuses on isolating the skill of a downscaling method assuming perfection on large-scale predictors given by a reanalysis, which represents the most reliable approximation to the reality for large-scale variables that we dispose of. For local variables, there are two possibilities: single stations or observational gridded data. The main advantage of using gridded data is their good spatial coverage, which makes them appropriate for regions of complex and inhomogeneous terrain. On the other hand, the use of observational gridded data introduces a new source of uncertainty depending on the technique adopted for its construction. The Iberian Peninsula was included in the COST VALUE intercomparison work, which had a greater spatial coverage but less sampling points over this particular region. In this study, we have opted for the high-resolution grid described at Section 2, because of the extremely complex topography of the Iberian Peninsula, in order to palliate the lack of spatial uniformity coming from the observational network. The five methods evaluated in this article are applied in Perfect Prognosis (PP) mode (see Gutiérrez *et al.*, 2019; Hertig *et al.*, 2019); when used to generate regional climate scenarios they use predictors from reanalysis for calibration and are applied over GCMs projections. We consider Experiments 2 and 3 extremely important for a complete evaluation, given that these methods are aimed to be applied to GCMs predictors under future (possibly out of range) conditions. Nevertheless, this article is restricted to Experiment 1, whereas Experiments 2 and 3 are planned for future work.

Two out of the five methods presented in this article, Analog (ANA) and Regression (REG) (see Table 2), belong to families that have been evaluated in the COST action VALUE and in Gutiérrez *et al.* (2013) and San-Martín *et al.* (2017), these two last studies over the same area of interest as this article. Artificial Neural Networks (ANN) have also been evaluated over the region following the VALUE framework (Baño-Medina *et al.*, 2020). As explained above, this article corresponds to the first of three experiments, and the complete methodology differs from the approaches adopted in the mentioned studies. In addition, the other two methods, Support Vector Machines (SVM) and Kernel Ridge Regression (KRR) have not been evaluated in the past over our region and need to be tested. There are four aspects that can be evaluated in this context: marginal, temporal, spatial and inter-variable (Maraun *et al.*, 2015; Maraun and Widmann, 2018). The evaluation presented in this article is restricted to marginal aspects, both mean values and extremes, which aligns with the analysis performed by Gutiérrez *et al.* (2019) and Hertig *et al.* (2019), respectively, at VALUE.

Therefore, the main objective of this article, along with Experiments 2 and 3, is to evaluate the new methods and to establish a comparison in order to decide which methods and under what configurations are more appropriate for the generation of downscaled climate projections over Spain to feed the PNACC-2.

The article is organized as follows. First, a description of the datasets is given at Section 2, followed by a brief introduction to the five downscaling methods and their different configurations in Section 3. Results of the evaluation are presented in Section 4, and finally main conclusions are summarized in Section 5.

2 | DATA

The following datasets have been used for downscaling maximum/minimum temperature and precipitation on a daily basis.

Predictors have been taken from the reanalysis ERA-Interim (Dee *et al.*, 2011) of the European Centre for Medium-Range Weather Forecasts (ECMWF) in the area (55.5°N, 30°N, 28.5°W, 15°E) with spatial resolution of $1.5^\circ \times 1.5^\circ$ (see study area at Figure 1) and as daily mean values (from 00, 06, 12 and 18 UTC). Predictors used for each variable are listed in Table 1. The choice of predictors could vary from region to region depending on the characteristics of the large-scale atmospheric circulation and the predictand to be downscaled. Any type of variable or index can be used as a predictor as long as it is reasonable to expect a

certain relationship with the predictand (Wetterhall, 2005). Often, in climate impact studies, such predictors are chosen as variables that are: (a) reliably simulated by GCMs, (b) strongly correlated with the predictand and (c) able to capture climate change signal. In order to scale all predictors and to palliate possible biases on GCMs when applied over them, all predictors are standardized using their own mean and standard deviation over the period 1980–2005. In addition, they are interpolated to each target point as a weighted average of the four nearest neighbours, being their weights the inverse of the distances.

Predictands (daily maximum/minimum temperature and 24 hr. accumulated precipitation) come from a high-resolution grid (0.05°) consisting of 16,156 points over Spain (mainland and Balearic Islands) developed by AEMET (Peral *et al.*, 2017). This grid has been generated using an adaptation of the HIRLAM Surface Analysis code (Navascués *et al.*, 2003; Rodríguez *et al.*, 2003), based on an Optimum Interpolation algorithm (Daley, 1991), applied over the very dense AEMET observational network (1,800 selected stations for temperature and 3,236 for precipitation).

The period used both for calibration and evaluation is 1980–2005, conditioned by the availability of data for Experiments 2 and 3, in order to make them inter-comparable. Independence between calibration and evaluation data is achieved by dividing the whole period in fivefolds (1980–1984, 1985–1989, 1990–1994, 1995–1999 and 2000–2005), so each of them is downscaled using the others for calibration and finally the fivefolds are joint to reconstruct the whole evaluation period.

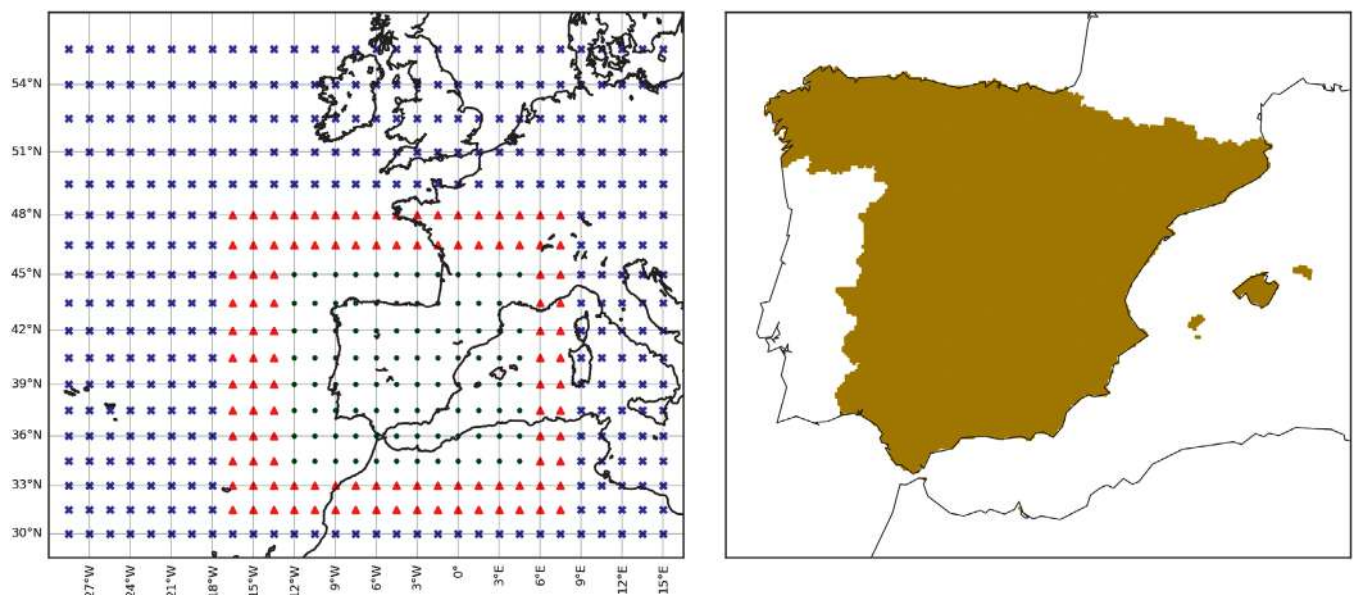


FIGURE 1 Area of study. On the left, grid points of reanalysis (predictors), in different colours to delimitate different weightings (larger weights in green and lower weights in blue) for the synoptic analogy (Analog method). On the right, the target region (predictands)

TABLE 1 Predictor variables for the different methods: Common variables for maximum/minimum temperature and precipitation (upper rows), specific variables for both temperatures (middle rows) and variables just for precipitation (lower rows)

	Variable	Name	Levels	Units	Type
Common	T	Temperature	500	K	Direct
	U	U wind component	500	m·s ⁻¹	Direct
	V	V wind component	500	m·s ⁻¹	Direct
	Ug	U geostrophic wind component	Mean sea level	m·s ⁻¹	Derived
	Vg	V geostrophic wind component	Mean sea level	m·s ⁻¹	Derived
	SLP	Mean sea level pressure	Mean sea level	Pa	Direct
Maximum/minimum temperature	T	Temperature	700, 850	K	Direct
	Ins	Theoretical insolation	–	hr	
Precipitation	Vtg	Vertical thermal gradient	Between 500 and 850 hPa	K	Derived
	Vog	Geostrophic vorticity	Mean sea level	s ⁻¹	Derived
	Vo	Vorticity	500	s ⁻¹	Derived
	Dg	Geostrophic divergence	Mean sea level	s ⁻¹	Derived
	D	Divergence	500	s ⁻¹	Derived
	SLP	Mean sea level pressure variation from the trend	previous day	Mean sea level	Pa

Note: Attending their availability, predictors are categorized as direct/derived from reanalysis outputs. Theoretical insolation is the theoretical number of hours of insolation as function exclusively of the latitude and the day of the year.

3 | DOWNSCALING METHODS AND DIAGNOSTICS

In this section, we provide a description of the downscaling methods and the methodology adopted for their evaluation.

3.1 | Downscaling methods

Five methods have been applied to both temperature and precipitation; some of them under different configurations (see Table 2):

1. Analog methods (Lorenz, 1969; Zorita and von Storch, 1999) are based on the assumption of similar local conditions under similar synoptic situations. They search synoptic analog situations in the past, for which Analog methods rely on the availability of long historical series. In general, Analog methods can maintain spatial coherence but on the other hand, they cannot predict values out of the observed range. Their simplicity and ability to capture many aspects of the local climate have made Analog methods widely applied in this field (Cubasch *et al.*, 1996; Wetterhall, 2005; Gutiérrez *et al.*, 2013; San-Martín *et al.*, 2017; Gutiérrez *et al.*, 2019; Hertig *et al.*, 2019). The Analog method presented in this article is an

TABLE 2 Downscaling methods grouped by three families: Analog, regression and machine Learning methods

Family	Temperature	Precipitation
Raw	RAW (ERA-Interim 1.5° × 1.5°)	RAW (ERA-Interim 1.5° × 1.5°)
Analog	1. ANA	Synoptic analogy: 1. ANA-SYN-1 (best analog) 2. ANA-SYN-N (average) 3. ANA-SYN-PDF (probability density function) Same but combined with local analogy: 4 ANA-LOC-1 5 ANA-LOC-N 6 ANA-LOC-PDF
Regression	2. REG	7 REG-LIN (linear) 8 REG-EXP (exponential) 9 REG-CUB (cubic)
Machine learning	3. ANN (Artificial Neural Network) 4. SVM (Support Vector Machines) 5. KRR (Kernel Ridge Regression)	10 ANN 11 SVM 12 KRR

Note: Additionally, ERA-Interim 1.5° × 1.5° has been included as 'RAW'.

adaptation of a method previously developed and applied at AEMET (Petisco de Lara, 2008a, 2008b; Amblar-Francés *et al.*, 2017). In this method, synoptic analogy is defined by the Euclidean distance of the following four variables: Ug, Vg, U500 and V500 (see Table 1), to which we refer as *synoptic analogy fields*, over the whole region. These variables are taken into account with different weights depending both on their level (surface/500 hPa) and their distance to the target area. The larger/lower weights are given to the nearer/farther points. Three different regions have been delimited (see Figure 1), being their relative weights 0/1/2 for surface and 1/4/8 for 500 hPa (Petisco de Lara, 2008a, 2008b; Ribalaygua *et al.*, 2013). For temperature, the first step consists of the search of analog days to the target day. Then, for each target point, a multiple linear regression with L2 regularization, *Ridge Regression* (Hoerl and Kennard, 1970; Tikhonov and Arsenin, 1977) is established with all predictors from the analog days. This particular Analog method has no problem in predicting values out of the observed range but does not yield spatially coherent fields. For precipitation there are six different configurations: (a) ANA-SYN-1: precipitation is given by the best analog to the target day, (b) ANA-SYN-N: precipitation is calculated as a weighted average of a selection of analog days, with weights depending on their analogy, (c) ANA-SYN-PDF: a probability density function is built based on a selection of analog days, with probabilities given by their analogy to the target day, and precipitation is given by a random analog day, (d) ANA-LOC-1, (e) ANA-LOC-N and (f) ANA-LOC-PDF: similar to the previous three options but using a combination of synoptic and local analogies instead of exclusively synoptic. Local analogy is given, for each target point, by the Euclidean distance of significant predictors along a selection of analog days. Significant predictors are previously determined, for each grid point and weather type, with a *k*-means clustering algorithm (Lloyd, 1957; MacQueen, 1967) and the Spearman Rank correlation coefficient.

2. Regression methods use linear and nonlinear relationships between predictors and predictands (Sailor and Li, 1999; Wilby *et al.*, 2002). The Regression method presented in this article is an adaptation of a method previously developed and applied at AEMET (Amblar-Francés *et al.*, 2017), which consists of a Multiple Linear Regression (MLR) for temperature and a Generalized Linear Model (GLM) for precipitation; based on the Statistical Downscaling Model-SDSM (Wilby *et al.*, 2002). MLRs and GLMs have been widely used to downscale temperature and precipitation because of their simplicity (Benestad, 2002; Huth, 2002; Huth, 2004; Benestad, 2005; Gutiérrez *et al.*, 2019; Hertig *et al.*, 2019). Nonetheless, some major drawbacks are their poor representation of observed variance and extreme events, or the assumption of normality of data (Wilby *et al.*, 2004), which makes them unsuitable for daily precipitation (Gutiérrez *et al.*, 2019). In the Regression method applied here, temperature is estimated by a *Ridge Regression*. For precipitation, a logistic regression with L2 regularization is used for the wet/dry classification problem under a probabilistic approach; the probability of precipitation given by the logistic regression is compared with a random number taken from the uniform distribution between 0 and 1 in order to classify as wet or dry. Then, for wet data, the amount of precipitation can be calculated using three different configurations: linear regression (REG-LIN), exponential regression (REG-EXP) and cubic regression (REG-CUB), all of them with L2 regularization. The terms exponential and cubic regressions refer to linear regressions applied over transformed predictands (their natural logarithm and cubic root, respectively).
3. ANN (McCulloch and Pitts, 1943; Rosenblatt, 1958) have gained wide recognition due to their capability to simulate complex nonlinear predictor/predictand relationships. ANN are supervised learning algorithms based on the human brain functioning and are composed of several nodes which try to imitate the behaviour of its neurons. Each of these nodes work as a perceptron (Rosenblatt, 1958), receiving an input and deciding whether to pass a binary signal or not to the connected nodes by an activation function. In a multi-layer perceptron (MLP), these nodes are organized in several layers, the input layer, the output layer and a set of hidden layers. Each node communicates with all the nodes on the neighbouring layer with different weights that are established during the calibration process with a backpropagation algorithm. ANN have proven to be a very powerful tool to simulate nonlinear classifications and regressions, and its major drawbacks are the fact that during their training they can get trapped in local minima, their high subjectivity and dependency on model architecture (Suykens, 2001) and the high computational cost of their training. ANN have been extensively applied to this field (Trigo and Palutikof, 1999; Sailor *et al.*, 2000; Snell *et al.*, 2000; Coulibaly *et al.*, 2005; Dibike and Coulibaly, 2006; Chadwick *et al.*, 2011; Mendes and Marengo, 2013). The ANN method presented in this article is an adaptation of a method previously developed and applied at AEMET (García-Valero, 2021). For temperature, it uses multilayer

perceptron regression (MLPR) with a rectified linear unit activation function and L2 regularization. For precipitation, the wet/dry classification is solved by a multilayer perceptron classifier (MLPC) with the same structure as the mentioned MLPR. The output of the MLPC is transformed to well calibrate probabilities using the *Platt scaling* method (Platt, 1999) and it is applied under the probabilistic approach explained at Regression method. Then, for wet data, a MLPR with the same structure is used to estimate the precipitation amount.

4. Another popular Machine Learning technique for both nonlinear classification and regression is SVMs (Boser *et al.*, 1992; Cortes and Vapnik, 1995; Vapnik, 1995). SVM are supervised learning algorithms originally designed for binary classification based on a combination of minimal errors and maximal margins around the boundary decision. The nonlinear problem can be addressed by the use of kernels, mapping original data from the input space to a higher dimensional space (*feature space*), where the problem becomes linear. Expensive high-dimensional computations, usually referred to as the curse of dimensionality, are avoided thanks to what is commonly known as the *kernel trick* (see Shawe-Taylor and Cristianini, 2004). A slightly different version of SVM can be applied for regressions. Both approaches need to solve the optimization of a convex quadratic programming problem, which lacks the inconvenience of local minimums. Their tuning is simpler than for ANN but the process is also computationally expensive. SVM have also been applied to downscaling climate projections in the past (Tripathi *et al.*, 2006; Yu and Liong, 2007; Anandhi *et al.*, 2008; Ghosh and Mujumdar, 2008; Chen *et al.*, 2010; Sachindra *et al.*, 2013; Hou *et al.*, 2014). The method presented in this article uses a Support Vector Regression (SVR) for temperature, with a *Gaussian* (or *radial basis function, RBF*) kernel and L2 regularization. For precipitation, a Support Vector Classifier (SVC) with *Gaussian kernel* and L2 regularization is used to solve the wet/dry problem, under the already explained probabilistic approach and after obtaining well-calibrated probabilities as explained for ANN. Then, for wet data, a SVR with *Gaussian kernel* and L2 regularization is used to estimate the precipitation amount.
5. The last method uses two specific forms of SVM: KRR (Vovk, 2013) and *Least-Square Support Vector Machine* (LS-SVM) (Suykens and Vandewalle, 1999). KRR is the combination of a Ridge Regression with a kernel to address nonlinear problems. LS-SVM can be used for classification and regression, and the main difference with classical SVM is that they find the solution by solving a set of linear equations instead of the convex

quadratic programming problem. A KRR with a Gaussian kernel and L2 regularization has been used for temperature. For precipitation, a LS-SVM with linear kernel and L2 regularization is used to solve the wet/dry problem, under the already explained probabilistic approach and after obtaining well-calibrated probabilities as explained for ANN. Then, for wet data, a KRR with *Gaussian kernel* and L2 regularization is used to estimate the precipitation amount.

Parameters such as L2 regularization term weight, number of hidden neurons, Gaussian kernel width, and so forth, have been tuned by a systematic cross validation process over a wide enough range of possible values.

Additionally, original ERA-Interim $1.5^\circ \times 1.5^\circ$ ('RAW') has been included in many diagnostics and figures in order to analyse the benefits of applying any type of downscaling.

3.2 | Diagnostics

The evaluation of the five downscaling methods is focussed on marginal aspects, through an analysis of the distributions of each variable and season. In order to enable a manageable comparison among all methods and to perform an analysis replicable at Experiments 2 and 3, only three simple indexes have been used to synthesize some basic features of both observed and simulated distributions. For maximum/minimum temperature, these three indexes are the mean value and the 10th and 90th percentiles. For precipitation, three different indexes have been chosen: total precipitation (PRCPTOT), total precipitation on very wet days (R95p) and number of wet days (R01). For each of these indexes, we have evaluated their interannual mean values. R01 is defined as the number of wet days (precipitation ≥ 1 mm), and R95p corresponds to the total precipitation on very wet days, being a very wet day defined by the 95th percentile of the wet days from the observed climatology on a reference period (1961–1990). For each of these indexes, bias and root mean square error (RMSE) have been computed and analysed. For precipitation, biases and RMSEs have been calculated over relative errors (%). Biases have been presented in the form of boxplots, where each box contains the quartiles of all grid points, the whiskers extend to a maximum of 1.5 times the interquartile range and outliers beyond this range are plotted individually. Bias maps have been included in Supporting Information (Figure S2–S10) as useful information for users though are not commented in the text. In addition to these marginal aspects, RMSEs for daily maximum/minimum temperatures and correlation for monthly accumulated precipitation have been calculated.

4 | RESULTS

The main results of each of the variables considered are presented in the following subsections.

4.1 | Maximum temperature

All methods clearly improve results of RAW (Figures 2 and 3). ANN, SVM and KRR achieve the best daily accuracies closely followed by ANA, all of them with RMSEs mostly between 1 and 1.5°C. Finally, REG gets the higher RMSEs, generally between 1.5 and 2°C (Figure S1).

Focussing on the marginal aspects through the mean values and the 10th and 90th percentiles, the most relevant results to highlight are:

1. 90th percentile. REG's biases in winter and autumn are much bigger than for the other methods, being ANA's biases for autumn also important (Figure 2). This is reflected in their RMSEs: 1.1°C for REG, 0.37°C for ANA and between 0.2 and 0.25°C for the others in winter, and 1.4°C for REG, 0.66°C for ANA and between 0.14 and 0.17°C for the others in autumn (Figure 3).

2. Mean value. The five methods capture satisfactorily the mean value, with biases generally smaller than 0.5°C and all of them present a slight positive bias in winter and negative in autumn, more marked for REG (Figure 2). RMSE for REG in winter (0.47°C) and in autumn (0.67°C) are also notably higher than for the other methods, with RMSEs between 0.22 and 0.32°C in winter and 0.24 and 0.32°C in autumn (Figure 3).
3. 10th percentile. Biases for the lower tail of the distributions are generally bigger than for their mean values in all methods (Figure 2). In summer, REG's RMSE (0.58°C) is very high compared to the other methods, which are between 0.18 and 0.23°C. On the other hand, its RMSE in winter is only 0.41°C whereas the rest are between 0.48 and 0.62°C, and in autumn, REG's RMSE is 0.26°C whereas ANN, SVM and KRR are between 0.36 and 0.44°C (Figure 3).

In summary, the improvements of performing any type of downscaling are clear, and ANN, SVM and KRR usually achieve smaller biases and RMSEs than the other methods. On the other hand, REG gets much higher RMSE than the other methods in some indexes and seasons.

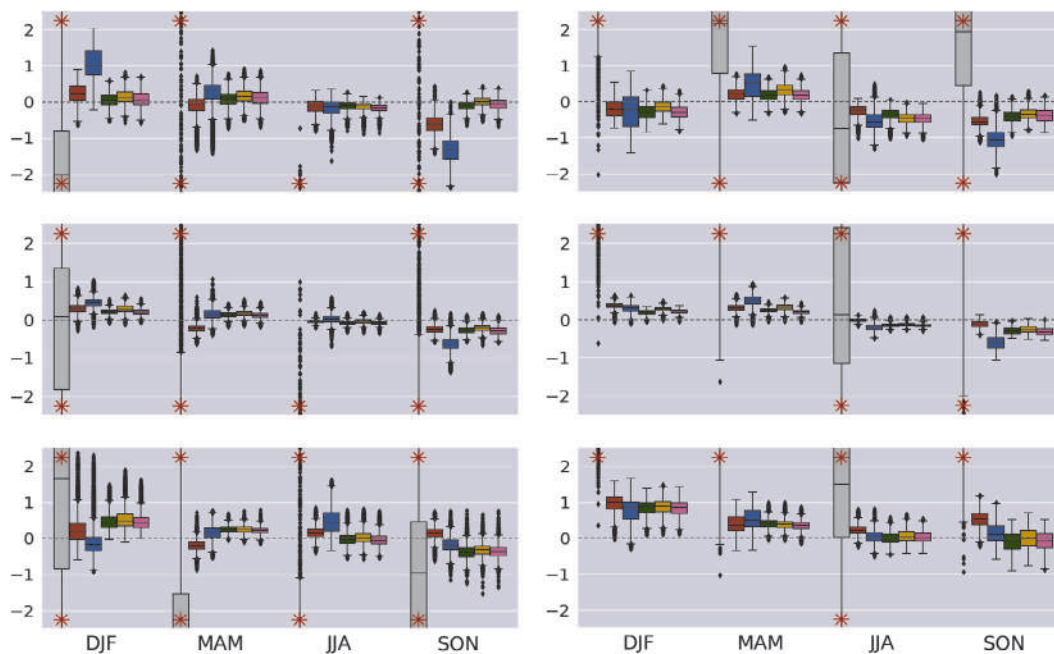


FIGURE 2 Bias (°C) for 90th percentile (upper row), mean value (middle row) and 10th percentile (lower row) corresponding to the daily maximum (left column) and minimum (right column) temperatures by season. Methods: RAW (grey), ANA (red), REG (blue), ANN (green), SMV (orange) and KRR (violet). Each box contains the quartiles of all grid points and the whiskers extend to a maximum of 1.5 times the interquartile range. Outliers beyond this range are plotted individually. Vertical scales have been limited for a good visualization of the ESD methods. A red asterisk indicates that values outside the plotted range. Same figure but for the whole range of values is included as Supporting Information (Figure S11)

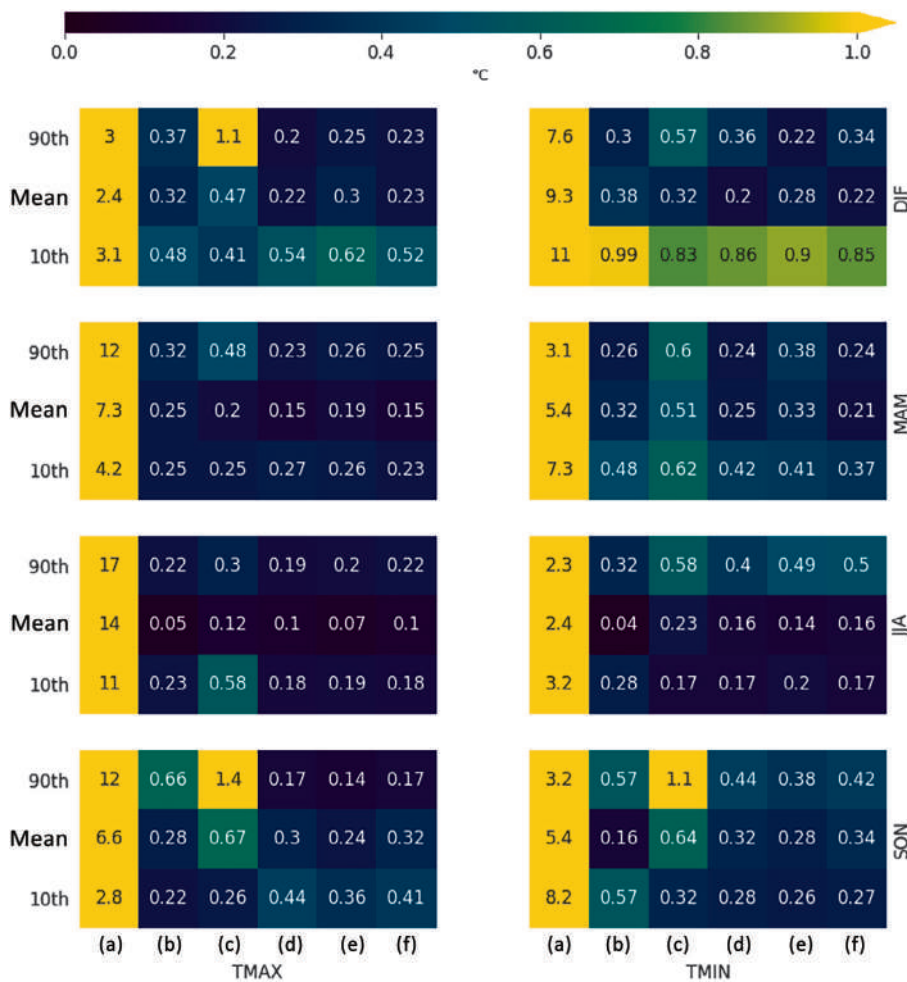


FIGURE 3 RMSE (°C) for 90th percentile, mean value and 10th percentile corresponding to daily maximum (left column) and minimum (right column) temperatures by season (each block corresponds to a different season). Methods: RAW (a), ANA (b), REG (c), ANN (d), SVM (e) and KRR (f)

4.2 | Minimum temperature

For minimum temperature, the benefits of downscaling are also out of question, achieving all methods clearly better results than RAW (Figures 2 and 3). Daily RMSEs for minimum temperature are higher than for maximum temperature for all methods (Figure S1). This is probably related to a known difficulty to downscale surface minimum temperatures in the region on episodes of lower troposphere thermal inversion when free-tropospheric temperatures are used as predictors (Gutiérrez *et al.*, 2013). Again, REG achieves the worst daily accuracies, with RMSEs mostly between 1.5 and 2.5°C, meanwhile the other methods stay between 1 and 2°C. Unlike for maximum temperature, this time ANA gets the best accuracies, closely followed by ANN, SVM and KRR.

Regarding minimum temperature, the same marginal aspects as for maximum temperature have been analysed:

1. 90th percentile. REG displays higher biases than the other methods in spring and autumn, and its RMSEs are systematically bigger than for the other methods

(Figure 3). All methods show a negative bias in summer, autumn and less marked in winter, whereas in spring, they display a slight positive bias (Figure 2).

2. Mean value. As it happened for maximum temperature, the mean value for minimum temperature is also well captured by all methods, with biases concentrated mostly between -0.5°C and $+0.5^{\circ}\text{C}$ (Figure 2). There is a generalized bias, positive in winter and spring and negative in autumn and, more slightly, in summer (Figure 2). REG bias in spring and autumn is notably bigger than for the other methods, and it reflects on its RMSE: 0.51°C in spring and 0.64°C in autumn, meanwhile the other methods reach RMSEs between 0.21 and 0.33°C and 0.16 and 0.34°C in the respective seasons. ANA is the one achieving the best results in summer and autumn, with RMSEs of 0.04°C and 0.16°C respectively versus the 0.14 – 0.23°C and 0.28 – 0.64°C accomplished by the other methods (Figure 3).
3. 10th percentile. In general, errors are bigger both than for the mean value and for the 10th percentile of maximum temperature. There is strong overestimation in winter, where all methods show biases concentrated

around almost $+1^{\circ}\text{C}$ (Figure 2). This is probably related to the above mentioned thermal inversions issue. In spring, all methods display an important positive bias as well. In autumn, ANA presents a marked positive bias (Figure 2) and much higher RMSE (0.57°C) than the other methods, which are between 0.26 and 0.32°C (Figure 3).

In summary, all methods achieve better results than RAW, and errors for minimum temperature are generally higher than for maximum temperature, especially the lower values (10th percentile in winter). Biases both for maximum and minimum temperatures are, in general, lower and with less spread for Machine Learning methods than for REG and ANA (Figure 2). This indicates they capture with more accuracy spatial variability for the three indexes analysed.

Results for maximum/minimum temperatures align with the ones found in the literature. Gutiérrez *et al.* (2013) reached better results for maximum temperature than for minimum temperature, as it is our case. At VALUE, the two-step Analog method also achieved slightly better results than MLRs (Gutiérrez *et al.*, 2019; Hertig *et al.*, 2019). Finally, Baño-Medina *et al.* (2020) proved how Neural Networks can improve results from MLRs.

4.3 | Precipitation

Precipitation is a much more complex variable and important differences are found among all methods:

1. PRCPTOT. All methods clearly improve results of not applying any downscaling (RAW), which leads to a generalized overestimation (Figure 4). All of them present quite similar results with small bias (Figure 4) and RMSE (Figure 5) except in autumn, where RMSEs are higher and there is a marked negative bias. None of the six options performs clearly better than the others. On Regression methods, REG-LIN clearly overcomes the other two options. REG-EXP and REG-CUB display important negative biases in all seasons (Figure 4) and their RMSEs are systematically higher than for REG-LIN (Figure 5). On Machine Learning methods, ANN and KRR clearly overcome SVM, which also presents important negative biases (Figure 4) and high RMSEs (Figure 5). When comparing the best methods of each of the three families, none of the six Analog methods, REG-LIN, ANN or KRR performs clearly better than the others. Nevertheless, Machine Learning methods, ANA-SYN-N and ANA-LOC-N present the higher correlations of monthly accumulated precipitation (Figure 6) both in

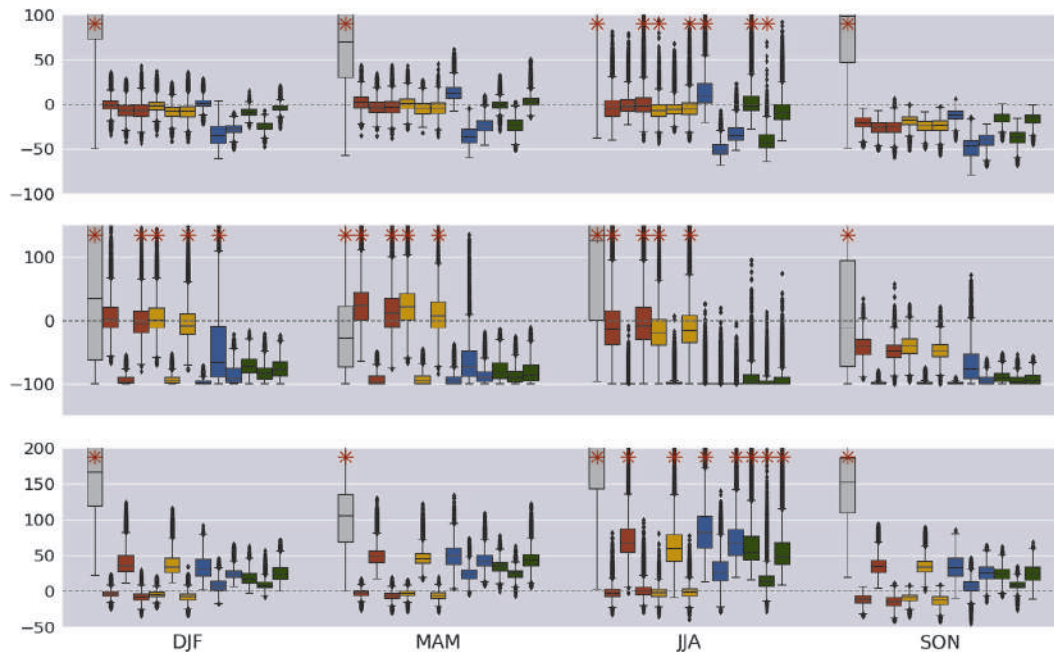


FIGURE 4 Relative bias (%) for PRCPTOT (upper row), R95p (middle row) and R01 (lower row) of precipitation by season. Methods coloured by families: RAW (grey), ANA-SYN-1, ANA-SYN-N and ANA-SYN-PDF (red), ANA-LOC-1, ANA-LOC-N and ANA-LOC-PDF (orange), REG-LIN, REG-EXP and REG-CUB (blue), ANN, SVM and KRR (green). Vertical scales have been limited for a good visualization of the ESD methods. A red asterisk indicates that values lie outside the plotted range. Same figure but for the whole range of values is included as Supporting Information (Figure S12)

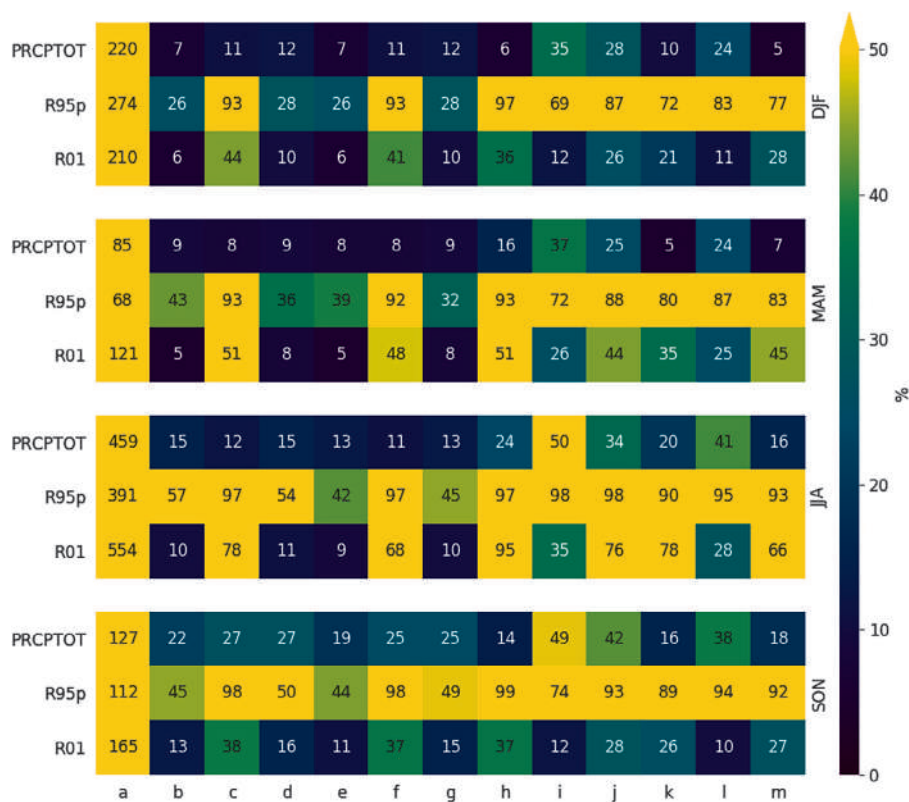


FIGURE 5 RMSE over relative errors (%) for PRCPTOT, R95p and R01 by seasons (each block corresponds to a different season) of precipitation. Methods: RAW (a), ANA-SYN-1 (b), ANA-SYN-N (c) ANA-SYN-PDF (d), ANA-LOC-1 (e) ANA-LOC-N (f), ANA-LOC-PDF (g), REG-LIN (h), REG-EXP (i), REG-CUB (j), ANN (k), SVM (l) and KRR (m)

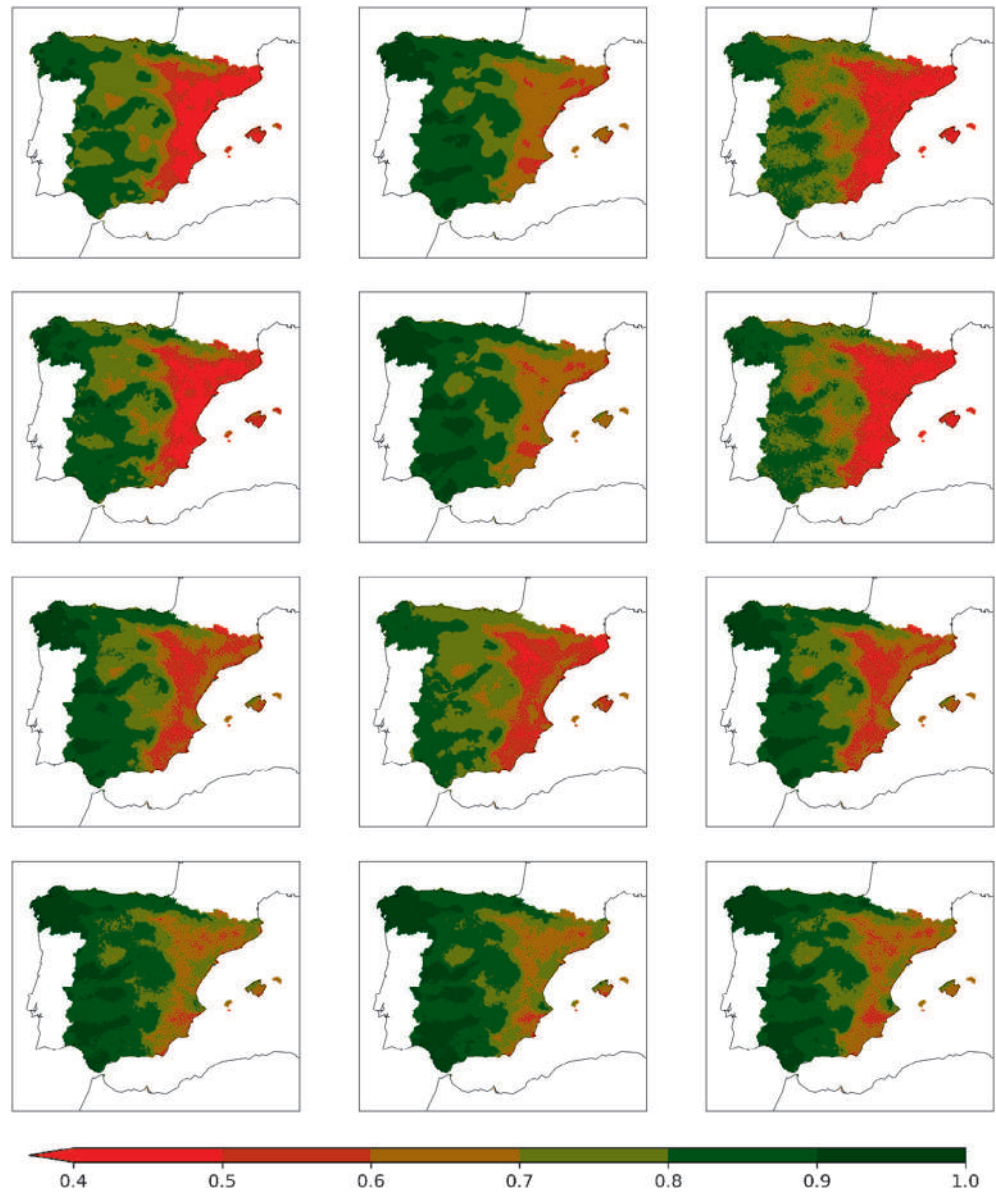
the Atlantic region (where all methods achieve good correlations) and in the Mediterranean region (where most methods get poorer correlations). In summary, considering only the mean total precipitation, Analog methods, REG-LIN, ANN and KRR get the best accuracies, but if we also take into account monthly accumulated precipitation, ANA-SYN-N, ANA-LOC-N, ANN and KRR overcome the others.

2. R95p. Noticeable differences are observed between ANA-SYN-N, ANA-LOC-N and the other four Analog methods, presenting the first two an important underestimation (Figure 4), what is expected due to the averaging they perform. Among Regression methods, the three of them present a systematic negative bias (Figure 4) and high RMSE (Figure 5) slightly less marked for REG-EXP. And among Machine Learning methods, again the three of them display a systematic underestimation, slightly more marked for SVM (Figure 4), which also exhibits the highest RMSE (Figure 5). When comparing the best representatives of each of the three families (ANA-SYN-1, ANA-SYN-PDF, ANA-LOC-1 and ANA-LOC-PDF for Analog methods, REG-EXP for Regression methods and ANN and KRR for Machine Learning methods), it can be seen that Analog methods overcome Regression and Machine Learning methods, with smaller RMSEs in all seasons (Figure 5). And as it happened with the mean total

precipitation, all 12 methods present certain underestimation in autumn (Figure 4). In summary, Analog methods (except for ANA-SYN-N and ANA-LOC-N) are the ones which best capture the intense precipitations described by the R95p index, followed by REG-EXP.

3. R01. All methods clearly improve results of not applying any downscaling (RAW), which leads to a generalized overestimation of the frequency of rainy days (Figure 4). Among Analog methods, there is a clear distinction between ANA-SYN-N and ANA-LOC-N with the rest. These two options present a clear overestimation (Figure 4), again expected by the averaging of analog days. The other four options present similar results, being ANA-SYN-1 and ANA-LOC-1 the ones with lower RMSEs (Figure 5). Among the Regression methods, all of them present a certain degree of overestimation (Figure 4), slighter for REG-EXP, the one with smaller RMSE (Figure 5). Machine Learning methods also display some overestimation, more marked for ANN and KRR (Figure 4). When comparing the best representatives of each family (ANA-SYN-1 and ANA-LOC-1 for Analog methods, REG-EXP for Regression methods and SVM for Machine Learning methods), Analog methods clearly overcome the other families, especially in spring and winter, with smaller biases (Figure 4) and lower RMSEs (Figure 5).

FIGURE 6 Correlation for monthly accumulated precipitation. First row: ANA-SYN-1, ANA-SYN-N and ANA-SYN-PDF. Second row: ANA-LOC-1, ANA-LOC-N and ANA-LOC-PDF. Third row: REG-LIN, REG-EXP and REG-CUB. Fourth row: ANN, SVM and KRR



To sum up, REG-EXP, REG-CUB and SVM get higher errors than the other methods. As for intense precipitations, ANA-SYN-1, ANA-SYN-PDF, ANA-LOC-1 and ANA-LOC-PDF achieve lower errors than the other methods. Finally, as regards precipitation occurrence ANA-SYN-1, ANA-SYN-PDF, ANA-LOC-1 and ANA-LOC-PDF present lower errors than the other methods, followed by REG-EXP and SVM.

Results for precipitation align with the ones found at VALUE: Gutiérrez *et al.* (2019) and Hertig *et al.* (2019) showed how, in general, Analog methods captured fairly well the mean precipitation, wet-day frequency and heavy precipitations, whereas GLMs showed a tendency to overestimate the wet-day frequency and to underestimate heavy precipitations, which agrees with our findings. Baño-Medina *et al.* (2020) proved how Neural Networks can improve results from GLMs, which aligns

with our results as well. On the other hand, our results appear to fall in certain contradictions with those found at San-Martín *et al.* (2017). The most remarkable differences are the generalized underestimation of the mean precipitation by their Analog method and a certain tendency to overestimate extreme events by their GLM. These conflicts might be partially explained by differences in ESD methods (and their particular implementations), predictors, periods and the observational grid.

5 | CONCLUDING REMARKS

The application of downscaling methods is particularly challenging in mainland Spain and the Balearic Islands due to its large regional spatio-temporal variabilities. In

this article, we have evaluated and intercompared five statistical downscaling techniques under present conditions and with perfect predictors mainly focusing on marginal aspects of the distributions of maximum/minimum temperatures and daily precipitation by season. This analysis has been performed through a set of three indexes for each variable, aiming to summarize some relevant features useful for impact studies, and the following conclusions have been reached:

Both for maximum/minimum temperatures and for precipitation, the benefits of applying any type of downscaling are clear. All methods greatly improve results from raw reanalysis, which introduces great biases in the local scale due to its averaging of big areas.

Maximum/minimum temperatures are well captured by the five downscaling methods, being results for their mean values better than for the tails of the distributions. Minimum temperature's results are generally worse than for maximum temperature, especially concerning the lower tail of the distributions. ANN, SVM and KRR usually overcome REG, emphasizing that predictors/predictand relationships have a non-linear component. Additionally, REG is also overcome by ANA, which suggests that those relationships are somehow dependent on the synoptic situation described by Ug, Vg, U500 and V500.

Precipitation is a much more complex variable, and the use of different approaches is very helpful and convenient. Among Analog methods, ANA-SYN-N and ANA-LOC-N present a very different behaviour compared to the other options. These two options achieve higher correlations for monthly accumulated precipitation but on the other hand, they fail in capturing heavy precipitation and number of wet/dry days. The addition of local analogy does not clearly improve the use of synoptic analogy exclusively, and none of the six options has proven to be clearly better than the simplest form, ANA-SYN-1. Among Regression methods, REG-LIN captures total precipitation more accurately than the other two options, while for intense precipitations and for the wet/dry classification it is overcome by REG-EXP. Machine Learning methods achieve higher correlations for monthly accumulated precipitation than the other families and, among them, ANN and KRR have displayed very similar results, generally more accurate than SVM, except for the wet/dry classification. When inter comparing all methods, there are options able to capture total precipitation satisfactorily inside each of the three families. On the other hand, for intense precipitations and the wet/dry classification there are four Analog options (ANA-SYN-1, ANA-SYN-PDF, ANA-LOC-1 and ANA-LOC-PDF) which achieve more accurate results than the other options and families.

Nevertheless, these conclusions might vary on Experiments 2 and 3. Analog methods require that synoptic

analog situations can be found in the calibration period, which cannot be assured for future projections. And the other methods might perform differently under extrapolation, especially those highly nonlinear (REG-EXP, ANN, SVM and KRR). Thus, it is extremely important to evaluate the behaviour of all of them out of the calibration range and when applied to imperfect predictors. For these reasons, we consider it important to wait for Experiments 2 and 3, in order to select and to have a consensus on the methods to feed the Spanish PNACC-2. Finally, we strongly recommend users to check conclusions from those future works in case significantly different results are reached.

ACKNOWLEDGEMENTS

We would like to acknowledge all the teams implicated on the ERA-Interim reanalysis and on the Python machine-learning library Scikit-learn (Pedregosa *et al.*, 2011). A special gratitude to María Candelas Peral García and Beatriz Navascués for their development of the high-resolution observational grid over Spain here used and to the numerous altruistic collaborators of the AEMET observational network. Finally, we would like to thank Eduardo Petisco de Lara and Pilar Amblar Francés for their early developments in analog and regression downscaling methods and two anonymous reviewers for their interest, good disposition and wise suggestions. One author (M.D.) has received funding from the MEDSCOPE project co-funded by the European Commission as part of ERA4CS, an ERA-NET initiated by JPI Climate, grant agreement 690462.

AUTHOR CONTRIBUTIONS

Alfonso Hernanz: Software; visualization; writing-original draft. **Juan García-Valero:** Writing-review & editing. **Marta Domínguez:** Writing-review & editing. **Petra Ramos:** Writing-review & editing. **María Pastor:** Writing-review & editing. **Ernesto Rodríguez-Camino:** Supervision; writing-review & editing.

ORCID

Alfonso Hernanz  <https://orcid.org/0000-0003-1091-0422>

Juan Andrés García-Valero  <https://orcid.org/0000-0002-3914-6328>

Marta Domínguez  <https://orcid.org/0000-0001-7840-5516>

Ernesto Rodríguez-Camino  <https://orcid.org/0000-0002-1565-2373>

REFERENCES

- Amblar-Francés, P., Casado-Calle, M.J., Pastor-Saavedra, M.A., Ramos-Calzado, P., and Rodríguez-Camino, E. (2017) Guía de escenarios regionalizados de cambio climático sobre España a partir de los resultados del IPCC-AR5. Available at: <https://>

- www.aemet.es/documentos/es/conocerlas/recursos_en_linea/publicaciones_y_estudios/publicaciones/Guia_escenarios_AR5/Guia_escenarios_AR5.pdf.
- Anandhi, A., Srinivas, V.V., Nanjundiah, R.S. and Nagesh Kumar, D. (2008) Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology*, 28, 401–420. <https://doi.org/10.1002/joc.1529>.
- Baño-Medina, J., Manzanar, R. and Gutiérrez, J.M. (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13, 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>.
- Benestad, R.E. (2002) Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble. *Climate Research*, 21, 105–125.
- Benestad, R.E. (2005) Climate change scenarios for northern Europe from multi-model IPCC AR4 climate simulations. *Geophysical Research Letters*, 32, L17704. <https://doi.org/10.1029/2005GL023401>.
- Boser, B., Guyon, I. and Vapnik, V. (1992). *A Training Algorithm for Optimal Margin Classifier. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory* 5. <https://doi.org/10.1145/130385.130401>
- Chadwick, R., Coppola, E. and Giorgi, F. (2011) An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlinear Processes in Geophysics*, 18, 1013–1028. <https://doi.org/10.5194/npg-18-1013-2011>.
- Charles, S.P., Bates, B.C., Smith, I.N. and Hughes, J.P. (2004) Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrological Processes*, Special Issue: Scale and Scaling in Hydrology, 18, 1373–1394. <https://doi.org/10.1002/hyp.1418>.
- Chen, H., Guo, J., Xiong, W., Guo, S. and Xu, C.-Y. (2010) Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin. *Advances in Atmospheric Sciences*, 27, 274–284. <https://doi.org/10.1007/s00376-009-8071-1>.
- Chiew, F.H.S., Teng, J., Vaze, J., Post, D.A., Perraud, J.M., Kirono, D.G.C. and Viney, N.R. (2009) Estimating climate change impact on runoff across Southeast Australia: method, results, and implications of the modeling method. *Water Resources Research*, 45(10).
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>.
- Coulibaly, P., Dibike, Y. B., & Anctil, F. (2005). Downscaling precipitation and temperature with temporal neural networks, *Journal of Hydrometeorology*, 6(4), 483–496. Available at: https://journals.ametsoc.org/view/journals/hydr/6/4/jhm409_1.xml [Accessed 1st February 2021].
- Cubasch, U., Von Storch, H., Waszkewitz, J. and Zorita, E. (1996) Estimates of climate change in southern Europe derived from dynamical climate model output. *Climate Research*, 7(2), 129–149.
- Daley, R. (1991) *Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series*. Cambridge: Cambridge University Press, p. 341.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beiljars, A.C.M., van den Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597.
- Dibike, Y.B. and Coulibaly, P. (2006) Temporal neural networks for downscaling climate variability and extremes. *Neural Networks*, 19(2), 135–144. <https://doi.org/10.1016/j.neunet.2006.01.003>.
- Ekström, M., Grose, M.R. and Whetton, P.H. (2015) An appraisal of downscaling methods used in climate change research. *WIREs Climate Change*, 2015(6), 301–319. <https://doi.org/10.1002/wcc.339>.
- García-Valero, J.A. (2021) Redes neuronales artificiales. Aplicación a la regionalización de la precipitación y temperaturas diarias. In: *AEMET Nota Técnica, Área de Evaluación y Modelización del Cambio Climático*. Spain: AEMET. https://www.aemet.es/documentos/es/conocerlas/recursos_en_linea/publicaciones_y_estudios/publicaciones/NT_34_Redets_neuronales_artificiales/NT_34_Redets_neuronales_artificiales.pdf.
- Ghosh, S. and Mujumdar, P. (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, 31, 132–146. <https://doi.org/10.1016/j.advwatres.2007.07.005>.
- Gutiérrez, J.M., Maraun, D., Widman, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanar, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Portoles, J., Raty, O., Raisanen, J., Hingray, B., Raynaud, D., Casado, M.J., Ramos, P., Zerener, T., Turco, M., Bosshard, T., Stepanek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A.M., Cardoso, R.M., Soares, P.M.M., Czernecki, B. and Page, C. (2019) An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*, 39, 3750–3785. <https://doi.org/10.1002/joc.5462>.
- Gutiérrez, J.M., San-Martín, D., Brands, S., Manzanar, R. and Herrera, S. (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, 26(1), 171–188.
- Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J.M., Wibig, J., Casanueva, A. and Soares, P.M.M. (2019) Comparison of statistical downscaling methods with respect to extreme events over Europe: validation results from the perfect predictor experiment of the COST action VALUE. *International Journal of Climatology*, 39, 3846–3867. <https://doi.org/10.1002/joc.5469>.
- Hoerl, A. and Kennard, R. (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67. <https://doi.org/10.1080/00401706.1970.10488634>.
- Hou, Y., Huang, X., Chen, H. and Xu, C.-Y. (2014) A statistical downscaling method based on least squares support vector machines. *Journal of Water Resources Research* (in Chinese with English abstract), 3, 72–77. <https://doi.org/10.12677/JWRR.2014.31012>.
- Huth, R. (2002) Statistical downscaling of daily temperature in Central Europe. *Journal of Climate*, 15, 1731–1742.

- Huth, R. (2004) Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate*, 17, 640–652.
- Lauri, H., De Moel, H., Ward, P.J., Räisänen, T.A., Keskinen, M. and Kummu, M. (2012) Future changes in Mekong River hydrology: impact of climate change and reservoir operation on discharge. *Hydrology and Earth System Sciences Discussions*, 9(5), 6569–6614.
- Lloyd, S.P. (1957) *Least-squares quantization in PCM*. Bell labs technical report. Eventually published in IEEE Trans. Inform. Thy. IT-28(2) 129, 1982.
- Lorenz, E.N. (1969) Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26(4), 636–646.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth BERKELEY Symposium on Mathematical Statistics and Probability*. Vol. 1, no. 14, pp. 281–297.
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M. and Hagemann, S. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), 764–773.
- Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M. and Thiele-Eich, I. (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3).
- Maraun, D. and Widmann, M. (2018) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge: Cambridge University Press.
- Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. (2015) Value: a framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1), 1–14.
- McCulloch, W.S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. <https://doi.org/10.1007/BF02478259>.
- Mendes, D. and Marengo, J. (2013) Temporal downscaling: a comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios. *Theoretical and Applied Climatology*, 100, 413–421. <https://doi.org/10.1007/s00704-009-0193-y>.
- Navascués, B., Rodríguez, E., Ayuso J.J. and Järvenoja, S. (2003) *Analysis of surface variables and parameterization of surface processes in HIRLAM. Part II: Seasonal assimilation experiment*. HIRLAM Technical Report 59. Available from <http://hdl.handle.net/20.500.11765/12003>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peel, M.C., Srikanthan, R., McMahon, T.A. and Karoly, D.J. (2015) Approximating uncertainty of annual runoff and reservoir yield using stochastic replicates of global climate model data. *Hydrology and Earth System Sciences*, 19, 1615–1639. <https://doi.org/10.5194/hess-19-1615-2015>.
- Peral, C., Navascués, B. and Ramos, P. (2017). Serie de precipitación diaria en rejilla con fines climáticos. Nota Técnica n° 24 AEMET. Available at: http://www.aemet.es/documentos/es/conocermas/recursos_en_linea/publicaciones_y_estudios/publicaciones/NT_24_AEMET/NT_24_AEMET.pdf.
- Petisco de Lara, S.E. (2008a) Método de regionalización de precipitación basado en análogos. Explicación y Validación. In: *AEMET Nota Técnica 3A, Área de Evaluación y Modelización del Cambio Climático*. Spain: AEMET.
- Petisco de Lara, S.E. (2008b) Método de regionalización de temperatura basado en análogos. Explicación y Validación. In: *AEMET Nota Técnica 3B, Área de Evaluación y Modelización del Cambio Climático*. Spain: AEMET.
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3), 61–74. Cambridge: MIT Press.
- Ribalaygua, J., Torres, L., Pórtoles, J., Monjo, R., Gaitán, E. and Pino, M.R. (2013) Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, 114(1–2), 253–269. <https://doi.org/10.1007/s00704-013-0836-x>.
- Rodríguez, E., Navascués, B., Ayuso, J.J. and Järvenoja, S. (2003) *Analysis of surface variables and parameterization of surface processes in HIRLAM. Part I: Approach and verification by parallel runs*. HIRLAM technical report 58. Available from <http://hdl.handle.net/20.500.11765/12004>.
- Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Rummukainen, M. (2010) State-of-the-art with regional climate models. *WIREs Climate Change*, 1(1), 82–96. <https://doi.org/10.1002/wcc.8>.
- Sachindra, D.A., Huang, F., Barton, A. and Perera, B.J.C. (2013) Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology*, 33, 1087–1106. <https://doi.org/10.1002/joc.3493>.
- Sailor, D., Hu, T., Li, X. and Rosen, J.N. (2000) A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. *Renewable Energy*, 19, 359–378. [https://doi.org/10.1016/S0960-1481\(99\)00056-7](https://doi.org/10.1016/S0960-1481(99)00056-7).
- Sailor, D.J. and Li, X. (1999) A semiempirical downscaling approach for predicting regional temperature impacts associated with climatic change. *Journal of Climate*, 12(1), 103–114. <https://doi.org/10.1175/1520-0442-12.1.103>.
- San-Martín, D., Manzanar, R., Brands, S., Herrera, S. and Gutiérrez, J.M. (2017) Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *Journal of Climate*, 30(1), 203–223.
- Schoof, J.T. (2013) Statistical downscaling in climatology. *Geography Compass*, 7(4), 249–265. <https://doi.org/10.1111/gec3.12036>.
- Shawe-Taylor, J. and Cristianini, N. (2004) *Kernel Methods for Pattern Analysis*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511809682>.
- Snell, S.E., Gopal, S. and Kaufmann, R.K. (2000) Spatial interpolation of surface air temperatures using artificial neural networks: evaluating their use for downscaling GCMs. *Journal of Climate*, 13(5), 886–895.

- Suykens, J. (2001) Support vector machines: a nonlinear modelling and control perspective. *European Journal of Control*, 7, 311–327. <https://doi.org/10.3166/ejc.7.311-327>.
- Suykens, J. and Vandewalle, J. (1999) Least squares support vector machine classifiers. *Neural Processing Letters*, 9, 293–300. <https://doi.org/10.1023/A:1018628609742>.
- Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems*. Washington, D.C; New York: V. H. Winston & Sons; John Wiley & Sons.
- Trigo, R. and Palutikof, J. (1999) Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. *Climate Research*, 13, 45–59. <https://doi.org/10.3354/cr013045>.
- Tripathi, S., Venkata, S. and Nanjundiah, R. (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330(3), 621–640. <https://doi.org/10.1016/j.jhydrol.2006.04.030>.
- Trzaska, S., and Schnarr, E. (2014). A Review of Downscaling Methods for Climate Change Projections: African and Latin American Resilience to Climate Change (ARCC). Available at: http://www.ciesin.org/documents/Downscaling_CLEARED_000.pdf.
- Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag.
- Vovk, V. (2013) Kernel ridge regression. In: *Empirical Inference*. New York City: Springer, pp. 105–116.
- Vrac, M., Stein, M., Hayhoe, K. and Liang, X.-Z. (2007) A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, 34. <https://doi.org/10.1029/2007GL030295>.
- Wetterhall, F. (2005). Statistical downscaling of precipitation from Large-scale Atmospheric circulation. Comparison of Methods and Climate Regions. Available at: <https://www.diva-portal.org/smash/get/diva2:166978/FULLTEXT01.pdf>.
- Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P. and Mearns, L. (2004) Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. Supporting material of the Intergovernmental Panel on Climate Change.
- Wilby R. L; Wigley, T. M. L. (1997). *Downscaling General Circulation Model Output: A Review of Methods and Limitations*. United Kingdom: Progress in Physical Geography: Earth and Environment 21 4 530-548.
- Wilby, R., Dawson, C. and Barrow, E.M. (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, 17, 145–157. [https://doi.org/10.1016/S1364-8152\(01\)00060-3](https://doi.org/10.1016/S1364-8152(01)00060-3).
- Yu, X. and Liang, S.-Y. (2007) Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology*, 332, 290–302. <https://doi.org/10.1016/j.jhydrol.2006.07.003>.
- Zorita, E. and von Storch, H. (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, 12, 2474–2489. [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).

SUPPORTING INFORMATION

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

How to cite this article: Hernanz, A., García-Valero, J. A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M. A., & Rodríguez-Camino, E. (2021). Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors. *International Journal of Climatology*, 1–15. <https://doi.org/10.1002/joc.7271>