RESEARCH ARTICLE

Evaluation of statistical downscaling methods for climate change projections over Spain: Future conditions with pseudo reality (transferability experiment)

Alfonso Hernanz[1] | Juan Andrés García-Valero[2] | Marta Domínguez[1] |
Ernesto Rodríguez-Camino[1]

[1]Spanish Meteorological Agency (AEMET), Madrid, Spain

[2]AEMET, Murcia, 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) and this is the last of three papers aimed to evaluate and intercompare five empirical/statistical downscaling (ESD) methods developed at AEMET: (a) Analog, (b) Regression, (c) Artificial Neural Networks, (d) Support Vector Machines and (e) Kernel Ridge Regression, in order to decide which methods and under what configurations are more suitable for that purpose. Following the framework established by the EU COST Action VALUE, in this experiment we test the transferability of these methods to future climate conditions with the use of regional climate models (RCMs) as pseudo observations. We evaluate the marginal aspects of the distributions of daily maximum/minimum temperatures and daily accumulated precipitation, over mainland Spain and the Balearic Islands, analysed by season. For maximum/minimum temperatures all methods display certain transferability issues, being remarkable for Support Vector Machines and Kernel Ridge Regression. For precipitation all methods appear to suffer from transferability difficulties as well, although conclusions are not as clear as for temperature, probably due to the fact that precipitation does not present such a marked signal of change. This study has revealed how an analysis over a historical period is not enough to fully evaluate ESD methods, so we propose that some type of analysis of transferability should be added in a standard procedure of a complete evaluation.

KEYWORDS
climate projections, evaluation, extrapolation, pseudo reality, Spain, statistical downscaling, transferability

1 | INTRODUCTION

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) and this is the last of three papers aimed to evaluate five empirical/statistical downscaling (ESD) methods developed at AEMET.

The methodology adopted for the complete evaluation follows the guidelines proposed by Vrac *et al.* (2007) and the EU COST Action VALUE (Maraun *et al.*, 2015) and consists of a set of three experiments: "Experiment 1: perfect predictor," "Experiment 2: global climate model (GCM) predictor" and "Experiment 3: pseudo reality." The first experiment (Hernanz *et al.*, 2021) showed good results by the five methods both for the mean values and the tails of maximum/minimum temperatures. As for precipitation, most methods were able to capture the total precipitation amount, but only a few methods and configurations seemed suitable for extreme events and for the precipitation occurrence. The second experiment (Hernanz *et al.*, 2021) showed how ESD methods presented a significant sensitivity to the use of imperfect predictors from GCMs and in this third paper we study the transferability of those methods to future conditions with the use of pseudo observations from regional climate models (RCMs) as predictands. Pseudo reality experiments are aimed to reveal whether predictors/predictand relationships found by ESD methods at a historical period are transferable to future climate conditions and whether predictors necessary to simulate the response to climate change have not been taken into account (Maraun *et al.*, 2015; Maraun and Widmann, 2018).

Pseudo reality experiments use RCMs not only to evaluate ESD methods but also to calibrate them. By calibrating and evaluating ESD methods with predictands of the same nature, small imperfections on GCM+RCMs are not expected to have a strong impact on the conclusions. This methodology somehow assumes that RCMs do not have transferability issues, which is not granted; although RCMs rely mainly on physical laws, they also contain empirical parameterizations which might vary under future climate conditions, and they are not usually evaluated in terms of their ability to capture the observed trends but in terms of their biases. Nevertheless, GCMs and RCMs here used have been widely evaluated over different regions of the globe at CMIP and CORDEX experiments, which enables certain confidence in their transferability to different climates. Pseudo reality experiments allow to check whether ESD methods are able to capture the climate change signal. However, these experiments represent a necessary but not sufficient condition, since there are differences between the real world and the pseudo reality world, and they cannot be used to evaluate those aspects which are not realistically simulated by the GCM+RCM, such as for example, extreme precipitation when using models that parameterize convection (Maraun and Widmann, 2018).

The use of pseudo observations to study transferability of ESD methods has been applied in the past. Charles *et al.* (1999) evaluated a nonhomogeneous hidden Markov model to downscale daily precipitation over Australia using pseudo observations from one RCM both in a historical period and in a $2 \times CO_2$ context. They found that its validation under present and $2 \times CO_2$ conditions were different and also that predictor selection played an important role in the transferability. Gaitan *et al.* (2014) evaluated a large ensemble of ESD methods to downscale daily precipitation over Canada in a historical period and under a future *SRES* A2 scenario (IPCC, 2000), also by using pseudo observations from one RCM. Among the numerous methods they evaluated there were different versions of Analog methods, Regression methods and Artificial Neural Networks, and they found that, although some methods performed similarly in present and future regarding the precipitation occurrence, errors related with the precipitation amount were bigger in future than in present for all methods. van der Linden and Mitchell (2009) also evaluated Analog methods, Regression methods and Artificial Neural Networks over Europe in a historical period and in a *SRES* A1B scenario (IPCC, 2000) using one RCM and they found that ESD methods achieved worse results as far in the future they were evaluated. Erlandsen *et al.* (2020) evaluated an ESD method based on "downscaling climate" instead of "downscaling weather," which consists in estimating parameters of the distributions instead of daily data (Benestad, 2021). They combined this approach with the use of common EOFs (Benestad, 2001; 2021) using a convective permitting RCM over the emissive scenario RCP8.5 (see IPCC, 2013) and found a significant sensitivity to the predictors choice and to the calibration period.

On the other hand, other approaches to study the transferability of ESD methods are possible. Gutiérrez *et al.* (2013) validated a large ensemble of ESD methods for maximum/minimum temperature over Spain by selecting the warmest years of a historical period. They found a significant underestimation by all methods compared to their validation in regular years. This underestimation was more marked in methods exclusively based on weather typing and analogs, and less marked in Multiple Linear Regression, both alone or combined with weather typing approaches. San-Martín *et al.* (2017) did the same for precipitation, validating over the driest years of a historical period, given that climate projections point

to a drier future climate in the region. They found a certain and similar overestimation by all methods, and they concluded that transferability was more sensitive to the set of predictors used than to the ESD method itself.

It is important to emphasize that, although the assumption of transferability is a well-known limitation of ESD methods, few evaluation studies include an analysis of transferability, and it is even less common to find pseudo reality studies using more than one RCM. For this evaluation we have broadened the methodology proposed by Vrac *et al.* (2007) to the comparison of seven different combinations of GCM+RCM, in order to allow an analysis of the sensitivity of the results to the GCM+RCM used.

The main objective of this paper is to evaluate the transferability of five ESD methods to future projected climate conditions and it 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, their different configurations and the methodology used for the analysis of results 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 and temporal periods have been used for this study.

Predictands (daily maximum/minimum temperature and 24 hr accumulated precipitation) come from RCMs listed in Table 1, driven by GCMs listed in Table 2. We have intentionally combined the same GCM with different RCMs and the same RCM with different GCMs, in order to allow the analysis of sensitivity to the GCM/RCM used, resulting in the seven combinations of GCM+RCM listed in Table 3 (two combinations have been discarded because of availability or reliability problems). These models, all of them participant in the EURO-CORDEX experiment (Jacob *et al.*, 2014), have been widely evaluated in this context (see Kotlarski *et al.*, 2014; Katragkou *et al.*, 2015; Dell'Aquila *et al.*, 2016; Vaittinada-Ayar *et al.*, 2016; Herrera *et al.*, 2020) and they have been selected because of their fairly good representation of the indexes used for this work. They cover the region of the study (Spain mainland and the Balearic Islands) with 3,357 grid

TABLE 1 Regional climate models (RCMs)

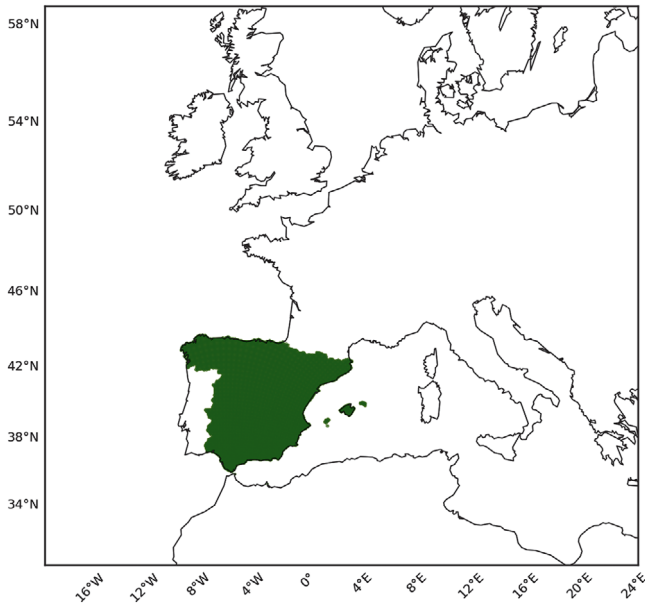| RCM	Institution	Reference
CNRM-ALADIN63	Centre National de Recherches Météorologiques (CNRM), France	Nabat *et al.* (2020)
DMI-HIRHAM5	Danish Meteorological Institute (DMI), Denmark	Bøssing Christensen *et al.* (2007)
KNMI-RACMO22E	Royal Netherlands Meteorological Institute (KNMI), Netherlands	Meijgaard *et al.* (2008)

TABLE 2 Global climate models (GCMs)

GCM	Institution	Reference
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique (CNRM-CERFACS), France	Voldoire *et al.* (2013)
IPSL-CM5A-MR	Institut National de l'Environnement Industriel et des Risques, Verneuil en Halatte, France/Institut Pierre Simon Laplace, CNRS (IPSL-INERIS), France	Dufresne *et al.* (2013)
MPI-ESM-LR	Max-Planck-Institut (MPI) for Meteorology, Germany	Giorgetta *et al.* (2013)

TABLE 3 GCM/RCM matrix of combinations

	CNRM-ALADIN63	DMI-HIRHAM5	KNMI-RACMO22E
CNRM-CM5	RCM1	RCM2	RCM3
IPSL-CM5A-MR		RCM4	RCM5
MPI-ESM-LR	RCM6		RCM7

points and a spatial resolution of 0.11° (see study area in Figure 1).

Predictors for calibration and evaluation come from the same GCMs used to drive those RCMs (see Table 2), all of them participants in the CMIP5 experiment (Taylor *et al.*, 2012), in the area (55.5°N, 30°N, 28.5°W,

15°E) with spatial resolution of 1.5° × 1.5° (see predictors domain at Figure S1, Supporting Information) and daily mean values. Their simulations correspond to the first realization, first initialization method and first physics (r1i1p1). In order to scale all predictors, they are standardized using their own mean and standard deviation over the period 1961–1990. 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. The sets of predictors used for each variable are listed in Table 4. Predictors in pseudo reality experiments are ideally taken from the RCM itself, because RCMs are constrained with the boundary conditions given by the driving GCM usually in large domains, so predictors from the GCM and from the RCM can be significantly different (Maraun and Widmann, 2018). However, we have used predictors from the GCMs, which allows us to analyse the impact of this effect.

For each GCM+RCM, the calibration period corresponds to 1961–1985, and there are two evaluation periods: 1986–2005 (present) and 2081–2100 (future). Data for calibration and for evaluation in a present period come from the historical run. Data used for the future evaluation period corresponds to the radiative forcing given by the RCP8.5 (see IPCC, 2013), which has been chosen for being the most extreme scenario in terms of climate change.

FIGURE 1 Area of study (Peninsular Spain and Balearic Islands) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 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 trend	Mean sea level pressure variation from the previous day	Mean sea level	Pa	Derived

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

Additionally, we have included some results from Experiment 1, in which predictors come from the reanalysis ERA-Interim (Dee *et al.*, 2011) of the European Centre for Medium-Range Weather Forecasts (ECMWF) and predictands are taken from a high-resolution observational grid developed by AEMET (Peral *et al.*, 2017) in 1980–2005.

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

The five ESD methods and their different configurations (see Table 5) are briefly presented here. For a more detailed description see Hernanz *et al.*, 2021. (a) Analog (ANA) methods (Lorenz, 1969; Zorita and von Storch, 1999) are based on the assumption of similar local conditions under similar synoptic situations, and one of their major drawbacks is their limitation to predict values outside of the observed range (Imbert and Benestad, 2005). This method uses Ug, Vg, U500 and

V500 (see Table 1) as large-scale fields and combines the synoptic analogy with local analogy in different ways. For precipitation, analog days can be selected following a nearest neighbour approach ("1"), a n-nearest neighbours approach ("N") or taking one randomly with a probability ("PDF") given by their analogy to the target day. It should be noticed that, for temperature, this method is a hybrid with a multiple linear regression (MLR), so this particular implementation does not suffer from the limitation commented above. (b) Regression (REG) consists in a MLR for temperature and a generalized linear model (GLM) for precipitation, based on the statistical downscaling model (SDSM; Wilby *et al.*, 2002). (c) Artificial Neural Networks (ANN) (McCulloch and Pitts, 1943; Rosenblatt, 1958) method uses a multilayer perceptron (Rosenblatt, 1958) both for temperature and precipitation. (4) Support Vector Machines (SVM) (Boser *et al.*, 1992; Cortes and Vapnik, 1995; Vapnik, 1995) uses different versions of SVMs for temperature and precipitation. (5) The last method consists on a combination of two specific types of SVMs: Kernel Ridge Regression (KRR) (Vovk, 2013) and Least-Square Support Vector Machine (LS-SVM) (Suykens and Vandewalle, 1999). None of the methods here presented make use of EOFs, but all of them relate predictors and predictands at grid point scale.

All methods make use of L2 regularization (Hoerl and Kennard, 1970; Tikhonov and Arsenin, 1977), and the tuning of the different parameters have been performed by cross-validation in a leave-one-out approach.

Other intercomparison studies of classical methods and machine learning (ML) techniques can be found in Zorita and von Storch (1999), Vandal *et al.* (2019) and Li *et al.* (2020).

3.2 | Diagnostics

Although Experiments 1 and 2 analysed both the mean values and the tails of the distributions for maximum/minimum temperatures, and three indexes for precipitation related to the total precipitation amount, intense precipitations and the precipitation occurrence, in this paper we only have use one index per variable in order to limit the analysis to the aspects best reproduced by RCMs. For maximum/minimum temperature we analyse their mean values (TXm and TNm, respectively) and for precipitation, its mean total amount (PRCPTOT). The precipitation occurrence has been discarded because of the well-known bias of RCMs to overestimate this magnitude with low intensities commonly known as *drizzle effect* (Gutowski Jr. *et al.*, 2003). And intense precipitation has also been excluded from the analysis because of possible

TABLE 5 Downscaling methods grouped by three families: Analog, Regression and machine learning methods

Family	Temperature	Precipitation
Analog	ANA	Synoptic analogy: ANA-SYN-1 (best analog) ANA-SYN-N (average) ANA-SYN-PDF (probability density function) Same but combined with local analogy: ANA-LOC-1 ANA-LOC-N ANA-LOC-PDF
Regression	REG	REG-LIN (linear) REG-EXP (exponential) REG-CUB (cubic)
Machine learning	ANN (Artificial Neural Network) SVM (Support Vector Machines) KRR (Kernel Ridge Regression)	ANN SVM KRR

misrepresentations by RCMs under climate change (Kendon *et al.*, 2014). For each of the selected indexes and for each GCM+RCM from Table 3, we have computed the bias, the mean error (ME) and the root mean square error (RMSE), in absolute terms for temperature and relative for precipitation, at both evaluation periods (present and future). Note that the three indexes correspond to temporal aggregations (mean or sum), so they result in a pair of pseudo-observed/downscaled values for each grid point in the whole period, which have been used to compute the biases. Thus, biases correspond to one value per grid point, while MEs and RMSEs summarize biases from all grid points in a single and spatially averaged value.

Additionally, a selection of cases of interest (combinations of GCM+RCMs and seasons) have been analysed through scatter plots of downscaled versus observed PRCPTOT, both in present and in future.

And finally, we have also analysed the climate change signal given by the pseudo observations and by the ESD methods for the three indexes; TXm, TNm and PRCPTOT, and we have included maps of the ensemble mean.

4 | RESULTS

The main results of each of the variables considered are presented in the following subsections. In order to allow a fluent analysis, we will use the terms good/bad transferability for similar/different behaviours between future and present, that is, for biases in the future of the same/different order to those in the present. This way, a method with low bias/ME/RMSE in future and high biases in present will be said to have bad transferability, while a method with high bias/ME/RMSE both in present and future will be said to have good transferability.

4.1 | Maximum/minimum temperature

For the mean values of maximum/minimum temperature, TXm and TNm respectively, the following results have been reached:

All methods display very low MEs in present for all GCM+RCMs, of the same order as those from Experiment 1 (Figure 2), which strengthens the methodology here used and the validity of using pseudo observations for temperature. This finding also lets us conclude that the impact of using predictors from GCMs instead of from RCMs (see section 2) is not significant. Nonetheless, this conclusion has been reached for a particular set of predictors and might not be the case for others. Furthermore, it should be noted that we are evaluating aggregated data, so at daily level the impact might be larger. In general, MEs are bigger in future than in present. All ESD methods display a wide range of MEs for the different GCM+RCMs in future, with higher values, in absolute terms, than those achieved in the historical period. Nevertheless, there are important differences among EDS methods, seasons and GCM+RCMs. SVM and KRR present the most remarkable transferability problems, with clear underestimations in JJA and SON for TNm and, less marked, although also significant, for TXm. ANN also displays an important underestimation for TNm in SON and, not so intensely, in JJA. And ANA and REG present more moderated transferability problems in general.
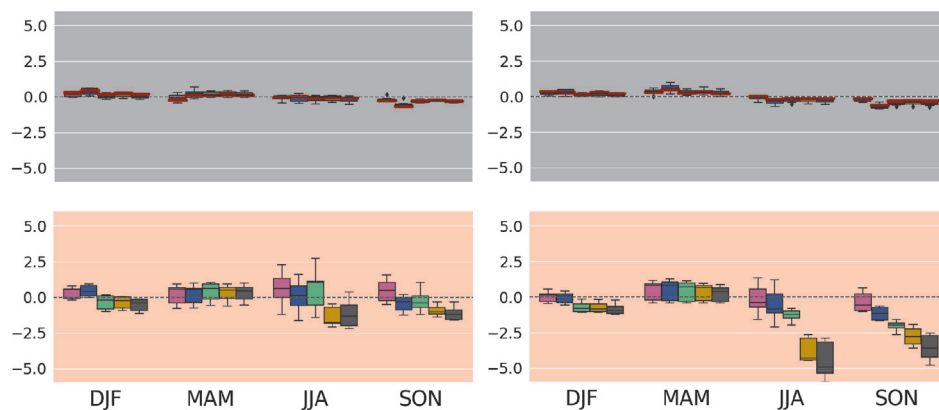
FIGURE 2 ME (°C) for mean value of the daily maximum (left column) and minimum (right column) temperatures by season under present conditions (1986–2005) at first row (grey background) and under future conditions (2081–2100, RCP8.5) at second row (red background). The methods are ANA (pink), REG (blue), ANN (green), SVM (orange) and KRR (grey). Each box contains the seven GCM+RCM combinations from Table 3 and red lines represent results from Experiment 1 (reanalysis+observational grid) [Colour figure can be viewed at wileyonlinelibrary.com]

It is important to bear in mind that MEs might be off-setting positive and negative spatial biases. These biases have been analysed but their sign and magnitude are quite dependent on the GCM+RCM, so no general conclusions can be drawn from them, apart from the fact that biases in future can reach extremely high values (see biases for RCM5 at Figure 3 as an example). RMSEs have also been analysed (not shown) but do not show additional marked errors apart from the ones revealed by the MEs.

Although the climate change signal has been analysed by seasons, we have included maps only for JJA (Figures 4 and S3), when this signal is stronger over the analysed area. In general, all methods show a positive sign of change in agreement to that shown by dynamical simulations, but differ in the intensity of change. REG captures the change in TXm and TNm very well, with quite low biases. ANA and ANN present some positive bias in the signal of change for TXm, while for TNm ANN presents a clear underestimation and ANA captures it with insignificant bias. And finally, SVM and KRR clearly underestimate the signal of change, both for TXm and TNm.

In summary, all methods present worse evaluation results in future than in present, which points to a lack of transferability by all of them. Nevertheless, two out of the five methods, SVM and KRR, have revealed very important transferability problems, much more marked than for the other methods.

4.2 | Precipitation

For precipitation, MEs in the historical period are, in general, of the same order and sign than those from Experiment 1, with the only exception of summer (Figure 5). In the other seasons, MEs over GCM+RCMs match perfectly with the ones over reanalysis for all ESD methods, reproducing even the general displacement to more negative MEs in SON, which supports the validity of RCMs as pseudo observations, and also the use of predictors from the GCMs instead of from the RCMs themselves (see section 2). Nevertheless, in JJA there is a slight mismatch, with more positive MEs over GCM+RCMs than over reanalysis. One possible cause comes from the fact that we are using relative errors, which makes the metric very sensitive to small imperfections during the dry season. Another possible explanation might be that summers in the historical evaluation period are slightly drier than in the calibration period (see Figure S1), so transferability issues start to emerge already. And finally, summer precipitation, which is basically convective, represents a special added difficulty both for ESD methods and RCMs, as downscaling links local effects to large-scale conditions, and convection is rarely a large-scale phenomenon.

MEs in future are similar to those in present in DJF and SON, with a slight trend to more positive MEs in MAM and a clear shift to overestimation in JJA. In this season, MEs display a wide spread, so JJA needs a deeper analysis over each GCM+RCM separately (see Figure 6). REG-EXP, REG-CUB and SVM are excluded from the discussion because of their systematic underestimation of PRCPTOT, even at Experiment 1. Figure 6 reveals a clear distinction between RCMs 1–3 and 4–7. Furthermore, when analysing biases over each GCM+RCM separately (not shown), RCMs 4, 5 and 6 display much higher biases in future JJA than the other GCM+RCMs. And finally, Figure S1 shows how RCMs 4, 5 and 6 present a stronger signal of change in PRCPTOT in JJA than the other GCM+RCMs. For these reasons we focus the analysis henceforth on three cases of interest (RCMs 4, 5 and 6 in JJA). In these three study cases, all methods appear to display a tendency to overestimate in future (Figure 6), less marked for KRR, whereas in the other combinations of GCM+RCM, where there is barely signal of change, biases by ESD methods are of the same order in present and future. Nevertheless, these three study cases correspond to very low PRCPTOT values given by the GCM+RCMs, so the use of relative errors can easily lead
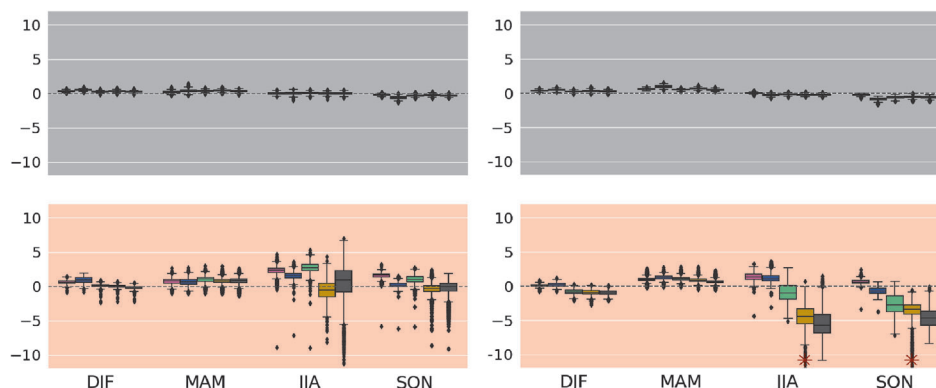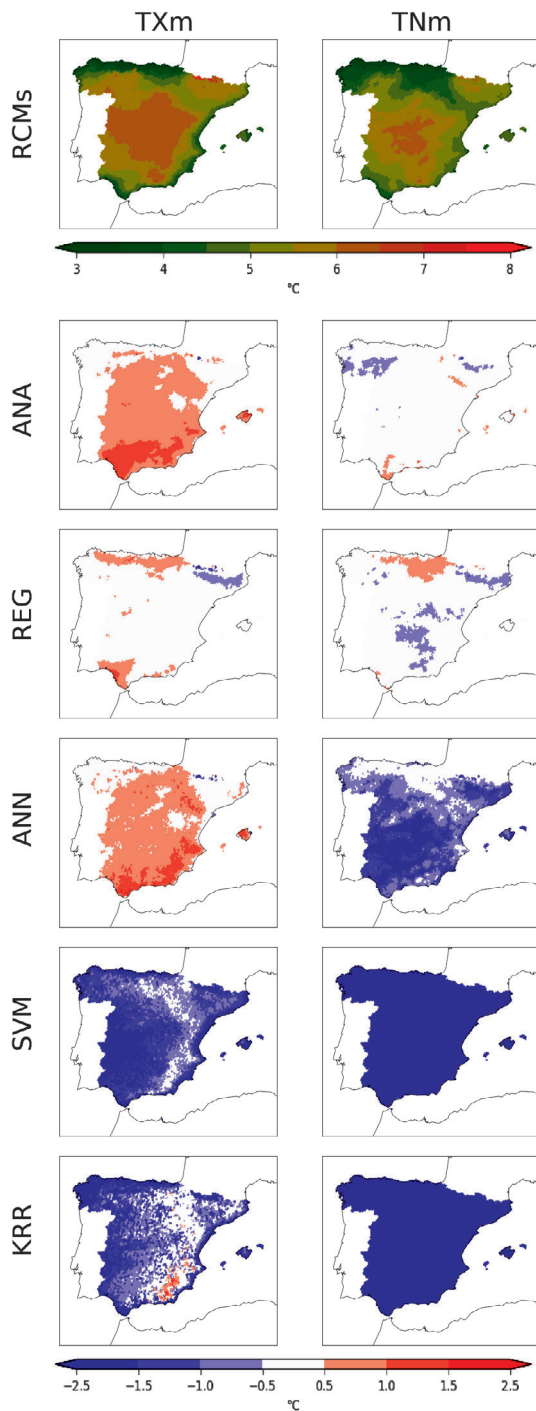

FIGURE 3 Same as Figure 2, but for bias (°C) over RCM5 (IPSL-CM5A-MR+KNMI-RACMO22E). Each box contains 3,357 grid points. A red asterisk indicates that values lie outside the plotted range [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 4 Change (°C) for TXm (left column) and TNm (right column) in JJA given by the pseudo reality (first row), and biases (°C) by the ESD methods: ANA (second row), REG (third row), ANN (fourth row), SVM (fifth row) and KRR (sixth row). The seven GCM+RCMs have been summarized by their ensemble mean. Maps of change by the ESD methods have been included in Figure S3 [Colour figure can be viewed at wileyonlinelibrary.com]

to very high MEs. Furthermore, as it has been mentioned, low MEs might be hiding positive and negative biases by compensating them. In order to avoid these
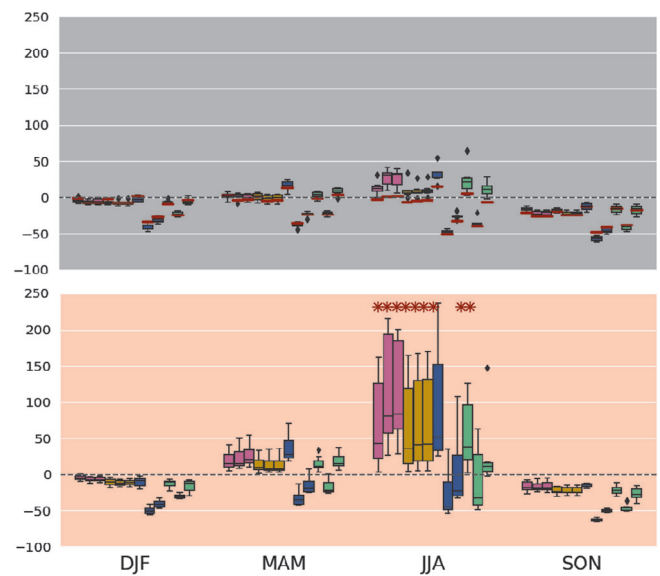
FIGURE 5 ME over relative errors (%) for PRCPTOT by season under present conditions (1986–2005) at first row (grey background) and under future conditions (2081–2100, RCP8.5) at second row (red background). Methods coloured by families: RAW (grey), ANA-SYN-1, ANA-SYN-N and ANA-SYN-PDF (pink), ANA-LOC-1, ANA-LOC-N and ANA-LOC-PDF (yellow), REG-LIN, REG-EXP and REG-CUB (blue), ANN, SVM and KRR (green). Each box contains the seven GCM+RCM combinations from Table 3, and red lines represent results from Experiment 1 (reanalysis+observational grid). A red asterisk indicates that values lie outside the plotted range [Colour figure can be viewed at wileyonlinelibrary.com]
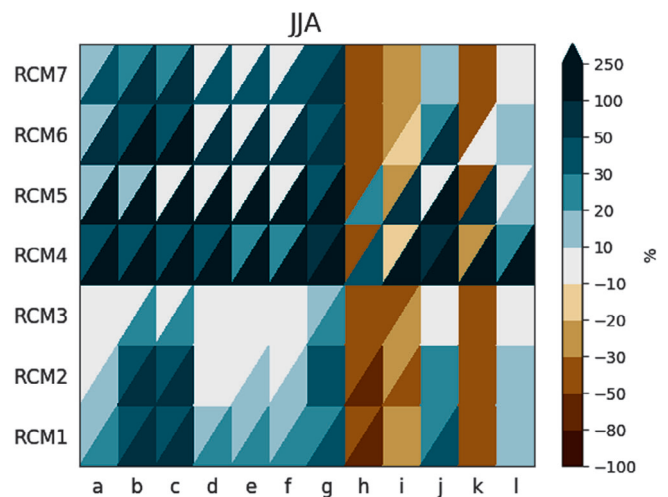
FIGURE 6 ME over relative errors (%) for PRCPTOT in JJA. Rows correspond to the seven GCM+RCM combinations from Table 3 (in order from the bottom to the top) and columns correspond to ESD methods: ANA-SYN-1 (a), ANA-SYN-N (b), ANA-SYN-PDF (c), ANA-LOC-1 (d), ANA-LOC-N (e), ANA-LOC-PDF (f), REGLIN (g), REG-EXP (h), REG-CUB (i), ANN (j), SVM (k) and KRR (l). Upper left triangle of each cell corresponds to present conditions (1986–2005) and bottom right corresponds to future conditions (2081-2100, RCP8.5) [Colour figure can be viewed at wileyonlinelibrary.com]

problems, we have performed an additional analysis in the form of scatter plots for ANA-SYN-1 (as a representative of the Analog family), REG-LIN, ANN and KRR. Figure 7 shows how ANA-SYN-1 and REG-LIN tend to overestimate PRCPTOT in the future period of the three selected cases, which confirms their difficulty to reproduce drier future climates. And for ANN and KRR, although they capture PRCPTOT with great accuracy in the present period in the three cases, their behaviours in future are very different. These two machine learning methods do not present a clear tendency to overestimation as did ANA-SYN-1 and REG-LIN, but their scatter plots reveal a significant spread around the diagonal, a sign of lack of accuracy.

In order to analyse the signal of change we have also focused on JJA (Figures 8 and S4). The Analog methods are not able to capture the dry conditions projected by the GCM+RCMs, and portray a future with barely a signal of change at all. Regression methods appear to capture the change in PRCPTOT better than Analog methods, although they do not present such dry conditions as the GCM+RCMs do, especially the REG-LIN option. And finally, of the three machine learning methods, ANN and KRR appear to capture the signal of change fairly well, while SVM presents a significant positive change in PRCPTOT in the south of Spain that clearly does not match the pattern given by the GCM+RCMs.

In summary, precipitation does not present as marked a signal of change as temperature does, so the analysis of transferability is focused on a selection of cases of interest, and they reveal transferability problems in all ESD methods. For Analog methods and for REG-LIN, a tendency to overestimate PRCPTOT in a drier future climate has been detected. For machine learning methods a certain deterioration when applied to a drier climate is also seen, especially for SVM, although with no systematic bias.
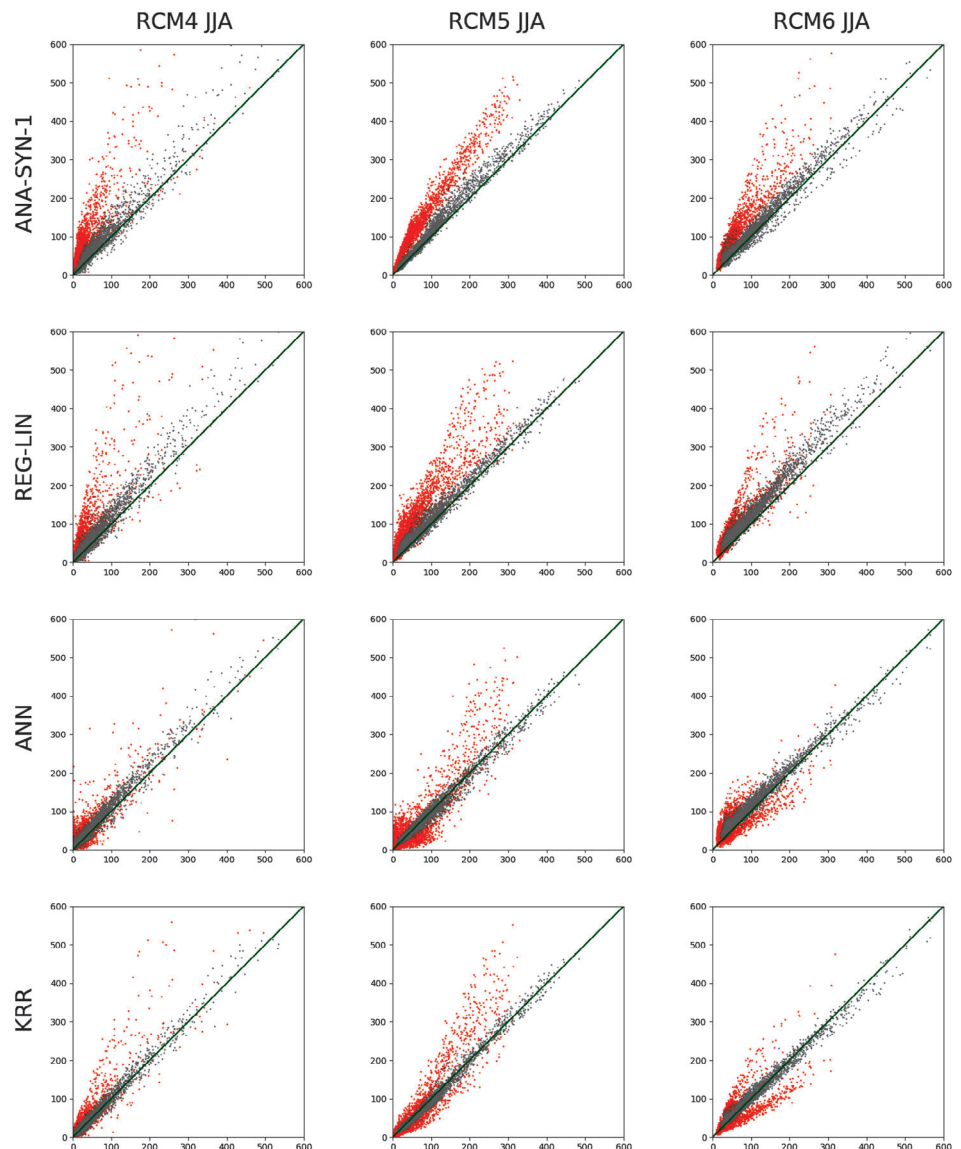
FIGURE 7 Statistical downscaling (y axis) versus pseudo observations (x axis) of PRCPTOT (mm) for a selection of special cases: RCM4 in JJA (first column), RCM5 in JJA (second column) and RCM6 in JJA (third column). Present conditions (1986–2005) in grey and future conditions (2081–2100, RCP8.5) in red. Methods by rows: ANA-SYN-1, REG-LIN, ANN and KRR [Colour figure can be viewed at wileyonlinelibrary.com]
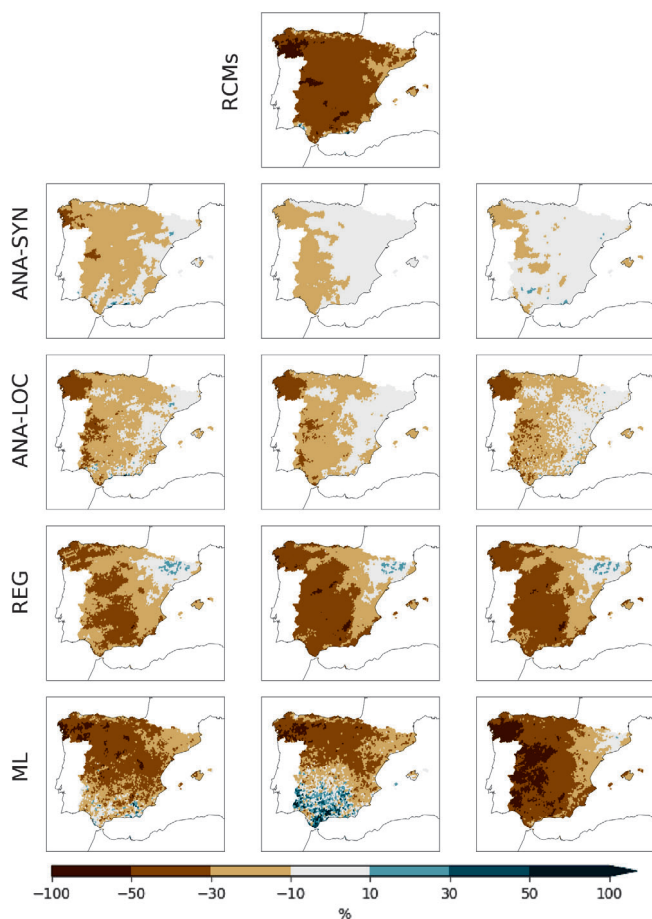
FIGURE 8 Relative change (%) for PRCPTOT in JJA given by the ensemble mean of the seven GCM+RCMs. Pseudo reality (first row), ANA-SYN-1, ANA-SYN-N and ANA-SYN-PDF (second row, from left to right), ANA-LOC-1, ANA-LOC-N and ANA-LOC-PDF (third row, from left to right), REG-LIN, REG-EXP and REG-CUB (fourth row, from left to right), and machine learning (ML) methods (fifth row, ANN, SVM and KRR from left to right). Maps of the bias by ESD methods have been included in Figure S4 [Colour figure can be viewed at wileyonlinelibrary.com]

5 | CONCLUDING REMARKS

In this paper we have evaluated the transferability to future climate conditions of five statistical downscaling techniques using pseudo observations from RCMs as predictands, and by comparing their evaluation results in a historical and a future period. We have extended the methodology proposed at Vrac *et al.* (2007) using seven GCM+RCM combinations. Additionally, results in the historical evaluation period have been compared with those from Experiment 1 in order to identify possible sources of uncertainty introduced by the use of pseudo observations.

For maximum/minimum temperatures, the five methods have revealed some transferability issues, which

aligns with the findings by van der Linden and Mitchell (2009) and Gutiérrez *et al.* (2013). This lack of transferability has been found to be very remarkable in the cases of SVM and KRR. Experiments 1 and 2 (Hernanz *et al.*, 2021, 2021b) showed how these two methods, which are able to reproduce complex nonlinear relationships and are based on different types of Support Vector Machines, could achieve fairly good results under present conditions, overcoming the linear method REG both with perfect and imperfect predictors, but this study has revealed some important transferability issues in them. This relates with the well-known problem of machine learning algorithms to deal with new situations to which they have not been trained, and calls in question their suitability for downscaling climate projections, as pointed out by Hsieh (2009). On the other hand, the other machine learning method, ANN, presents transferability issues not as marked, generally similar to those by ANA and REG with few exceptions. Considering conclusions from Experiments 1 and 2, in which ANA and ANN usually reached better results than REG, it seems reasonable to use both of them to elaborate the climate projections, so the uncertainty introduced by the downscaling technique is taken into account.

For precipitation, being the signal of change not as marked as for temperature, transferability problems are not as easy to detect. The study of the three cases with the most marked change in PRCPTOT has revealed difficulties for all methods to represent a drier future climate, which aligns with the findings by Gaitan *et al.* (2014), van der Linden and Mitchell (2009) and San-Martín *et al.* (2017). Analog methods display a positive bias in the three cases, confirming the difficulty for these methods to represent different future climate conditions. Also REG-LIN has shown the same overestimation when applied to a drier future climate, and the machine learning algorithms, ANN and KRR, have revealed a certain deterioration in capturing PRCPTOT under future drier conditions when compared with their results in the historical period. The other methods, REG-EXP, REG-CUB and SVM, have not been analysed because of their lack of accuracy for PRCPTOT at Experiment 1. Experiments 1 and 2 showed how Analog methods were able to capture the total precipitation amount, the precipitation occurrence and intense precipitations, while transfer function methods appear only suitable for the total precipitation amount. Nevertheless, with the transferability problems revealed by all methods, it seems reasonable to use, at least, one method of each family to generate the climate projections. No Analog method has proved clearly better than the simplest form, ANA-SYN-1, and for Regression methods, REG-LIN has proved the best configuration. For machine learning techniques,

although ANN and KRR have achieved similar results in the three experiments, the fact that KRR displays much more marked transferability issues than ANN for temperature suggests that ANN might also be a better choice for precipitation.

Literature indicates that predictor selection plays a key role in the transferability of ESD methods (see, e.g., Parding *et al.*, 2019). Being the purpose of this evaluation the comparison of ESD methods, and considering the high computational cost of the whole methodology, we have limited it to a unique set of predictors for each variable. Nevertheless, a systematic analysis replicating these three experiments but for different sets of predictors might constitute an interesting future work.

Additionally, it should be noticed that all ESD methods here evaluated operate at a daily scale ("downscaling weather"), but "downscaling climate" appears to be a promising approach, as parameters of the distributions are usually easier to predict than daily states and transferability issues might be palliated (Erlandsen *et al.*, 2020; Benestad, 2021).

And finally, it is important to point out that transferability problems might not come exclusively from ESD, but also from the GCMs themselves. Correlation between different parameters simulated by GCMs might be non-stationary (Wilby and Wigley, 2000; Vrac *et al.*, 2021), and also GCMs parameterizations may have been adjusted and calibrated making use of observational dataset with the corresponding transferability issues.

In summary, this study, together with Experiments 1 and 2, has allowed us to thoroughly evaluate five ESD methods over the studied region, revealing their strengths and weaknesses. This particular experiment, the third one, has highlighted how ESD methods can perform very differently under present and future climate conditions, which aligns with the existing literature regarding this issue, and supports the idea that a complete evaluation must include some type of transferability analysis.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Alfonso Hernanz: Software; visualization; writing-original draft. **Juan García-Valero:** Writing-review & editing. **Marta Domínguez:** Writing-review & editing. Ernesto **Rodriguez-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

Benestad, R.E. (2001) A comparison between two empirical downscaling strategies. *International Journal of Climatology*, 21, 1645–1668. https://doi.org/10.1002/joc.703.

Benestad, R.E. (2021) A Norwegian approach to downscaling. *Geoscientific Model Development Discussion* (in review). https://doi.org/10.5194/gmd-2021-176.

Boser, B., Guyon, I. and Vapnik, V. (1992) A training algorithm for optimal margin classifier. In: *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, Vol. 5. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/130385.130401.

Bøssing Christensen, O., Drews, M., Hesselbjerg Christensen, J., Dethloff, K., Ketelsen, K., Hebestadt, I. and Rinke, A. (2007) *The HIRHAM Regional Climate Model. Version 5 (beta).*

Copenhagen: Danish Climate Centre, Danish Meteorological Institute. Technical report number: 06-17. Available at: http://www.dmi.dk/dmi/tr06-17.

Charles, S., Bates, B., Whetton, P. and Hughes, J. (1999) Validation of downscaling models for changed climate conditions: case study of southwestern Australia. *Climate Research*, 12, 1–14. https://doi.org/10.3354/cr012001.

Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20, 273–297. https://doi.org/10.1007/BF00994018.

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., Beeljars, 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, L., 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. https://doi.org/10.1002/qj.828.

Dell'Aquila, A., Mariotti, A., Bastin, S., Calmanti, S., Cavicchia, L., Deque, M., Djurdjevic, V., Dominguez, M., Gaertner, M. and Gualdi, S. (2016) Evaluation of simulated decadal variations over the Euro-Mediterranean region from ENSEMBLES to Med-CORDEX. *Climate Dynamics*, 51, 857–876. https://doi.org/10.1007/s00382-016-3143-2.

Dufresne, JL., Foujols, MA., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N. and Vuichard, N. (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, 40, 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Erlandsen, H.B., Parding, K.M., Benestad, R., Mezghani, A. and Pontoppidan, M. (2020) A hybrid downscaling approach for future temperature and precipitation change. *Journal of Applied Meteorology and Climatology*, 59(11), 1793–1807. https://doi.org/10.1175/JAMC-D-20-0013.1.

Gaitan, C., Hsieh, W. and Cannon, A. (2014) Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada. *Climate Dynamics*, 43, 1–17. https://doi.org/10.1007/s00382-014-2098-4.

Giorgetta, M.A., Jungclaus, J., Reick, C.H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T.,

Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K.D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B. (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5, 572–597. https://doi.org/10.1002/jame.20038

Gutiérrez, J.M., San-Martín, D., Brands, S., Manzanas, R. and Herrera, S. (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, 26(1), 171–188. https://doi.org/10.1175/JCLI-D-11-00687.1.

Gutowski, W.J., Jr., Decker, S.G., Donavon, R.A., Pan, Z., Arritt, R. W. and Takle, E.S. (2003) Temporal spatial scales of observed and simulated precipitation in central US climate. *Journal of Climate*, 16, 3841–3847. https://doi.org/10.1175/1520-0442(2003)016<3841:TSOOAS>2.0.CO;2.

Hernanz, A., García-Valero, J.A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M.A. and 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.

Herrera, S., Soares, P.M.M., Cardoso, R.M. and Gutierrez, J.M. (2020) Evaluation of the EURO-CORDEX regional climate models over the Iberian Peninsula: observational uncertainty analysis. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032880. https://doi.org/10.1029/2020JD032880.

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.

Hsieh, W. (2009) *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627217.

Imbert, A. and Benestad, R. (2005) An improvement of analog model strategy for more reliable local climate change scenarios. *Theoretical and Applied Climatology*, 82, 245–255. https://doi.org/10.1007/s00704-005-0133-4.

IPCC. (2000) *Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

IPCC. (2013) In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (Eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge and New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R.,

Weber, B. and Yiou, P. (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14(2), 563–578. https://doi.org/10.1007/s10113-013-0499-2.

Katragkou, E., Garciá-Diéz, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R.M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P.M.M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A. and Jacob, D. (2015) Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. *Geoscientific Model Development*, 8(3), 603–618. https://doi.org/10.5194/gmd-8-603-2015.

Kendon, E., Roberts, N., Fowler, H., Roberts, M., Chan, S. and Senior, C. (2014) Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4, 570–576. https://doi.org/10.1038/nclimate2258.

Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V. (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, 7(4), 1297–1333. https://doi.org/10.5194/gmd-7-1297-2014.

Li, X., Li, Z., Huang, W. and Zhou, P. (2020) Performance of statistical and machine learning ensembles for daily temperature downscaling. *Theoretical and Applied Climatology*, 140, 571–588. https://doi.org/10.1007/s00704-020-03098-3.

Lorenz, E.N. (1969) Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26(4), 636–646.

Maraun, D. and Widmann, M. (2018) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107588783.

Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. (2015) Value: a framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1), 1–14. https://doi.org/10.1002/2014EF000259.

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.

Meijgaard, E., Ulft, L.H., Berg, W.J., Bosvelt, F.C., Hurk, B., Lenderink, G. and Siebesma, A.P. (2008) *The KNMI regional atmospheric model RACMO version 2.1*. KNMI. Technical report number: 302.

Nabat, P., Somot, S., Cassou, C., Mallet, M., Michou, M., Bouniol, D., Decharme, B., Drugé, T., Roehrig, R. and Saint-Martin, D. (2020) Modulation of radiative aerosols effects by atmospheric circulation over the Euro-Mediterranean region. *Atmospheric Chemistry and Physics*, 20, 8315–8349. https://doi.org/10.5194/acp-20-8315-2020.

Parding, K.M., Benestad, R., Mezghani, A. and Erlandsen, H.B. (2019) Statistical projection of the North Atlantic storm tracks. *Journal of Applied Meteorology and Climatology*, 58(7), 1509–1522. https://doi.org/10.1175/JAMC-D-17-0348.1.

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. https://doi.org/10.5555/1953048.2078195.

Peral, C., Navascués, B. and Ramos, P. (2017) *Serie de precipitación diaria en rejilla con fines climáticos*. Madrid, Spain: AEMET. Nota Técnica n° 24. Available at: http://www.aemet.es/documentos/es/conocermas/recursos_en_linea/publicaciones_y_estudios/publicaciones/NT_24_AEMET/NT_24_AEMET.pdf.

Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.

San-Martín, D., Manzanas, 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. https://doi.org/10.1175/JCLI-D-16-0366.1.

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.

Taylor, K., Ronald, S. and Meehl, G. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society.*, 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1.

Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems*. Washington, DC: V. H. Winston & Sons.

Vaittinada-Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M. and Gallardo, C. (2016) Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations. *Climate Dynamics*, 46(3–4), 1301–1329. https://doi.org/10.1007/s00382-015-2647-5.

van der Linden, P. and Mitchell, J.F.B. (2009) *ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project*. Technical report. Exeter: Met Office Hadley Center.

Vandal, T., Kodra, E. and Ganguly, A.R. (2019) Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137, 557–570. https://doi.org/10.1007/s00704-018-2613-3.

Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*. Berlin-Heidelberg: Springer.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L. and Chauvin, F. (2013) The CNRM-CM5.1 global climate model: description and basic evaluation. *Climate Dynamics*, 40, 2091–2121. https://doi.org/10.1007/s00382-011-1259-y

Vovk, V. (2013) Kernel ridge regression. In: *Empirical Inference*. New York, NY: 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, L18701. https://doi.org/10.1029/2007GL030295.

Vrac, M., Thao, S. and Yiou, P. (2021) Changes in temperature-precipitation correlations over Europe: Are climate models

reliable? *Climate Dynamics* (in press). https://doi.org/10.21203/rs.3.rs-1008080/v1.

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.

Wilby, R. and Wigley, T. (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology*, 20, 641–661. https://doi.org/10.1002/(SICI)1097-0088(200005)20:6<641::AID-JOC501>3.0.CO;2-1.

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.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hernanz, A., García-Valero, J. A., Domínguez, M., & Rodríguez-Camino, E. (2021). Evaluation of statistical downscaling methods for climate change projections over Spain: Future conditions with pseudo reality (transferability experiment). *International Journal of Climatology*, 1–14. https://doi.org/10.1002/joc.7464