Statistical Downscaling in the Tropics and Midlatitudes: A Comparative Assessment over Two Representative Regions

ALFONSO HERNANZ[a], CARLOS CORREA[a], MARTA DOMÍNGUEZ[a], ESTEBAN RODRÍGUEZ-GUISADO[a], AND
ERNESTO RODRÍGUEZ-CAMINO[a]

[a] *Spanish Meteorological Agency (AEMET), Madrid, Spain*

ABSTRACT: Statistical downscaling (SD) of climate change projections is a key piece for impact and adaptation studies due to its low computational expense compared to dynamical downscaling, which allows exploration of uncertainties through the generation of large ensembles. SD has been extensively evaluated and applied in the extratropics, but few examples exist in tropical regions. In this study, several state-of-the-art methods belonging to different families have been evaluated for maximum/minimum daily temperature and daily accumulated precipitation (both from the ERA5 at 0.25°) in two regions with very different climates: Spain (midlatitudes) and Central America (tropics). Some key assumptions of SD have been tested: the strength of the predictor–predictand links, the skill of different approaches, and the extrapolation capability of each method. It has been found that relevant predictors are different in both regions, as is the behavior of statistical methods. For temperature, most methods perform significantly better in Spain than in Central America, where transfer function (TF) methods present important extrapolation problems, probably due to the low variability of the training sample (present climate). In both regions, model output statistics (MOS) methods have achieved the best results for temperature. In Central America, TF methods have achieved better results than MOS methods in the evaluation in the present climate, but they do not preserve trends in the future. For precipitation, MOS methods and the extreme gradient boost machine learning method have achieved the best results in both regions. In addition, it has been found that, although the use of humidity indices as predictors improves results for the downscaling of precipitation, future trends given by statistical methods are very sensitive to the use of one or another index. Three indices have been compared: relative humidity, specific humidity, and dewpoint depression. The use of the specific humidity has been found to lead to trends given by the downscaled projections that deviate seriously from those given by raw global climate models in both regions.

KEYWORDS: Extratropics; Tropics; Climate change; Downscaling; Machine learning

1. Introduction

Global warming is unequivocal according to the Intergovernmental Panel on Climate Change (IPCC), and current and future climate change strongly depend on human actions (IPCC 2021). The effects of climate change in different regions are a fundamental piece for impact and adaptation studies. Nevertheless, global climate models (GCMs), the main tool to simulate future climate, often have too coarse a resolution for this purpose, and some type of downscaling is needed (Charles et al. 2004; Wilby et al. 2004; Schoof 2013). There are two main strategies for downscaling: dynamical downscaling and statistical downscaling (SD). Dynamical downscaling relies on the same physical principles as GCMs, and it mainly consists of nesting a high-resolution regional climate model in a GCM. On the other hand, SD is based on the existence of empirical–statistical relationships between large-scale predictors and local-scale predictands, which

are assumed to be maintained under future climate change. Both approaches present advantages and disadvantages, and they have been widely reviewed (Wilby and Wigley 1997; Charles et al. 2004; Wilby et al. 2004; Rummukainen 2010; Trzaska and Schnarr 2014). For example, SD does not always ensure physical consistency, opposite to dynamical downscaling. On the other hand, SD is in general less computationally expensive compared to dynamical downscaling. Thus, SD allows exploration of uncertainties through large ensembles, which has promoted its use for impact and adaptation studies (Trzaska and Schnarr 2014).

There is a huge variety of statistical downscaling models (SDMs), belonging to different families (see Wilby et al. 2004; Trzaska and Schnarr 2014; Maraun and Widmann 2018). In the perfect prognosis (PP) approach, SDMs are calibrated establishing statistical relationships between predictors given by a reanalysis and observations, and then, they are applied over predictors simulated by GCMs. PP relies on the assumptions that predictors are well simulated by GCMs both in present and future climates and that a strong link between predictors and predictands exists. On the other hand, the model output statistics (MOS) approach corrects biases from GCMs by adjusting different aspects of the simulated distributions and usually assuming the stationarity of model biases in the future. Finally, weather generators (WGs) are stochastic models able to produce synthetic series matching their marginal and temporal aspects with climatological statistics conditioned on properties given by GCM simulations.
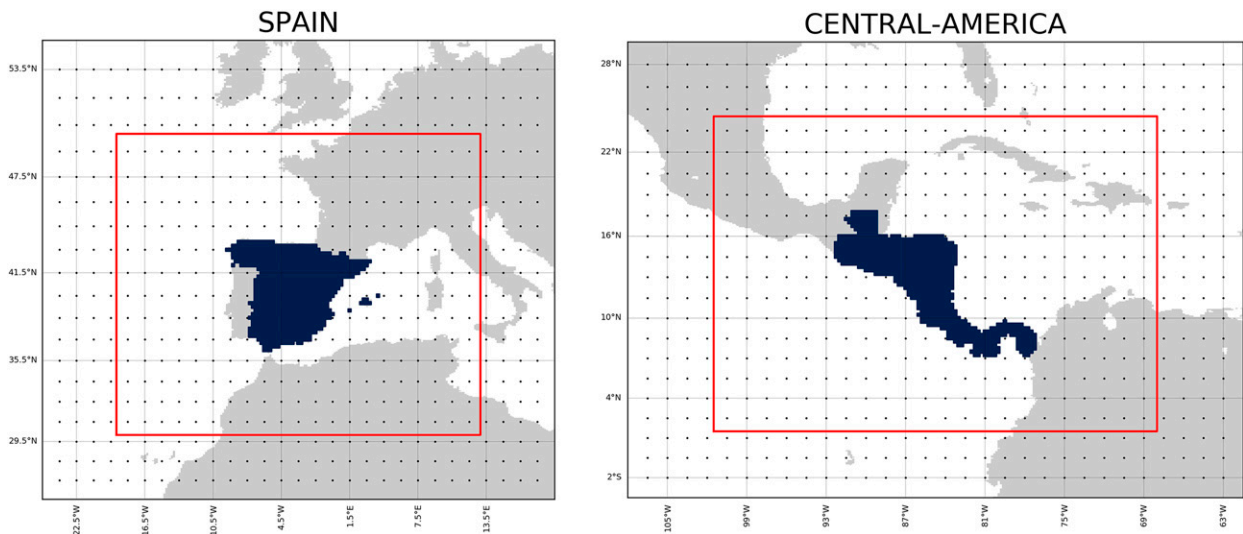
FIG. 1. Target regions (blue) and spatial domains for synoptic analogy fields (red outline) in (left) Spain and (right) Central America.

The European Union (EU) Cooperation in Science and Technology (COST) Action VALUE (Maraun et al. 2015) proposed a comprehensive framework for evaluating and intercomparing the different SD approaches. This framework consists of the evaluation of different aspects (marginal, temporal, spatial, and intervariable) under different experiments: using "perfect" predictors from a reanalysis, using "imperfect" predictors from GCMs, and using predictands from regional climate models as "pseudo-observations." In this context, a large ensemble of SDMs belonging to different families was evaluated under a perfect predictor experiment approach (i.e., using perfect predictors from a reanalysis), and the key findings and remaining questions were summarized by Maraun et al. (2019). This comprehensive evaluation exercise was performed over different stations all over Europe, capturing very different climates. Nevertheless, no tropical region was included in it.

SD has been extensively evaluated and applied in the extratropics, and comprehensive studies such as the aforementioned VALUE have intercompared the main families of SDMs. While SD has also been evaluated and applied in tropical regions (see, e.g., Cavazos and Hewitson 2005; Ramseyer and Mote 2016), no intercomparison exercise between SDMs representative of the different families has been found. Some key limitations for SD in the tropics have been suggested. As pointed out by Manzanas et al. (2015), while in extratropical regions a large fraction of the local climate variability can be explained by well-simulated large-scale structures, characterized by the quasigeostrophic coupling between wind and mass fields, such as extratropical cyclones and their associated fronts, atmospheric drivers in the tropics operate in finer scales and are poorly simulated.

In this study, we have performed a comparative evaluation of the different families of SDMs over one tropical region and another one in the midlatitudes, with the purpose of analyzing the suitability of SD in the tropics and which methods are more skillful in each region. The objective of this paper is to present a first overview on the matter to be used as a general guide by the downscaling community.

This work is organized as follows. In section 2, the datasets used are described. In section 3, a brief introduction to the SDMs employed and the diagnostics performed is given. In section 4, results are presented. In section 5, the main conclusions are summarized.

2. Data

For this study, we have chosen two regions especially vulnerable to climate change, where a future drier climate is projected by GCMs: Spain, as part of the Mediterranean Sea region, and Central America (see Fig. 1). This choice has been motivated as an extension of a comprehensive evaluation recently performed for Spain (Hernanz et al. 2022a,d,e) to Central America, where the Spanish Meteorological Agency (AEMET) collaborates in the generation of regional climate change information through the EUROCLIMA+ project (Comisión Europea 2021; https://euroclimaplus.org/).

As predictands, we have used maximum/minimum daily temperatures and daily accumulated precipitation from the fifth major global reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; ERA5; Hersbach et al. 2020) with a spatial resolution of 0.25°, consisting of 845 points over Spain and 828 over Central America. Both regions present complex topography, so the resolution here used (0.25°) is not expected to accurately reproduce local climate in the finer scales. While the use of a reanalysis introduces an additional source of uncertainty, the lack of available observations has motivated its use instead. Predictors have been taken from the same reanalysis but with a resolution of 1.5° and as daily averages using data from the 0000, 0600, 1200, and 1800 UTC and from the 10 GCMs listed in Table 1, all of them participants in phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016), interpolated from their original resolutions to 1.5° using a bilinear interpolation and as mean daily values.

TABLE 1. GCMs from CMIP6: Model name, institution, horizontal resolution, and reference.

Model	Institution	Resolution	References
ACCESS-CM2	CSIRO and Bureau of Meteorology (BoM), Australia	$1.9° \times 1.3°$	Bi et al. (2020)
CanESM5	Canadian Centre for Climate Modelling and Analysis, Canada	$2.8° \times 2.8°$	Swart et al. (2019)
EC-EARTH3	EC-Earth Consortium, Europe	$0.7° \times 0.7°$	Döscher et al. (2022)
INM-CM5-0	Institute of Numerical Mathematics, Russia	$2° \times 1.5°$	Volodin et al. (2017)
INM-CM4-8	Institute of Numerical Mathematics, Russia	$2° \times 1.5°$	Volodin et al. (2013)
IPSL-CM6A-LR	L'Institut Pierre-Simon Laplace (IPSL), France	$2.5° \times 1.3°$	Boucher et al. (2020)
MIROC6	Research Center for Environmental Modeling and Application, Japan	$1.4° \times 1.4°$	Tatebe et al. (2019)
MPI-ESM1-2-LR	Max Planck Institute (MPI) for Meteorology, Germany	$0.9° \times 0.9°$	Mauritsen et al. (2019)
MPI-ESM1-2-HR	Max Planck Institute (MPI) for Meteorology, Germany	$1.9° \times 1.9°$	Müller et al. (2018)
MRI-ESM2-0	Meteorological Research Institute, Tsukuba, Japan	$1.1° \times 1.1°$	Yukimoto et al. (2019)

SDMs have been trained in 1979–2005 and applied over reanalysis in 2006–20 for evaluation and also over GCMs in 2015–2100, under the illustrative scenario SSP5-8.5 (see IPCC 2021). A reference period (1981–2010) has been used to standardize predictors and as a base period for future climate change.

The initial pool of potential predictors is listed in Table 2, and after an analysis of their correlations with the target variables, Table 3 shows the list of selected predictors for each region and target variable. For the analog method (see section 3), different fields have been used in each region. For Spain, where synoptic pressure structures strongly condition the large-scale flow and the local weather, geopotential at 500 hPa has been used (see, e.g., Dünkeloh and Jacobeit 2003) over the domain 49°N, 18°W; 29.5°N, 12°E. For Central America, the zonal and meridional wind components at 700 and 250 hPa and the specific humidity at 700 hPa have been used instead, based on Ribalaygua et al. (2018), over the domain 23.5°N, 100.5°W; 1°N, 70.5°W. See the study areas in Fig. 1. These domains have been selected by adding a large-enough area around the study region so the synoptic situations are captured, but for a deeper analysis, the impact on considering different synoptic domains might constitute an interesting study. To analyze the role of humidity, different humidity indices have been combined with the core of predictors listed in Table 3. The suffixes "hur," "hus," and "Dtd" have been added to the SDM names when including relative humidity, specific humidity, and dewpoint depression, respectively.

To characterize the extremely different climates of the two regions, Fig. 2 shows their annual cycle for monthly averaged maximum/minimum temperature and monthly accumulated precipitation. In Spain, temperatures show a marked seasonality, with maximum and minimum temperatures ranging from 10° and 2°C in winter to 30° and 16°C in summer, respectively. On the other hand, in Central America, temperatures are almost constant throughout the year, around 25°–30°C for maximum temperature and 20°C for minimum temperature. Precipitation in Spain presents a marked dry season in summer with a minimum around 20 mm month^{-1} and two peaks in spring and autumn with a maximum of almost 80 mm month^{-1}. For Central America, the precipitation regime is completely different. A strong wet season with up to 300 mm month^{-1} extends from May to October, and a dry season is present for the rest of the year, reaching a minimum of around 50 mm in February and March.

3. Method

In the following sections, the downscaling methods and diagnostics are described.

a. Downscaling methods

Several SDMs belonging to the different families have been used in this study (see Table 3). The election of these specific methods is subjective, and other methods or versions of the different families are of course possible. Nevertheless, the methods included here should cover the main approaches commonly used in statistical downscaling. These methods and/or very similar versions of them have been used in previous evaluation exercises in

TABLE 2. Initial pool of predictors.

Name	Description
psl	Mean sea level pressure
uas; vas	Eastward and northward surface wind components
tas	Surface temperature
ua1000, ua850, and ua500; va1000, va850, and va500	Eastward and northward wind components at 1000, 850, and 500 hPa
ta1000, ta850, and ta500	Temperature at 1000, 850, and 500 hPa
zg1000, zg850, and zg500	Geopotential at 1000, 850, and 500 hPa
hur1000, hur850, and hur500	Relative humidity at 1000, 850, and 500 hPa
hus1000, hus850, and hus500	Specific humidity at 1000, 850, and 500 hPa
Dtd1000, Dtd850, Dtd500	Dewpoint depression at 1000, 850, and 500 hPa
vort1000, vort850, and vort500; div1000, div850, and div500	Vorticity and divergence at 1000, 850, and 500 hPa
K_index; TT_index	Instability indices: K index and total totals index

TABLE 3. Selected predictors for each target variable and region.

	Spain	Central America
tasmax; tasmin pr	ta1000, ta850, and tas (only if indicated as sfc) psl, zg1000, zg850, zg500, K_index, TT_index, and pr (only for XGB)	ta1000, ta850, and tas (only if indicated as sfc) ua1000, ua850, va1000, va850, vort850, div850, K_index, TT_index, and pr (only for XGB)
	hur1000, hur850, and hur700 (only if indicated as hur) hus1000, hus850, and hus700 (only if indicated as hus) Dtd1000, Dtd850, and Dtd700 (only if indicated as Dtd)	hur1000, hur850, and hur700 (only if indicated as hur) hus1000, hus850, and hus700 (only if indicated as hus) Dtd1000, Dtd850, and Dtd700 (only if indicated as Dtd)

Spain (Gutiérrez et al. 2013; San-Martín et al. 2017; Hernanz et al. 2022d) and Europe (Maraun et al. 2015; Gutiérrez et al. 2019), and results here presented are, in general, consistent with the mentioned studies. Because of the particularities of the downscaled variables, different methods are used for temperature and precipitation.

1) RAW

The "RAW" method is not really a downscaling method, but a nearest-neighbor interpolation, included to analyze the added value of the statistical downscaling.

2) MOS

Model output statistics methods correct model biases in different ways. The MOS methods included here are different versions of quantile mapping, in which the adjustment of the simulated distributions is done quantile by quantile. MOS methods use as input the target variable itself, but in coarse resolution, and perform an adjustment on its distribution to correct systematic biases. For each target point, the predictor is taken, interpolating the four nearest grid points using a bilinear interpolation. The major drawback of MOS methods is the assumption of model bias stationarity, although some methods do not rely on this assumption. Two different versions of quantile mapping have been included. The empirical quantile mapping (QM) method, as in Themeßl et al. (2011), compares the observed and simulated distributions in a

historical period to detect systematic biases. Then, it assumes stationarity on the bias under future conditions and applies the detected corrections to the future simulated series. This version is known to affect trends for variables with a marked signal of change. The quantile delta mapping (QDM) method (Cannon et al. 2015), where a delta change is applied for each quantile of the simulated and observed series so trends are preserved in all quantiles and no assumption on the transferability of biases is made, compares the simulated distributions in the future and a historical period, detecting the delta change projected for each quantile. Then, it applies those delta changes to the observed distribution. Both QM and QDM operate at point level (i.e., a different adjustment is done for each target point).

3) ANALOGS

Analog methods (Lorenz 1969; Zorita and von Storch 1999) are based on the assumption that large-scale synoptic patterns condition the local weather, and they search for analog synoptic situations in the past. One major drawback of analog methods is that they cannot predict values outside of the observed sample, which makes them unsuitable for temperature. There is also a huge variety of analog versions. We have limited the use of analogs to precipitation (although for temperature, a hybrid method, which could be categorized as a two-step analog approach, has been included). This method, ANA-SYN-1NN, consists of a simple "best analog" approach, in which the
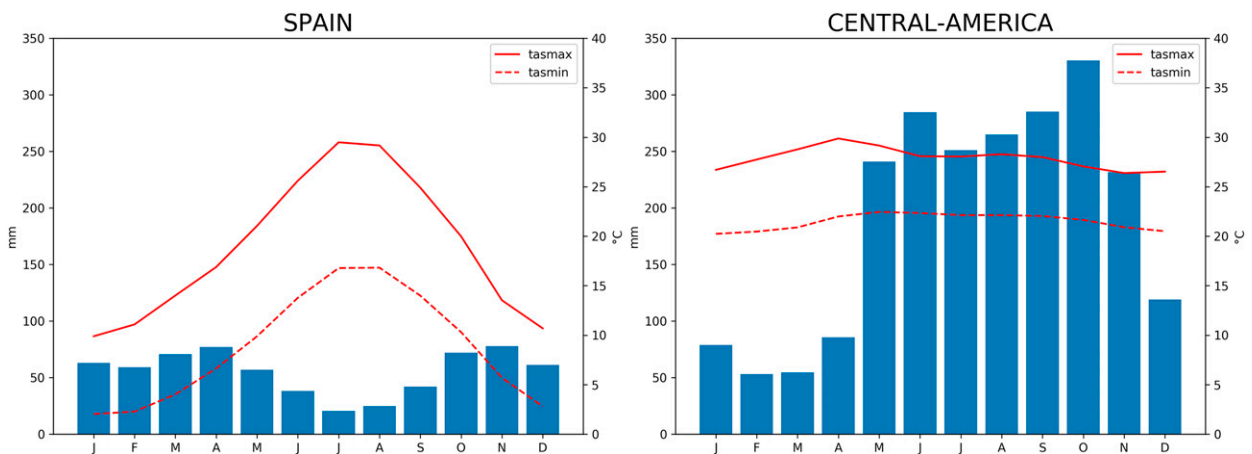

FIG. 2. Monthly observed maximum/minimum temperatures and precipitation climatology (spatially averaged) for (left) Spain and (right) Central America in the reference period (1981–2010).

target variable is taken from the most similar analog day in the historical database. The similarity of the analog days is measured over the synoptic analogy fields (see the spatial domain used in Fig. 1), previously transformed using a principal component analysis (PCA), keeping 95% of their explained variance, and using a Euclidean distance as similarity metric. Thus, for each coarse-resolution simulated day, this method performs the following steps in order: 1) transform the synoptic analogy fields using PCA, 2) search for similar days in the past, and 3) select the most similar day and assign its high-resolution observations to the problem day.

4) TF

Transfer function (TF) methods are based on the existence of statistical relationships between large-scale predictors and local predictands (Sailor and Li 1999; Wilby et al. 2002), which are detected and calibrated in the present and applied over future simulations. These relationships can be tackled using simple linear models or complex nonlinear machine learning approaches. In the particular implementations presented here, predictors are taken, for each target point, from the four nearest grid points using a bilinear interpolation. A different relationship is calibrated for each target point. TF major drawbacks are the assumptions of stationarity of the predictor–predictand relationships and of predictors being well simulated by GCMs.

5) LINEAR METHODS

For linear methods, two classic methods have been included: multiple linear regression (MLR) for temperature and the generalized linear model (GLM-LIN; Wilby et al. 2002) for precipitation. MLR performs a linear regression between all predictors and each predictand. For precipitation, due to its dual nature (precipitation occurrence and intensity), two models are combined. The precipitation occurrence is modeled by a logistic regression using a threshold of 0.1 mm to distinguish between dry and wet days. Then, the precipitation intensity for wet days is obtained by an MLR (forcing to positive values). Additionally, for temperature, a hybrid multiple linear regression based on analog days (MLR-ANA) approach has been included. This method performs the following steps in order: 1) for each simulated day, it transforms the synoptic analogy fields using PCAs; 2) it searches for similar days in the past; and 3) for each target point of the problem day, it calibrates a multiple linear regression using only the analog days. This way, it iterates through all simulated days and through all target points for each simulated day. These methods are limited to capture linear relationships, but predictor–predictand relationships can be nonlinear. For this reason, there is a large variety of machine learning techniques that can be used instead.

6) MACHINE LEARNING

For machine learning methods, we have included artificial neural networks (ANNs) for temperature and extreme gradient boost (XGB) for precipitation. ANNs (McCulloch and Pitts 1943; Rosenblatt 1958) are supervised learning algorithms based on the biological neurons' behavior, imitating them by nodes, which work as perceptrons (Rosenblatt 1958). The ANN used here consists of a multilayer perceptron, where these nodes are organized in several layers, the input layer, the output layer, and a set of hidden layers, which communicate with adjacent layers. Each node or neuron receives a signal from the nodes of the previous layer. The node adds the received signals with different weights, and then, the combined signal is transformed through an activation function. If the result exceeds a certain threshold, the node sends a signal to the next layer. Otherwise, the node sends no signal to the next layer. Different activation functions are possible, the sigmoid and the rectified linear function being some common choices. The training of the network consists in establishing the weights for each pair of neurons, which is achieved by an iterative process in which the signals are propagated forward and errors are sent backward until a certain criterion is reached. During their training, ANN algorithms can get trapped in local minima, which is one of their major drawbacks, along with their high computational training cost. On the other hand, ANNs have proven to be a very powerful tool to tackle complex nonlinear problems. For precipitation, a different machine learning algorithm has been used: XGB (Chen and Guestrin 2016). XGB is an algorithm based on decision trees, in which a set of decision trees is sequentially trained, each using as input data the errors of the previous tree. This method starts with a single tree, which is usually not enough to reach good results. Then, a second tree is added to the sequence, and this second tree is trained with the residuals from the first tree. More trees are added to the ensemble until residuals reach certain criteria. This method is less computationally demanding than ANNs. Additionally, it can combine predictors of different natures (continuous and categorical), and it does not need predictors to be standardized. Thus, this method allows incorporation of precipitation as predictor. On the other hand, one important limitation of this method is that it cannot predict values out of the training range. This method, similar to GLM-LIN, first tackles the wet/dry classification using a threshold of 0.1 mm and then the precipitation intensity on wet days. Different methods have been used for temperature and precipitation. XGB has been chosen for precipitation because of its capability of incorporating predictors of different natures, with no standardization needed, which allows the use of the low-resolution precipitation variable itself. On the other hand, XGB cannot predict values out of the observed range, which has promoted the use of ANNs for temperature instead. Machine learning methods are sensible to the use of different architectures, hyper parameters, etc., and they require a process of tuning. For this study, we have used the default configuration of the open source software pyClim-SDM (Hernanz et al. 2022c; available at https://github.com/ahernanzl/pyClim-SDM/), which has been established after several evaluation exercises with different datasets. Nevertheless, different setups might obviously lead to different results.

7) WEATHER GENERATOR

There are also a huge variety of WGs (see, e.g., Wilks and Wilby 1999). For temperature, we have used a parametric WG based on downscaling monthly statistics and then generating
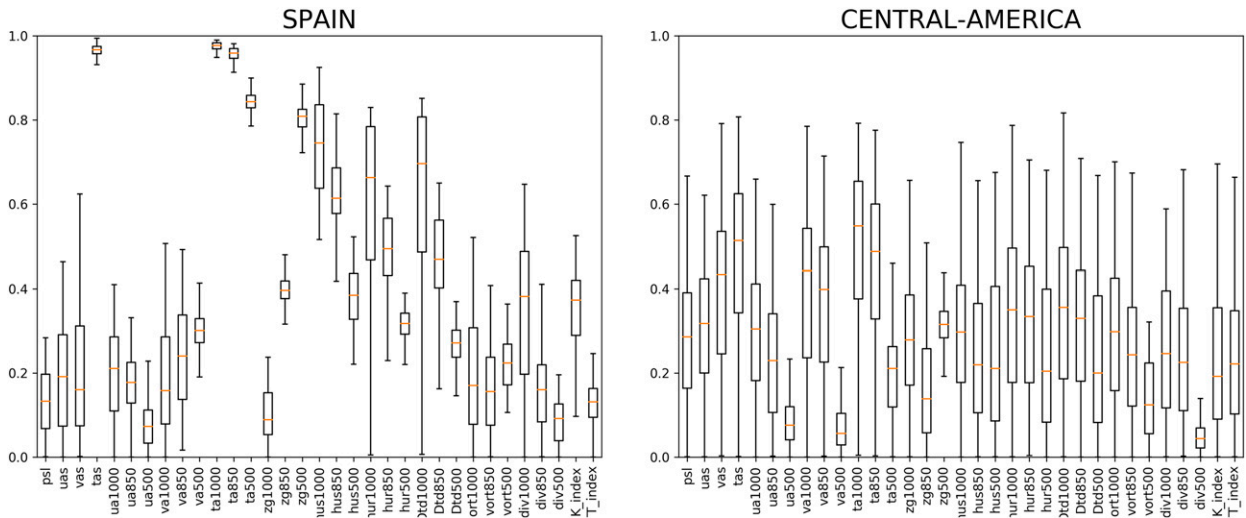
FIG. 3. Spearman's rank correlation coefficients between daily series of the initial pool of predictors from Table 2 and maximum temperature in (left) Spain and (right) Central America, calculated in the training period (1979–2005). Each box contains one value per point.

daily data using a normal distribution (WG-PDF; Erlandsen et al. 2020; Benestad 2021). In this method, first the monthly mean temperatures are calculated both for the observed and simulated series for each target point. Then, a linear regression is established between the monthly values. And finally, daily data are produced following a normal distribution for each month. A different calibration is established for each target point. For precipitation, WG-NMM has been used, which consists of a nonparametric weather generator following a first-order two-state (wet/dry) Markov chain (see Richardson 1981). Both the transition probabilities and the empirical distributions used for the intensity are conditioned on the precipitation given by the reanalysis/models. For this method, the simulated series is first used to classify in intervals of precipitation. For each interval, the observed series is then used to calculate the probabilities of transition between dry and wet days and the empirical distributions for the intensity of precipitation. Then, for each day of the simulated series to be downscaled, the wet/dry state is defined by the state of the previous day and by the probability of transition to the new state, and the intensity of precipitation is generated following the corresponding empirical distribution. A different calibration is established for each target point, using predictors from the four nearest grid points bilinearly interpolated.

b. Diagnostics

This study analyzes three main aspects related to statistical downscaling: 1) the link between different potential predictors and each predictand, 2) the skill of the different statistical methods under the present climate, and 3) their capability to preserve future trends given by GCMs.

First, a search for relevant predictors is performed, using as metric the Spearman's rank correlation coefficient in 1979–2020 between a pool of potential predictors (Table 2) and each of the three target variables. We have included among the potential predictors not only direct model output variables, such as

temperature, wind components, humidity, etc., but also derived predictors, such as vertical stability or low-level convergence, physically linked with predictands like convective precipitation. The correlation is calculated for the daily series of each point. Figures 3 and 4 show the correlation of all points in the form of boxplots. After this analysis, relevant predictors for each target variable and region have been selected to train the downscaling methods (Table 3).

Then, SDMs have been evaluated in a historical period using predictors from a reanalysis, and different metrics have been included. First, SDM accuracy at the daily level has been evaluated. For maximum and minimum temperatures, the root-mean-square error (RMSE) has been calculated between the downscaled and observed daily series for each point. Figure 5 shows those RMSEs for all points in the form of boxplots. Then, in order to reduce the score to a single value so SDMs can be more easily compared, the spatial average of all grid points has been computed (Table 4). For precipitation, due to its non-Gaussian distribution and the large number of zeros, RMSE is not a suitable score. Instead, the Spearman's rank correlation coefficient between the downscaled and the observed daily series of each point has been calculated, as well as the bias in the variance of the daily series. The bias in the variance is calculated in relative terms, as $100 \times$ (var_down–var_obs)/var_obs. These two metrics are shown for all points in the form of boxplots (Fig. 6). To compare SDMs, correlations have been spatially averaged in Table 5. Then, the bias in the mean values over the whole testing period for the three target variables has been calculated. This bias has been computed for each point as mean_down–mean_obs for maximum and minimum temperatures and as $100 \times$ (mean_down–mean_obs)/mean_obs for precipitation. The bias for all points is shown in the form of boxplots in Fig. 7. For precipitation, three other marginal aspects have been evaluated: the wet day frequency (R01), the mean precipitation on wet days (SDII), and the intense precipitations (R95p). A wet day is defined as a day with
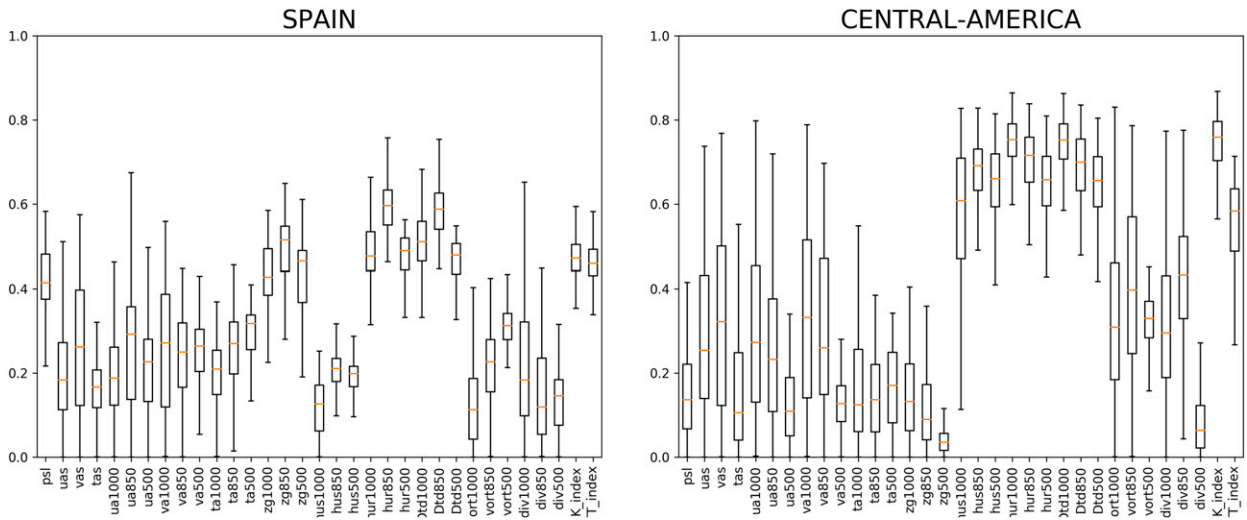
FIG. 4. As in Fig. 3, but for precipitation.

precipitation greater than or equal to 1 mm, and R95p corresponds to the total precipitation on very wet days, with a very wet day defined by the 95th percentile of the wet days in the reference period. For each of these three indices, the bias has been calculated in relative terms (same as for the mean precipitation), and they have been included in the form of boxplots in Fig. 8.

Finally, SDM behavior in the future climate has been analyzed. Although SDMs can reduce the scale of GCMs to finer resolutions, they should preserve trends on spatial scales GCMs operate on. The analysis performed here consists of a comparison of the trends given by raw GCMs with trends given by SDMs, spatially averaged in large regions. For this purpose, SDMs have been applied over the 10 GCMs in Table 1,

both in a reference period and under SSP5-8.5. For each GCM and SDM, the anomaly compared to a reference period has been compared for each point. Then, those anomalies have been spatially averaged for the whole region, and they have been presented in the form of evolution graphs (Figs. 9–12), comparing the multimodel ensemble of raw GCMs with the multimodel ensemble for each individual SDM. Good results in this trend analysis do not guarantee good results in finer scales. Nevertheless, this simple analysis is enough to raise significant modifications in trends from raw GCMs by several downscaling methods. It should be clarified that this analysis does not intend to provide reliable future projections, because no evaluation of GCMs have been performed. The only objective of this analysis is to
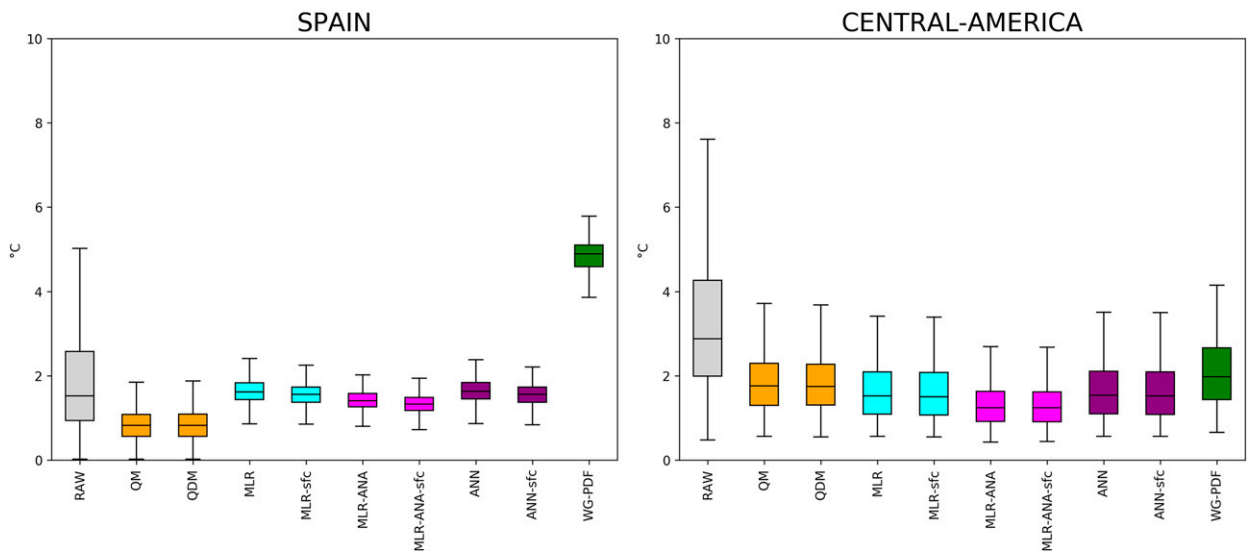

FIG. 5. RMSE of temporal series of daily maximum temperature (°C) in (left) Spain and (right) Central America in the testing period (2006–20). Each box contains one value per point.

TABLE 4. SDM methods used for maximum and minimum temperatures (T column) and precipitation (P column).

Family	Short name	Name	T	P
—	RAW	No downscaling; nearest grid point	x	x
MOS	QM	Quantile mapping	x	x
	QDM	Quantile delta mapping	x	x
Analogs	ANA-SYN-1NN	Analog		x
Linear methods (TF)	MLR	Multiple linear regression	x	
	MLR-ANA	Multiple linear regression based on analog days	x	
	GLM-LIN	Generalized linear model		x
Machine learning (TF)	ANN	Artificial neural networks	x	
	XGB	Extreme gradient boost		x
Weather generator	WG-PDF	Weather generator conditioned on monthly statistics	x	
	WG-NMM	Weather generator based on a nonhomogeneous Markov model		x

ensure that SDMs are able to preserve trends given by GCMs, no matter how realistic or unrealistic those trends are.

4. Results

In the following sections, results for the three main lines of the study (search for relevant predictors, evaluation of SDMs under present conditions, and evaluation of future trends) are presented.

a. Predictor selection

First, the strength of the predictor–predictand relationships has been evaluated using the Spearman's rank correlation coefficient. This has been done for the initial pool of predictors from Table 2 with each of the target variables. For maximum temperature (Fig. 3), significant differences are appreciated between Spain and Central America. While in Spain there is a strong (and expected) correlation with temperature both on surface and in low-pressure levels (1000 and 850 hPa), in Central America correlations are much lower. As has been mentioned in section 1, the link between large-scale structures and local weather is weaker in the tropics, so lower correlations were expected. Additionally, these poor correlations can be partially explained because of the low variance for temperature in Central America (Fig. 2). For minimum temperature, results are similar (Fig. S1 in the online supplemental material). For precipitation, Fig. 4 reveals important differences between the two study regions. While in Spain the most relevant predictors are those related to pressure structures [mean sea level pressure (psl); geopotential height (zg) at different levels], humidity (relative humidity and dewpoint depression), and instability (K index and total totals index), in Central America pressure is not relevant, and most of the predictive power relies on humidity (now also including the specific humidity) and instability indices. This was also expected, as it is known that large-scale pressure structures condition the generation of extratropical cyclones and their associated fronts, which can explain a large proportion of the precipitation in Spain. And humidity and instability were also expected to be relevant predictors in Spain due to their role in convective precipitation. The correlation with instability indices in Spain is higher during JJA (not shown), when precipitation is mainly of a convective nature. On the other hand, results in Central America

were also expected, because it is known that precipitation in that region is mainly convective, so the role of humidity and instability is very important, and vorticity and divergence in the low levels also gain relevance. One interesting finding is that, comparing relevant predictors for precipitation in both regions, correlations are higher in Central America (opposite of temperature). This is a good indicator for the potential success in statistical downscaling of precipitation in Central America.

After this initial exploration of the potential predictors (Table 2), a selection for each region has been made (Table 3). For maximum and minimum temperatures, temperatures at 1000 and 850 hPa have been used. Additionally, because surface temperature is influenced by atmosphere–ground interactions and exchanges of heat in different forms, which might cause differences to emerge among GCMs due to different physics schemes, we have decided to analyze its inclusion as a predictor separately. Thus, the suffix "sfc" in the downscaling method names indicates the use of surface temperature as a predictor (in addition to the mentioned temperature in low levels). For precipitation, different sets of predictors have been used for each region. In Spain, the predictors used are the mean sea level pressure; geopotential at 1000, 850, and 500 hPa; and instability indices (K and total totals). In Central America, the predictors used are the wind components at 1000, 850, and 500 hPa; vorticity and divergence at 850 hPa; and the same instability indices. One specific method (XGB) incorporates precipitation as a predictor as well, a highly non-Gaussian variable problematic for regressions in general, but not for this particular method. Two main differences exist between the two TF methods used for precipitation: 1) GLM is a linear model, but XGB can simulate nonlinear relationships, and 2) XGB has been fed with an additional predictor, the precipitation itself. Thus, differences between these two SDMs (shown in sections 4b and 4c) cannot be attributed exclusively to the methods themselves, but also to the use of different information. On the other hand, while having deprived XGB from the precipitation input would have made a fairer comparison with the GLM, the specific nature of XGB allows incorporation of that information, which is an important strength of this method that must be highlighted. And humidity indexes have been explored separately in order to analyze both their added value and their impact on future trends because of known issues in the transferability of the statistical methods
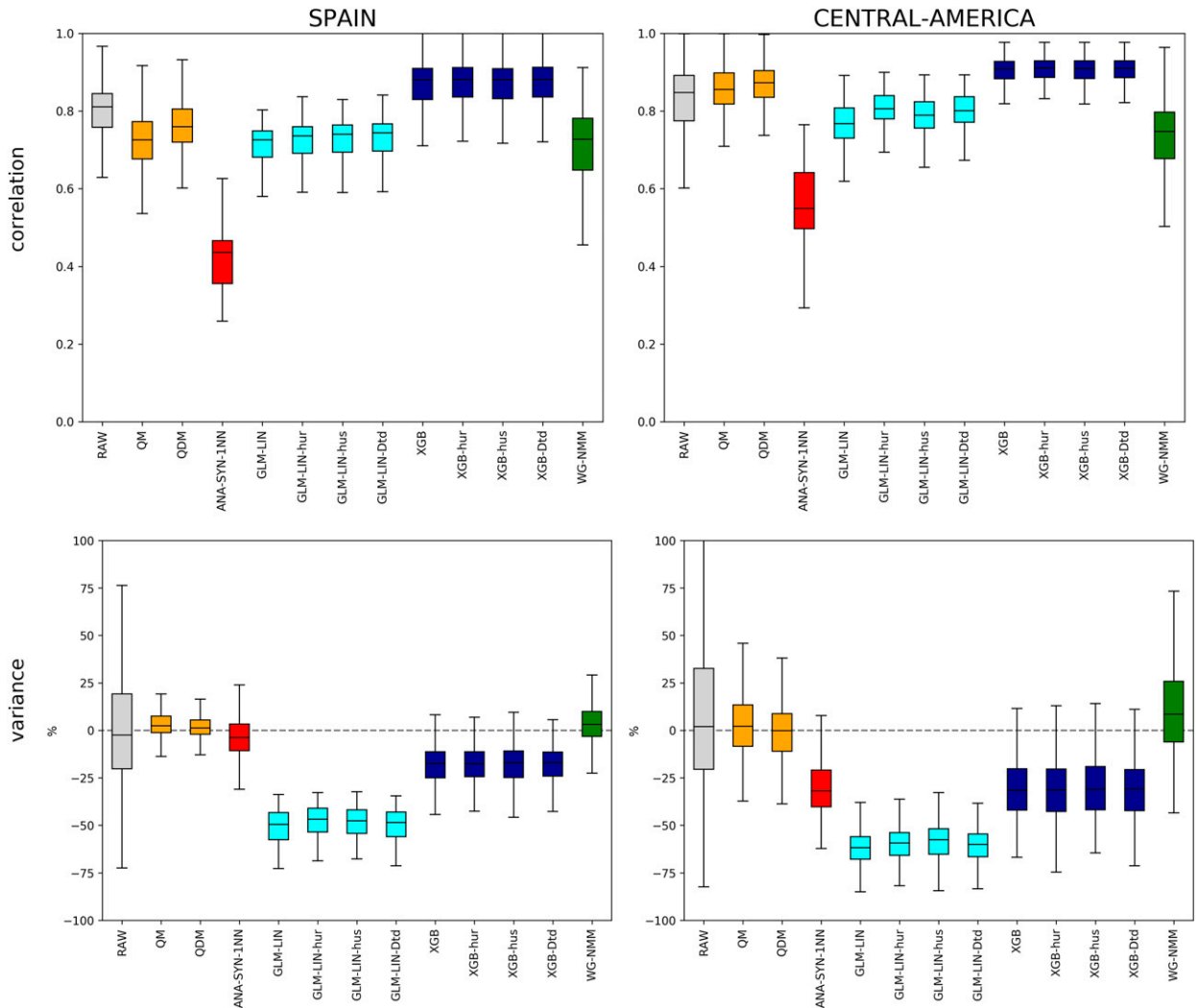
FIG. 6. (top) Spearman's rank correlation coefficient between the downscaled and observed series and (bottom) relative bias (%) in the variance of the downscaled series for daily precipitation in (left) Spain and (right) Central America in the testing period (2006–20). Each box contains one value per point.

due to the use of one or another humidity index (Fu et al. 2018). Thus, the suffixes hur, hus, and Dtd in the downscaling method names indicate the use (in addition to the mentioned predictors for each region) of relative humidity, specific humidity, and dewpoint depression, respectively, at 1000, 850, and 700 hPa.

b. Evaluation of SDMs under present conditions

Once SDMs have been trained using these predictors, they have been evaluated in a historical period over a reanalysis. For maximum temperature, the sizes of the boxes at Fig. 5 reveal how the performance of each method is more sensitive to each specific location in Central America than in Spain. While errors in Spain, by each method, are similar for all points, in Central America errors extend to larger ranges. In Spain, WG-PDF presents larger errors than in Central America. This is expected due to its stochastic nature (it produces daily

data based on monthly statistics, so bad scores at daily level are expected) and the low variability for temperature in Central America (which explains why errors in Central America are lower than in Spain). For minimum temperature, results are similar (Fig. S2 in the online supplemental material). Then, these RMSEs have been spatially averaged and are presented in Table 5. In Spain, methods with lower RMSEs are MOS methods, with errors of 0.8°C both for maximum and minimum temperatures. Transfer function methods present larger errors, similar for MLR and ANN (and slightly better for MLR-ANA and with the inclusion of surface temperature as a predictor). For Central America, errors by MOS methods are larger than in Spain, especially for maximum temperature (1.8°–1.9°C). And the lowest errors are achieved by MLR-ANA (1.3°C for maximum temperature and 0.7°C for minimum temperature), with no difference when including surface temperature as a predictor.

TABLE 5. Spatially averaged RMSE (°C) for maximum and minimum daily temperatures by different methods (rows) and in the two regions (columns).

	Spain		Central America	
	tasmax	tasmin	tasmax	tasmin
RAW	1.8	1.8	3.5	4.0
QM	0.8	0.8	1.9	1.0
QDM	0.8	0.8	1.8	1.0
MLR	1.6	1.9	1.6	0.9
MLR-sfc	1.6	1.5	1.6	0.9
MLR-ANA	1.4	1.6	1.3	0.7
MLR-ANA-sfc	1.3	1.3	1.3	0.7
ANN	1.7	1.9	1.7	1.0
ANN-sfc	1.6	1.5	1.7	0.9
WG-PDF	4.8	3.6	2.1	1.3

For precipitation, the correlation and the variance of the daily series have been analyzed (Fig. 6). In both regions, XGB gets the highest correlations, followed by some MOS methods. On the other hand, XGB shows an underestimation of the variance, while MOS methods reproduce it better. Nevertheless, the underestimation of the variance is a limitation common to TF methods that can be alleviated through a posterior bias correction. Table 6 shows the spatially averaged correlations for each method and region. The highest correlations are achieved by XGB in both regions, and the use of humidity predictors improves correlations only moderately

(from 0.868 to 0.871–0.873 in Spain and from 0.906 to 0.906–0.908 in Central America). On the other hand, for the GLM-LIN, the use of humidity predictors improves correlations more clearly (from 0.710 to 0.728–0.732 in Spain and from 0.766 to 0.786–0.805 in Central America). This can be explained due to the use of precipitation as a predictor by XGB, so it is less sensitive to other predictors.

Then, we have analyzed the bias in the mean values of the three target variables. Figure 7 shows how all methods clearly improve the representation of the mean values, both for temperature and precipitation, by the low-resolution reanalysis. For temperature, all methods achieve very similar biases, close to 0°C, and for precipitation, the analog method and the GLM present higher biases than the other methods (and lower when the relative humidity or the dewpoint depression are included). For precipitation, three more marginal aspects have been evaluated: R01, SDII, and R95 (Fig. 8). The number of wet days and the mean precipitation on them are well simulated by the analog method and XGB. On the other hand, extreme precipitation is best captured by MOS and WG methods. The underestimation of the extreme precipitation by TF methods is a well-known issue due to their underestimation of the variance. On the other hand, results for R01 and SDII do not align with Gutiérrez et al. (2019), where a large variety of statistical methods were intercompared over several locations along Europe. Thus, no general conclusion should be made about these indices, as the performance of SDMs appears to be sensitive to the datasets used.
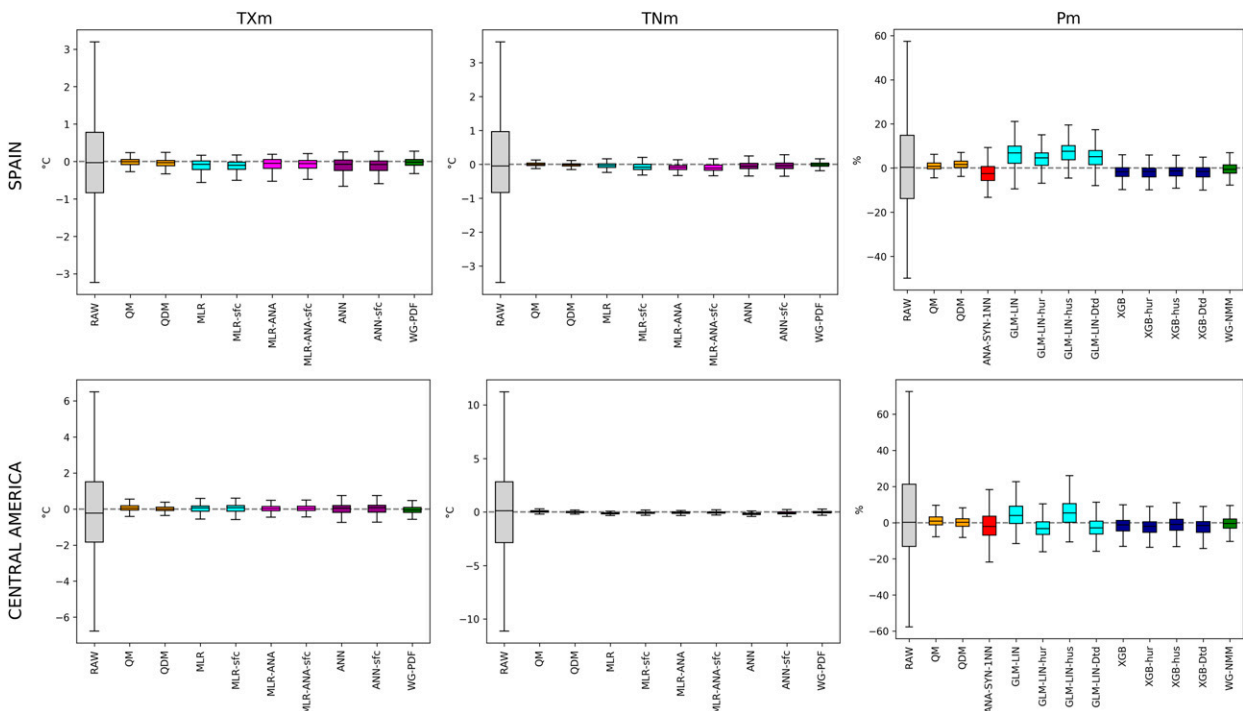

FIG. 7. Bias [absolute for temperature (°C) and relative for precipitation (%)] for the mean values (averaged over the whole testing period, 2006–20) of (left) maximum temperature, (center) minimum temperature, and (right) precipitation in (top) Spain and (bottom) Central America. Each box contains one value per point.
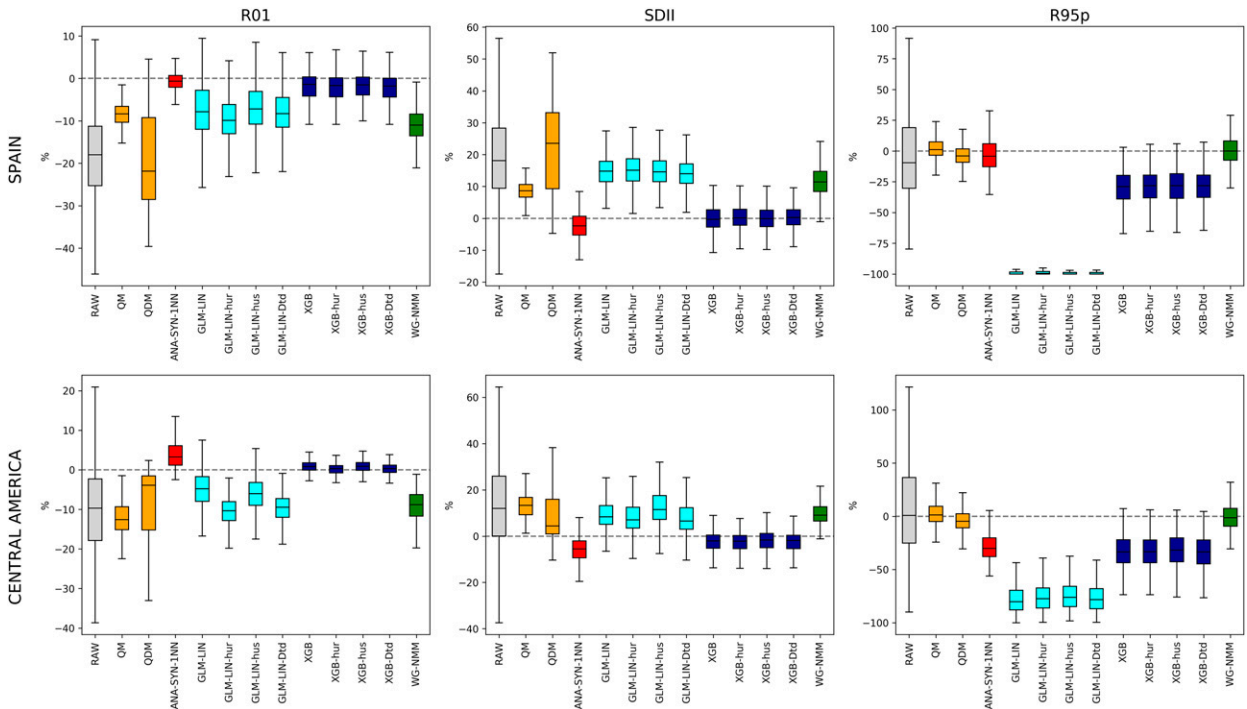
FIG. 8. Relative bias (%) for the mean values (averaged over the whole testing period, 2006–20) of (left) the number of wet days, (center) the mean precipitation in wet days, and (right) the total precipitation on very wet days in (top) Spain and (bottom) Central America. Each box contains one value per point.

c. Evaluation of future trends

Finally, the impact that statistical downscaling can have on long-term trends has been analyzed. Figure 9 shows trends given by raw GCMs and by each SDM for maximum temperature under SSP5-8.5 in Spain. All SDMs preserve trends by GCMs, and the use of surface temperature as a predictor for TF methods does not appear to affect the trend (while it improved results in the historical evaluation). On the other hand, in Central America (Fig. 10), most methods modify trends significantly. Among the MOS methods, the simple empirical quantile mapping shows important deviations from the desired trends, while for QDM, trends are preserved very accurately. Transfer function methods, despite their good results in the historical evaluation, present trends are extremely deviated from the ones given by raw GCMs. The underestimation of the signal of change displayed by the linear methods (MLR and MLR-ANA, with and without sfc) can be explained by the low variability in the training dataset. Linear relationships are calibrated in a sample with low variance and cannot be extrapolated to the high values expected under the future climate. The same problem affects ANN (and ANN-sfc). These methods are calibrated with a training dataset, and then, they are applied to predictors with values out of the calibration range, and in that case, nonlinear methods based on machine learning algorithms can behave extremely wrong (Hernanz et al. 2022b). And finally, the WG-PDF slightly deviates from the desired trend, probably because of the same explanation. For minimum temperature, results are similar (Figs. S3 and S4 in the online supplemental material).

For precipitation, Fig. 11 shows trends in Spain by each SDM. MOS, XGB, and WG methods preserve trends quite accurately (although with slight differences among them). On the other hand, the analog method deviates significantly from the desired trend, and the GLM-LIN is very sensitive to the use of one or another humidity predictor. While its trend without humidity and with relative humidity or dewpoint depression is close to the one given by raw GCMs (although with more spread, i.e., uncertainty), when the specific humidity is included its behavior in the future is extremely wrong. This can be explained by the fact that in a future warmer atmosphere, more water vapor can be stored without reaching saturation, so the statistical relationship between specific humidity and precipitation is expected to change. Additionally, the specific humidity is a variable with a strong signal of change, and its projected values often lie out of the observed range, which, as has been mentioned, can be a problem for statistical methods. The reason why XGB is less sensitive to this specific humidity issue is that XGB uses precipitation as a predictor, so it relies less on the other predictors. In Central America (Fig. 12), more marked differences emerge among the MOS methods, with QDM being the one that better preserves trends. The analog method, as occurred in Spain, deviates from the desired trend, in this case even changing the sign of the future change. The GLM-LIN is again sensitive to the use of one or another humidity predictor. In this case, if no humidity index is included, the method does not project any change, while raw GCMs project a decrease of the precipitation by the end of the century. When using the relative
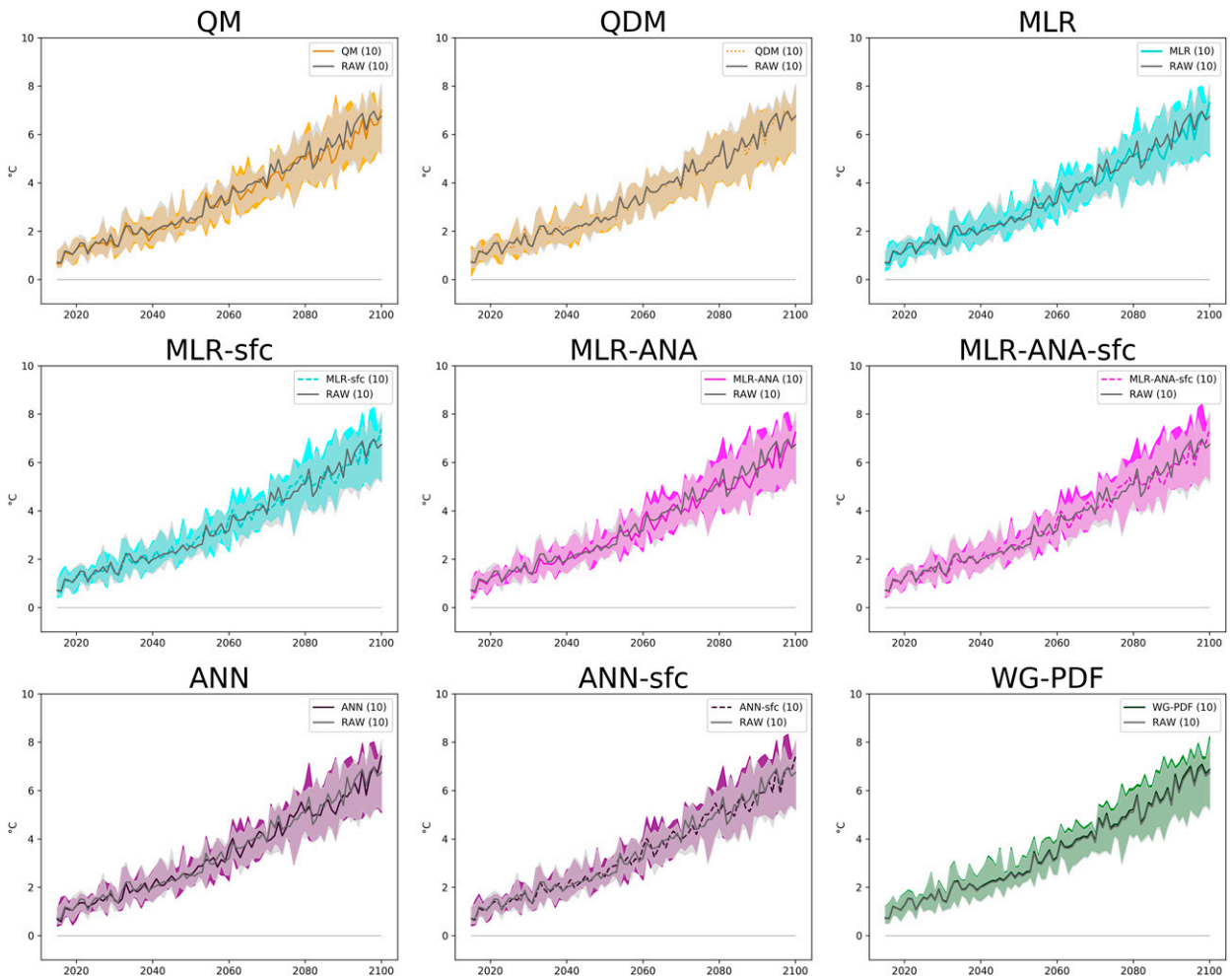
FIG. 9. Change in maximum temperature (°C) by raw GCMs [RAW (gray)] and each SDM applied over them, under SSP5-8.5, compared to the mean values in the reference period (1981–2010), spatially averaged in Spain and using annual means. The shaded area expands over the interquartile range around the median of the multimodel ensemble.

humidity, the signal of change is intensified, and with the specific humidity, a very nonrealistic behavior pointing to a large increase in the precipitation is seen. Finally, with the use of Dtd, the trend given by raw GCMs is better preserved. Thus, although the use of any of the three proposed humidity predictors improved results in the historical evaluation, the use of one or another can seriously affect trends, which aligns with Fu et al. (2018). This issue is less marked for XGB due to the important role of precipitation as a predictor in this method, but while in Spain the use of specific humidity did not set XGB apart from raw GCMs, in Central America it does. This is probably because of a more important role of the specific humidity in the explanation of tropical precipitation than in the extratropics. And WG, as well as in Spain, preserves trends given by raw GCMs perfectly.

5. Conclusions

Statistical downscaling has been extensively evaluated in extratropical regions, but evaluation studies intercomparing the main families of statistical downscaling methods are nonexistent in the tropics. In this study, a comparison of several state-of-the-art SDMs belonging to different families has been performed over two completely different climatic regions, Spain (midlatitudes) and Central America (tropical region). The following conclusions about the behavior of SDMs in both regions have been reached.

- Relevant predictors differ in both regions. Whereas in Spain the role of large-scale pressure structures is key, in Central America the humidity and instability appear to be more important.
- Statistical downscaling improves results given by low-resolution data.
- Statistical downscaling of temperature is more skillful in Spain than in Central America.
- Transfer function methods for temperature do not extrapolate well to future warmer conditions in Central America, probably because of the low variance in the observed climate.
- The use of humidity as a predictor improves results in a historical evaluation but can affect future trends. The sensitivity
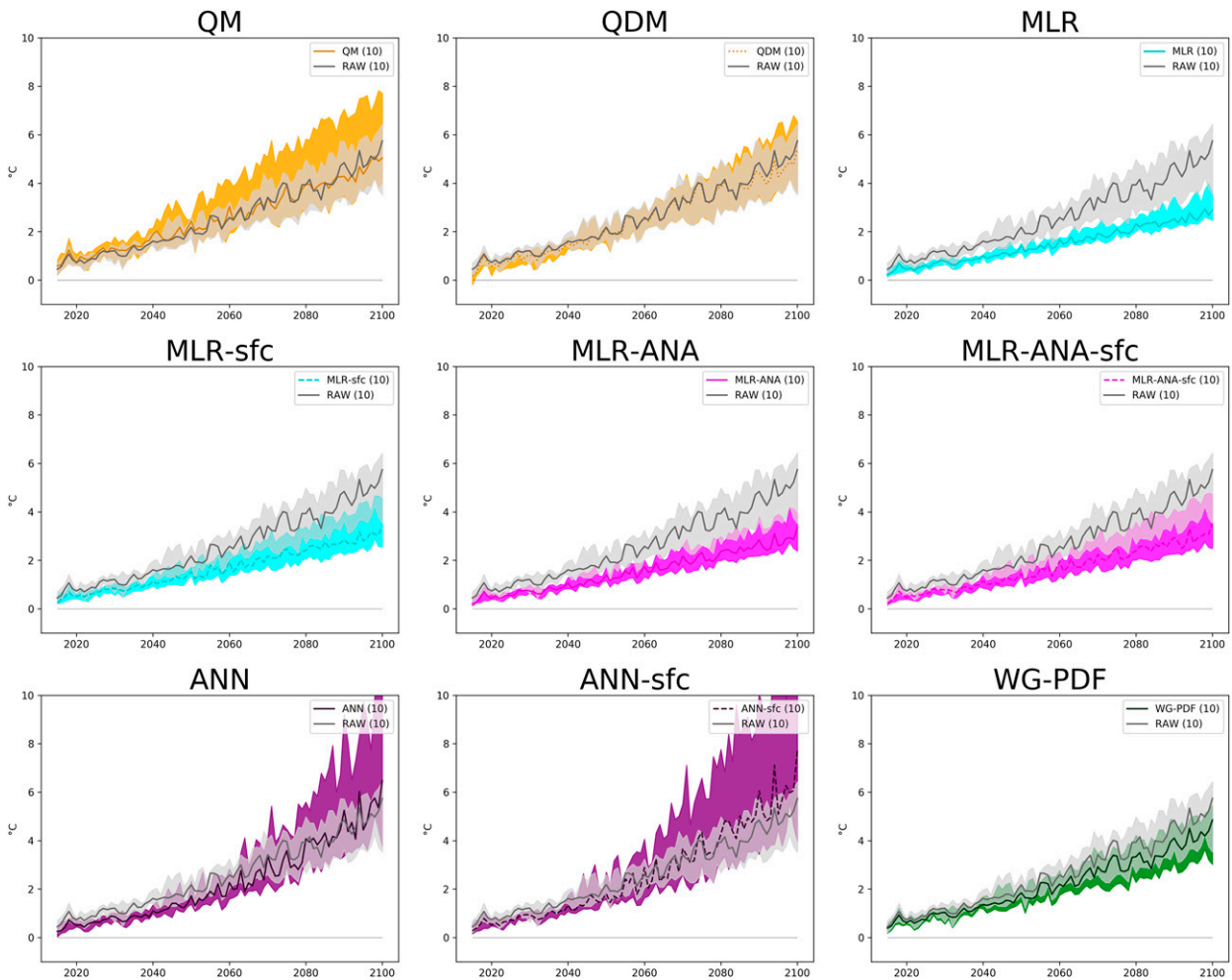
FIG. 10. As in Fig. 9, but in Central America.

for different methods and for each region is different, with specific humidity (compared to the relative humidity and the dewpoint depression) appearing to be the most problematic variable. It can be argued that GCMs might be missing the effects of an increase in specific humidity in precipitation due to their coarse resolution and/or their generally poor representation of convection. Convective rainfalls are important both in Central America and in Spain (especially in the Mediterranean region). Thus, when comparing trends given by raw GCMs with GLM-hus, it could be argued that maybe GLM-hus is the one projecting the correct signal of change. In our opinion, while it is theoretically possible that GCMs are misrepresenting convective situations that lead to important amounts of precipitation, and the GLM-hus might be improving the signal of change by raw GCMs, we find it difficult to trust that signal for various reasons. 1) The transferability of the statistical relationships between specific humidity and precipitation is problematic due to the capability of a warmer atmosphere to store more water vapor without reaching saturation. 2) Statistical relationships can present problems extrapolating outside their calibration range, especially for

nonlinear relationships. And specific humidity is frequently projected outside of the observed range in the future. 3) Similar conclusions have been found using other statistical methods (Fu et al. 2018). 4) The signal of change by GLM-hus is extremely different to the one given by raw GCMs, even of a different sign. On the other hand, it is true that GCMs might be misrepresenting convection and missing part of the precipitation. In this sense, trends given by raw GCMs might be wrong to some degree. It is difficult to analyze this possibility, however, because the use of downscaling methods, even convection-permitting regional climate models, is limited. Since there is no feedback from the downscaling methods to the GCMs, when a precipitation event missed by a GCM is well captured by a downscaling method; the amount of atmospheric humidity that should be subtracted from the GCMs will still be present and available to produce following precipitation events. Thus, it is likely that the downscaled simulation will produce an artifact overestimation of the precipitation amounts. In our opinion, this limitation could only be solved by the use of high-resolution convection-permitting GCMs. In the meantime, the dewpoint depression, an index implicitly accounting
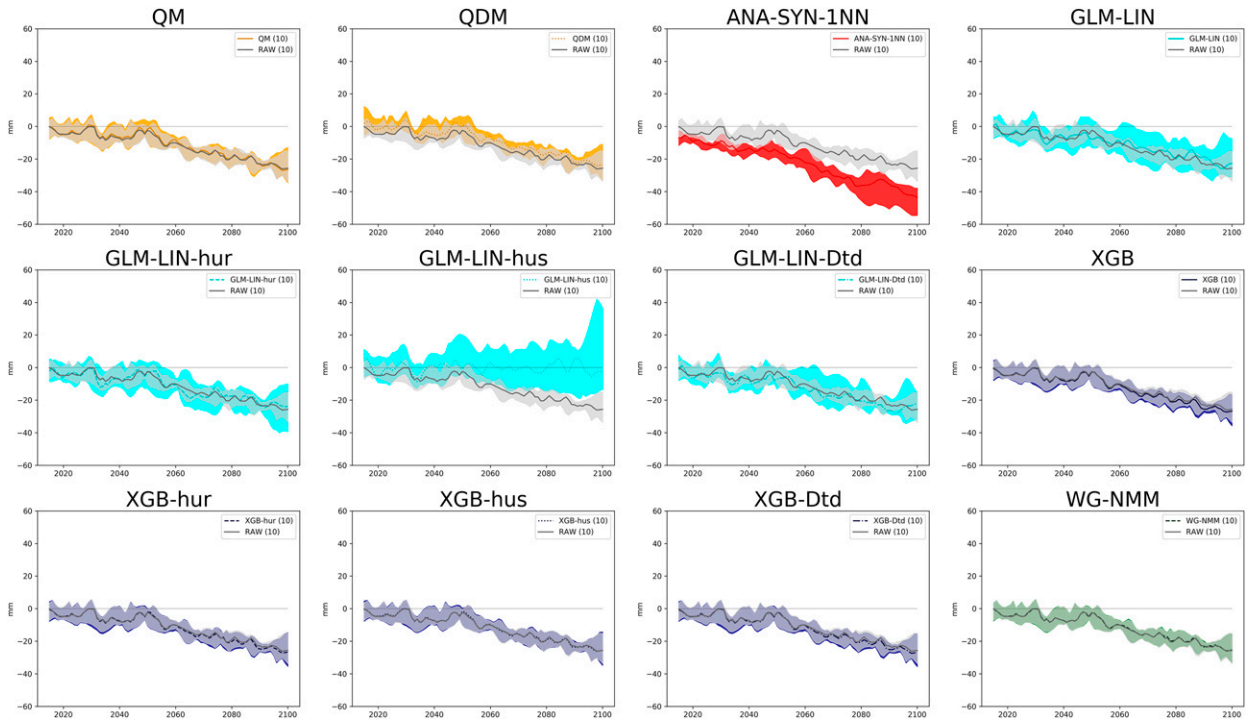
FIG. 11. As in Fig. 9, but for precipitation and as relative change (%).

for both the saturation and the amount of water vapor, appears to be a more reliable predictor.

• These conclusions have been reached for these specific SDMs and setups but might vary for other SDMs, sets of predictors, or regions. Regional variables of interest in the tropics, such as, for example, the Galvez–Davison index, have been neglected here for a fair and basic comparison with the extratropics. Nonetheless, their use is expected to
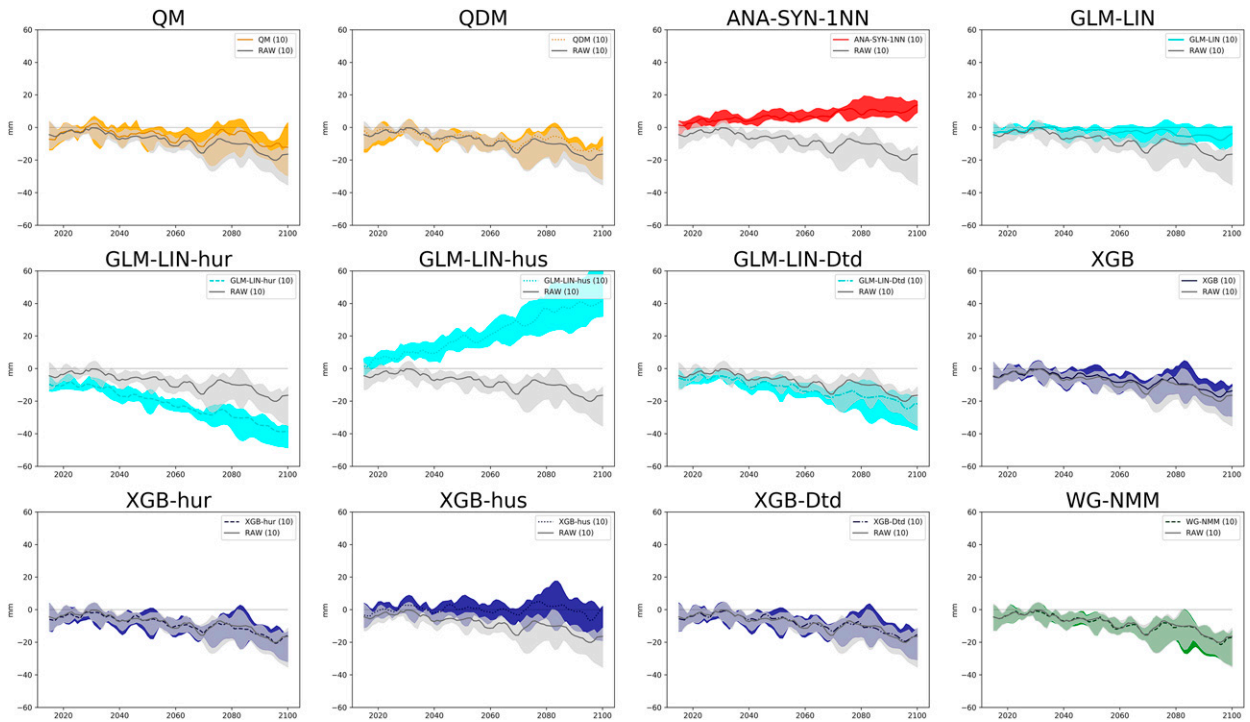
FIG. 12. As in Fig. 9, but for precipitation and as relative change (%) and in Central America.

TABLE 6. As in Table 5, but for Spearman's rank correlation coefficient and precipitation.

	Spain	Central America
RAW	0.798	0.826
QM	0.730	0.852
QDM	0.769	0.865
ANA-SYN-1NN	0.422	0.561
GLM-LIN	0.710	0.766
GLM-LIN-hur	0.724	0.805
GLM-LIN-hus	0.728	0.786
GLM-LIN-Dtd	0.732	0.799
XGB	0.868	0.906
XGB-hur	0.873	0.908
XGB-hus	0.871	0.906
XGB-Dtd	0.873	0.908

improve model performance, especially for the nonlinear methods, in the tropics.

In summary, SD has been found to behave significantly differently in Spain and Central America, although the benefits of SD are clear in both regions. Some tropical regions are especially vulnerable to climate change, so reliable regional information is needed for adaptation and impact studies; further analysis of the suitability of different SDMs and their specific setups for those regions would be desirable. Furthermore, conclusions reached here might vary for other tropical and extratropical regions, so the inclusion of different tropical regions in comprehensive studies such as the one performed in the EU COST Action VALUE targeting a wide variety of extratropical locations might be needed.

REFERENCES

Benestad, R. E., 2021: A Norwegian approach to downscaling. *Geosci. Model Dev. Discuss.*, https://doi.org/10.5194/gmd-2021-176.

Bi, D., and Coauthors, 2020: Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. *J. South. Hemisphere Earth Syst. Sci.*, **70**, 225–251, https://doi.org/10.1071/ES19040.

Boucher, O., and Coauthors, 2020: Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002010, https://doi.org/10.1029/2019MS002010.

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1.

Cavazos, T., and B. C. Hewitson, 2005: Performance of NCEP–NCAR reanalysis variables in statistical downscaling of daily precipitation. *Climate Res.*, **28**, 95–107, https://doi.org/10.3354/cr028095.

Charles, S. P., B. C. Bates, I. N. Smith, and J. P. Hughes, 2004: Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrol. Processes*, **18**, 1373–1394, https://doi.org/10.1002/hyp.1418.

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, Association for Computing Machinery, 785–794, https://dl.acm.org/doi/10.1145/2939672.2939785.

Comisión Europea, 2021: Escenarios de cambio climático regionalizados para la planificación de medidas de adaptación. Conceptos básicos, herramientas de visualización y buenas prácticas. Serie de Estudios Temáticos EUROCLIMA+19, 76 pp., https://adaptecca.es/sites/default/files/documentos/mnal21006esn.pdf.

Döscher, R., and Coauthors, 2022: The EC-EARTH3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci. Model Dev.*, **15**, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022.

Dünkeloh, A., and J. Jacobeit, 2003: Circulation dynamics of Mediterranean precipitation variability 1948–98. *Int. J. Climatol.*, **23**, 1843–1866, https://doi.org/10.1002/joc.973.

Erlandsen, H. B., K. M. Parding, R. Benestad, A. Mezghani, and M. Pontoppidan, 2020: A hybrid downscaling approach for future temperature and precipitation change. *J. Appl. Meteor. Climatol.*, **59**, 1793–1807, https://doi.org/10.1175/JAMC-D-20-0013.1.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Fu, G., S. P. Charles, F. H. S. Chiew, M. Ekström, and N. J. Potter, 2018: Uncertainties of statistical downscaling from predictor selection: Equifinality and transferability. *Atmos. Res.*, **203**, 130–140, https://doi.org/10.1016/j.atmosres.2017.12.008.

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

——, and Coauthors, 2019: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol.*, **39**, 3750–3785, https://doi.org/10.1002/joc.5462.

Hernanz, A., J. A. García-Valero, M. Domínguez, and E. Rodríguez-Camino, 2022a: Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with imperfect predictors (global climate model experiment). *Int. J. Climatol.*, **42**, 6793–6806–, https://doi.org/10.1002/joc.7611.

——, ——, ——, and ——, 2022b: A critical view on the suitability of machine learning techniques to downscale climate change projections: Illustration for temperature with a toy experiment. *Atmos. Sci. Lett.*, **23**, e1087, https://doi.org/10.1002/asl.1087.

——, C. Correa, J. A. García-Valero, M. Domínguez, E. Rodríguez-Guisado, and E. Rodríguez-Camino, 2022c: Service for generation of statistical downscaled climate change projections supporting national adaptation strategies. SSRN, 4247430, 23 pp., https://doi.org/10.2139/ssrn.4247430.

——, J. A. García-Valero, M. Domínguez, P. Ramos-Calzado, M. A. Pastor-Saavedra, and E. Rodríguez-Camino, 2022d: Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors. *Int. J. Climatol.*, **42**, 762–776, https://doi.org/10.1002/joc.7271.

——, ——, ——, and E. Rodríguez-Camino, 2022e: Evaluation of statistical downscaling methods for climate change projections over Spain: Future conditions with pseudo reality (transferability experiment). *Int. J. Climatol.*, **42**, 3987–4000, https://doi.org/10.1002/joc.7464.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

IPCC, 2021: Summary for policymakers. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 3–32.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

Manzanas, R., S. Brands, D. San-Martín, A. Lucero, C. Limbo, and J. M. Gutiérrez, 2015: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *J. Climate*, **28**, 4171–4184, https://doi.org/10.1175/JCLI-D-14-00331.1.

Maraun, D., and M. Widmann, 2018: *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 347 pp.

——, and Coauthors, 2015: Value: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3**, 1–14, https://doi.org/10.1002/2014EF000259.

——, M. Widmann, and J. M. Gutiérrez, 2019: Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *Int. J. Climatol.*, **39**, 3692–3703, https://doi.org/10.1002/joc.5877.

Mauritsen, T., and Coauthors, 2019: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *J. Adv. Model. Earth Syst.*, **11**, 998–1038, https://doi.org/10.1029/2018MS001400.

McCulloch, W. S., and W. Pitts, 1943: A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, **5**, 115–133, https://doi.org/10.1007/BF02478259.

Müller, W. A., and Coauthors, 2018: A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *J. Adv. Model. Earth Syst.*, **10**, 1383–1413, https://doi.org/10.1029/2017MS001217.

Ramseyer, C. A., and T. L. Mote, 2016: Atmospheric controls on Puerto Rico precipitation using artificial neural networks. *Climate Dyn.*, **47**, 2515–2526, https://doi.org/10.1007/s00382-016-2980-3.

Ribalaygua, J., E. Gaitán, J. Pórtoles, and R. Monjo, 2018: Climatic change on the Gulf of Fonseca (Central America) using two-step statistical downscaling of CMIP5 model outputs. *Theor. Appl. Climatol.*, **132**, 867–883, https://doi.org/10.1007/s00704-017-2130-9.

Richardson, C. W., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.*, **17**, 182–190, https://doi.org/10.1029/WR017i001p00182.

Rosenblatt, F., 1958: The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, **65**, 386–408, https://doi.org/10.1037/h0042519.

Rummukainen, M., 2010: State-of-the-art with regional climate models. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 82–96, https://doi.org/10.1002/wcc.8.

Sailor, D. J., and X. Li, 1999: A semiempirical downscaling approach for predicting regional temperature impacts associated with climatic change. *J. Climate*, **12**, 103–114, https://doi.org/10.1175/1520-0442(1999)012<0103:ASDAFP>2.0.CO;2.

San-Martín, D., R. Manzanas, S. Brands, S. Herrera, and J. M. Gutiérrez, 2017: Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *J. Climate*, **30**, 203–223, https://doi.org/10.1175/JCLI-D-16-0366.1.

Schoof, J. T., 2013: Statistical downscaling in climatology. *Geogr. Compass*, **7**, 249–265, https://doi.org/10.1111/gec3.12036.

Swart, N. C., and Coauthors, 2019: The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev.*, **12**, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019.

Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019.

Themeßl, M. J., A. Gobiet, and A. Leuprecht, 2011: Empirical–statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, **31**, 1530–1544, https://doi.org/10.1002/joc.2168.

Trzaska, S., and E. Schnarr, 2014: A review of downscaling methods for climate change projections: African and Latin American

Resilience to Climate Change (ARCC) project. USAID, 56 pp., http://www.ciesin.org/documents/Downscaling_CLEARED_000.pdf.

Volodin, E. M., N. A. Diansky, and A. V. Gusev, 2013: Simulation and prediction of climate changes in the 19th to 21st centuries with the Institute of Numerical Mathematics, Russian Academy of Sciences, model of the Earth's climate system. *Izv. Atmos. Oceanic. Phys.*, **49**, 347–366, https://doi.org/10.1134/S0001433813040105.

——, and Coauthors, 2017: Simulation of modern climate with the new version of the INM RAS climate model. *Izv. Atmos. Oceanic Phys.*, **53**, 142–155, https://doi.org/10.1134/S0001433817020128.

Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr. Earth Environ.*, **21**, 530–548, https://doi.org/10.1177/030913339702100403.

——, C. W. Dawson, and E. M. Barrow, 2002: SDSM—A decision support tool for the assessment of regional climate change impacts. *Environ. Modell. Software*, **17**, 145–157, https://doi.org/10.1016/S1364-8152(01)00060-3.

——, S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. Zenodo, https://doi.org/10.5281/zenodo.1438320.

Wilks, D. S., and R. L. Wilby, 1999: The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr. Earth Environ.*, **23**, 329–357, https://doi.org/10.1177/030913339902300302.

Yukimoto, S., and Coauthors, 2019: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *J. Meteor. Soc. Japan*, **97**, 931–965, https://doi.org/10.2151/jmsj.2019-051.

Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.