Climate change projections and extremes for Costa Rica using tailored predictors from CORDEX model output through statistical downscaling with artificial neural networks

Dánnell Quesada-Chacón | Klemens Barfus | Christian Bernhofer

Institute of Hydrology and Meteorology, Technische Universität Dresden, Dresden, Germany

Correspondence
Dánnell Quesada-Chacón, Institute of Hydrology and Meteorology, Technische Universität Dresden, Dresden, Germany. Email: dannell.quesada@tu-dresden.de

Abstract
Despite intense research on climate change (CC), regional studies for Central America, which is considered a CC hot spot, remain scarce. The information provided by general circulation models (GCMs) is too coarse to accurately reproduce local-scale climatic features, which are needed for impact assessment. Thus, downscaling techniques are employed to address this scale mismatch. Costa Rica is the present case study, for which suitable predictors were tailored for downscaling related to regional climatic characteristics, such as the Inter-Tropical Convergence Zone, El Niño Southern Oscillation, the Caribbean Low-Level Jet, and the Mid-Summer Drought. Statistical downscaling models were calibrated for precipitation, maximum and minimum temperature, using the perfect prognosis methodology by means of station data, ERA-INTERIM reanalysis and artificial neural networks, yielding satisfactory results. As found in several studies, the temperature models replicated more accurately the statistics of the observed datasets. However, here, through the implemented approach and the tailored predictors, the precipitation models conveyed an improvement compared to standard methods. Projected daily climate was obtained employing CORDEX data under the RCP8.5 scenario for the central region of the country. Overall, the changes in climate estimated by the end of the 21st century agree with coarser-scale projections. Finally, projected climate extremes indices were calculated and rendered further details on the intensity of future CC by the end of the century.

KEYWORDS
artificial neural networks, climate change, climate extremes indices, CMIP5, CORDEX, Costa Rica, perfect prognosis, statistical downscaling
1 | INTRODUCTION

Central America (CAM) has been identified as “the main emerging tropical hot-spot”, with a reduction in total precipitation and an increase in its variability by the end of the 21st century, due to climate change (CC) (Giorgi 2006). The latter could exacerbate existing socioeconomic struggles and affect the high biodiversity present in the developing countries situated on this isthmus. Aguilar et al. (2005) compared data from more than 200 CAM stations between 1961 and 2003 and found that the area is certainly warming in recent decades. Temperature extremes also vary accordingly. Additionally, three countries in the region (Honduras, Nicaragua, and Guatemala) were among the top 10 in the long-term climate risk index (Kreft et al. 2016).

CAM is a narrow isthmus with a very spatially variable climate due to factors such as the steep mountain range that extends almost across its full length, the Inter-Tropical Convergence Zone (ITCZ), the Mid-Summer Drought (MSD), El Niño Southern Oscillation (ENSO), and the Caribbean Low-Level Jet (CLLJ). The convergence of the trade winds near the Equator results in a low-pressure zone, which coupled with almost constant solar radiation throughout the year causes humid air to rise in the region, triggering a great number of thunderstorms and precipitation along the ITCZ band (Barry and Chorley 2009).

The region possesses a well-defined rainy season between May and October (Hastenrath 1967), which is roughly when the ITCZ crosses the area. This period exhibits a bimodal distribution with two precipitation peaks during May–June and September–October, with a relative minimum between July and August (Magaña et al. 1999). This relatively dry period is called MSD.

There have been multiple efforts to explain and categorize ENSO (e.g., Walker, 1932; Troup, 1965; Bjerknes, 1966). Wolter (1989) found it to be the most influential circulation mode, even outside the Pacific Ocean. In CAM, El Niño is related with droughts, loss of crops and cattle, while the counterpart of the phenomenon, La Niña, is associated with heavy rainfall and floods (Intergovernmental Panel of Climate Change (IPCC) 1997).

The CLLJ has shown to be a dominant climatologic feature of the region (Amador 1998; Magaña et al. 1999; Whyte et al. 2007; Wang 2007; Amador 2008). It is defined as the maximum speed of easterly wind at the 925 hPa level in the Caribbean region. The CLLJ possesses a semi-annual variation with peaks in boreal winter and summer (January and July), and minimums in spring and fall. During July the Caribbean basin experiences a maximum of the CLLJ and sea level pressure (SLP), combined with a relative minimum of precipitation (MSD) and minimum tropical cyclogenesis (Wang 2007).

The study area is Costa Rica (CR), which is a country with several climate zones, complex terrain, different precipitation characteristics, high biodiversity contrasts within small distances and an energy matrix consisting of approximately ~75% hydropower, thus highly dependent on precipitation. The focus region is the country’s Great Metropolitan Area, where more than half of its population resides. Consequently, it is necessary to understand the multiple implications of CC in the region.

General circulation models (GCMs) are the most powerful tools for building future CC projections (Huebener et al. 2007). The GCMs employed in climate studies and projections are computed at a spatial resolution of few hundred kilometres and possess significant differences when compared to the observations (Flato et al. 2013), which renders their information unsuitable as input for impact models (von Storch 1995). Because of this coarseness, coupled with limitations in the representation of small-scale processes and the heterogeneity of their boundary characteristics, GCMs cannot precisely describe certain characteristics of regional and global climate (Hidalgo and Alfaro 2014). It is therefore necessary to resolve this spatial discrepancy through a downscaling methodology, named dynamical and statistical. The goal of downscaling is to project the synoptic and large-scale information embedded in low resolution variables to the regional and local scale (von Storch et al. 1993), whereby its result represents climate more accurately while accounting for regional characteristics such as local topography.

Dynamical downscaling (DD), or nested model downscaling, refers to the use of initial and boundary conditions from GCM output to drive high resolution regional climate models (RCMs) (Hallett 2001). Statistical downscaling (SD) uses the GCM or RCM output as predictors of the locally observed data (predictands) by applying statistically derived relationships to obtain the future climate projections (CPs). This linkage is obtained through a transfer function (TF) between scales (Hewitson and Crane 1996). SD requires lower computing power, compared to DD. Multiple SD methods have been described in order to tackle this scale mismatch issue (e.g., von Storch et al., 1993; Hewitson and Crane, 1996; Wilby and Wigley, 1997; Cavazos and Hewitson, 2005).

High resolution CPs are needed to accurately evaluate and rapidly adapt to CC and its likely consequences on crucial subjects such as water availability, agriculture, society, economy and biodiversity. Significant efforts have been made in the region to reduce the scale at which the CPs are available. The Coordinated Regional Climate Downscaling Experiment (CORDEX 2018) offers
multiple RCM output variables based on CMIP5 (Taylor et al. 2012b) projections with a spatial resolution of 0.44° (~50 km) at the CAM domain.

Hidalgo et al. (2013) used the Bias Correction and Spatial Downscaling (BCSD) method to downscale 30 monthly GCMs from the A1B emission scenario to a 0.5° × 0.5° grid in CAM. These models were used to project temperature, precipitation and runoff. This study found a median reduction in yearly precipitation between 5–10% and 10–30% in runoff for the period 2050–2099. It also revealed an average increase of ~4°C in southern CAM (where CR is) during the same interval. Nevertheless, it was stated that the output of the study is not suitable to depict extreme weather or for impact assessment studies, due to the spatial and temporal resolution.

Imbach et al. (2018) employed the Eta Regional Climate Model to downscale HADGEM2-ES projections for the period 2021–2050 assuming the representative concentration pathway (RCP) 4.5 scenario, to an 8 km resolution for CAM. An overall reduction in precipitation was found, especially during the rainy season. Precipitation extremes showed decreasing trends except in CR and Panamá. Increasing temperature was found throughout the region, being greater in the northern part of the isthmus.

Stennett-Brown et al. (2017) performed station scale SD using the statistical downscaling model (SDSM) (Wilby et al. 2002) for Belize and several islands in the Caribbean. The explained variance ($R^2$) between modelled and observed predictands had values ranging from 4 to 60% for temperature and 1–11% for precipitation, which are common performance measurements for this type of model, nevertheless, they convey high uncertainty. Additionally, climate extreme indices (CEI) (Karl and Easterling 1999) were calculated and rendered an increase of consecutive dry days (CDD), increase (decrease) in warm (cool) days and nights. SDSM provides a set of predetermined predictors, from coarse GCM output, which are loaded to the program. This feature might exclude predictors that could transmit a physical forcing to the local climate. Therefore, there is a possibility to improve these models by including tailored predictors related to regional climatic characteristics.

Ribalaygua et al. (2018) performed local SD at the highly biodiverse Gulf of Fonseca between El Salvador, Honduras, and Nicaragua, where an increase in precipitation and temperature was identified by the end of the 21st century. In this study, coarse reanalysis and GCM output were used to obtain the CPs. Different sets of predictors were analysed which contained mainly wind components (at the pressure levels 1,000, 700, 500, and 250 hPa), geostrophic wind components (1,000 and 500 hPa) and specific humidity at 700 hPa. To our knowledge, the last two mentioned examples are the only studies of local-scale CPs available in the region. Due to their particularities and distances, their outcomes cannot be extrapolated to CR.

The aim of this study is to apply SD at the focus region to generate local-scale CPs of temperature and precipitation for the 21st century at daily temporal resolution, using the CMIP5 climate models. In order to improve the overall performance of the methodology, a group of tailored predictors was proposed to describe the climatic particularities of the region, related to the ITCZ, MSD, ENSO, CLLJ, and convective precipitation. Artificial neural networks (ANNs) were employed to build the TFs between scales, using reanalysis data to fit the observed station data. Several ANNs models were trained, under several conditions, for each predictand. The best-performing models were selected, according to their performance emulating the historical data, by means of statistical metrics and suitable skill scores.

The chosen TFs were coupled with the output from CORDEX to project the local climate. This specific set of CPs were employed because it has been found that RCMs reproduce, due to their higher spatial resolution, better than GCMs, the topographic features and the variability of observed temperature and rainfall (Karmalkar et al. 2011; Onyutha et al. 2016). The daily CPs allow a comprehensive survey of the forthcoming climate conditions, which combined with a CEI analysis, renders valuable information for decision makers and engineers regarding water availability, agriculture, energy production, and general impact assessment studies.

This paper is structured as follows: Section 2 presents the employed station data, reanalysis and GCMs/RCMs. In Section 3 we describe the downscaling methodology, the selection of predictors, the model used to create the TFs, the tools employed to analyse the data and the experimental workflow. Section 4 presents the results and discussion related to the model calibration process, the choice of models and predictors, the CPs, the extreme climate analysis, and the advantages and constraints of the used methodology. Lastly, Section 5 renders a summary of the investigation, its conclusions and further research outlook.

## 2 | DATA

The observed station data used in this study were supplied by the Costa Rican National Meteorological Institute (IMN). Daily observations of maximum (Tmax) and
minimum temperatures (Tmin) and precipitation were provided for various stations in the central region of the country, from which an excerpt was taken to apply the SD methodology (see Table 1).

Overall, the temperature shows relatively small changes throughout the year, which is characteristic of the tropics. Nevertheless, a minimum is observed during boreal winter (see Figure 1). For precipitation, a bimodal distribution is seen through the rainy season with a relative minimum around July and August (MSD) along with the dry season from November to April.

The reanalysis data used in this project is ERA-INTERIM (Dee et al. 2011), with a spatial resolution of 0.75°. It provides data from 1979 and will be continuously updated until 2019. Therefore, the observed station data employed was from 1979 until 2017, depending on the record length of each station (Table 1). Several climate variables were extracted for its use as predictors (described in Section 3.1.1) to generate the TFs.

The projected predictors were obtained from the dynamically downscaled models of the CORDEX initiative available for the CAM domain. Ten CORDEX CAM runs are available with daily temporal resolution, which were downscaled using the RCM models RCA4 and RegCM4-3 (CORDEX 2018) to a spatial resolution of 0.44°. This domain embodies the entire area of interest of this study. The boundary GCMs were selected due to their proven skill modelling near surface temperature, precipitation and ENSO teleconnections, according to Hidalgo and Alfaro (2014), taking into account their availability in the online databases (ESGF) (Cinquini et al. 2014).

Two CORDEX runs were selected, the first is based on the model developed by the EC-EARTH consortium, run label r12i1p1, overall ranked #14 (Hidalgo and Alfaro 2014). The second is based on MPI-ESM-LR, provided by the Max Planck Institute for Meteorology, run label r1i1p1, ranked #15 (Hidalgo and Alfaro 2014). Both GCMs were downscaled with the RCA4 model of the Swedish Meteorological and Hydrological Institute (SMHI). For further details on the run labels see Taylor et al. (2012a). The analysed RCP scenario is the 8.5 Wm⁻² (SMHI). For further details on the run labels see Taylor et al. (2012a). The SMHI provided additional variables not available in the databases, such as zonal wind at 925 hPa.

### METHODS

#### 3.1 Statistical downscaling

SD implies establishing quantitative relationships between large or mesoscale atmospheric variables (predictors) and local surface variables (predictands) (Wilby et al. 2004). Most SD methods create a link between the predictors x and the predictands y by a statistical model \( F(.) \) or TF (Maraun and Widmann 2018). The calibration of the model is done with the observed data \( y_{obs} \). The simulation of the local climate \( y_{sim} \) is obtained through the use of the model \( F(.) \) and a simulated array of predictors \( x_{sim} \). Here, a perfect prognosis (PP) approach was used to generate the TFs, which requires temporal correspondence between predictors and predictands, therefore the need of reanalysis data.

In order to properly use the PP methodology, the following assumptions (Maraun and Widmann, 2018, p. 142) should be satisfied: (1) The predictors have to be realistically and bias free simulated. Also, under CC conditions, the response of the predictors to external drivers should be credibly represented. (2) Predictors that explain a large portion of the local variability on the temporal scale of interest should be selected. For CPs, it is necessary to include all relevant predictors that determine future changes. (3) The influence of the predictors on the predictand, and also the interaction between predictors, needs to be incorporated in the downscaling model. Under CC conditions, it is likely that unseen climate will happen, thus, the model should allow for at least moderate extrapolation. Therefore, suitable predictors that explain the statistical climate aspects of interest are required.

The last two premises cover what is known as time-invariance or stationarity assumption: “if all predictors relevant for CC are included, and their influence on the predictand is sensibly modelled, also beyond observed states, the model is valid in a future climate” (Maraun and Widmann 2018). In practice, these models are merely estimations of the mesoscale processes, resulting in an approximation of the time-invariance. Also, the assumption that future predictors should be in the range of the observed predictors has been proved to be invalid for key

| Name                          | IMN code | Lat. (°) | Long. (°) | Elev. (m.a.s.l.) | Start           | End              |
|-------------------------------|----------|----------|-----------|------------------|-----------------|------------------|
| Linda vista, El Guarco        | 73018    | 9.83     | −83.97    | 1400             | January 1, 1960 | February 28, 2018|
| La Argentina, Grecia          | 84003    | 10.03    | −84.35    | 760              | January 1, 1960 | October 31, 2015 |
| Rancho Redondo                | 84018    | 9.97     | −83.97    | 1780             | January 1, 1960 | October 31, 2011 |
predictors (Wilby et al. 2004). Thus, the assumption that the model is appropriate for moderate extrapolations is made. The following subsections describe the selection of predictors and the TF.

### 3.1.1 Predictor selection

The selection of predictors depends on the region and season to be downscaled, in order to properly convey the basic requirements for a predictor: informative and a highly predictive power (Maraun et al. 2010). Under CC conditions measures of humidity are needed to obtain the change of the water-holding capacity of the atmosphere due to global warming (Wilby and Wigley 1997). Another factor to take into account is the availability of the desired predictors in the GCMs/RCMs databases, since not all the desired variables at the different pressure levels can be obtained. This reduces the range of predictors to be selected. A group of predictors based on Cavazos and Hewitson (2005) was selected, where predictors for both temperature and precipitation for different regions in the world were analysed. Vertical wind was discarded in this study since it is not available in the CORDEX database.

Also, since a large portion of the local precipitation is related to convective processes, tropospheric stability indices (TSI) will be employed as predictors. TSI are indices that provide simple metrics to characterize tropospheric conditions that promote convection (Barfus and Bernhofer 2015). In this study, three different TSI were calculated and used as predictors in the SD procedure. The indices are the total totals index (TT), the vertical total index (VT) (Miller 1972), and the severe weather index (SWEAT) (Miller et al. 1971).

Lastly, a group of predictors related to regional climatic characteristics, such as the CLLJ and ENSO, was tailored for the study region. To include the influence of the CLLJ, the averaged zonal wind at 925 hPa, sea surface temperature (SST) and sea level pressure (SLP) were calculated, within the area 12.5–17.5°N and 80–70°W, as defined by (Wang, 2007, Figure 2). Since this area is several hundreds of kilometres away from the study region, a predictor with a one-day delay in the zonal wind (925 hPa) was included to compensate the distance travelled by the CLLJ.

Wolter and Timlin (1993) defined the multivariate ENSO index (MEI) as the first principal component of six observed fields: SLP, zonal and meridional surface wind components; SST, near-surface air temperatures; and total cloudiness. It is calculated on a bi-monthly basis between the coordinates 5°N–5°S and 170–120°W. Due to the MEI's time scale, its relative low percentage of total explained variance (Wolter and Timlin 1993), and the great distance from CR, a region between the Equator and the Pacific coast of CR was proposed to directly evaluate the influence of the six variables that compose MEI on the predictands. This region is defined by the coordinates 4–9°N and 89–84°W, shown in Figure 2 as Pacific. The purpose of this selected area is to obtain predictors that contain information about the large and medium-scale climatic drivers related with both ENSO and the ITCZ.

Table 2 contains the details of the raw surface and atmospheric predictors, and in which set it was employed, hereafter known as predictor set one (PS1).
and two (PS2). The final selection of predictors, consisting of 38 different ones, is the union of the predictors shown in Table 2 with the lagged zonal wind at 925 hPa in the CLLJ region (used in PS1), the tropospheric stability indices (PS1) and the geometrical thickness between the 850 and 500 hPa levels (th8) used in both PS1 and PS2. PS1 contains 35 predictors and PS2, 24, which is based only on Cavazos and Hewitson (2005). A large number of the predictors in PS2 are also employed in many other studies (e.g., Goly et al., 2014; Ribalaygua et al., 2018).

In addition to the PSs described, a third set of predictors (PS3) was employed for each predictand. This set consists of the 12 highest correlated predictors, from the

### TABLE 2  Selected surface and atmospheric predictors

| Predictor/predictor set | Surface | Pressure levels (hPa) | Short name |
|-------------------------|---------|-----------------------|------------|
|                         |         | 1 2 1 2 1 2 1 2 1 2 1 2 |            |
| 2m T                    |         | L P L                 | T2m        |
| Max 2m T                |         | L L                   | Tx2m       |
| Min 2m T                |         | L L                   | Tn2m       |
| 2m Dew T                |         | P                     | TD2m       |
| Geopotential            | L L L L L L |                      | L z        |
| Relative humidity       | L L     | L                     | rh         |
| Sea level pressure      | C, P    | L                     | SLP        |
| Sea surface T           | C, P    | L                     | SST        |
| Specific humidity       | L L L L L L |                    | q         |
| Total cloud cover       | L, P L  | L                     | TTC        |
| Total precipitation     | L L     | L                     | Pr         |
| U wind                  | P L C   | L L L L L L L L     | U          |
| V wind                  | P L     | L L L L L L L L     | V          |

Note: The columns 1 and 2 refer to the predictor sets PS1 and PS2, respectively. Abbreviations: T, temperature; U, zonal wind; V, meridional wind; C, averaged in the CLLJ area; P, averaged in the proposed region in the Pacific Ocean; L, locally interpolated to each station.

FIGURE 2  Location of tailored areal predictors and the analysed stations. Also, to illustrate the spatial resolution of the reanalysis (+) and CORDEX (×) data, a small section of their grids is added.
overall set, for each individual predictand. Thus, there were numerous combinations of PS3.

### 3.1.2 Transfer function

The predictor-predictand relationships were modelled using ANNs, which have demonstrated to adequately determine empirical relationships between atmospheric variables and surface climate parameters (e.g., Hewitson and Crane, 1994; Cavazos, 2000). Nevertheless, some mechanisms that favour precipitation belong to smaller scale processes than the ones captured by GCMs/RCMs, hence, a parametrization of these phenomena is performed. Therefore, the linkages derived with ANNs are affected by this generalization. This renders known issues reproducing the statistics and occurrence of precipitation (Coppola et al. 2006).

The ANNs models ran in R language (R Core Team 2018) employing the “neuralnet” R package (Fritsch and Guenther 2016). The dataset was divided in 90% to calibrate the ANNs and 10% independent validation data, known as Cal/Val datasets. Since there are no general rules on ANN architecture, several combinations were tested to optimize the models.

The activation function employed was Sigmoid, which enables the linear architecture nature of ANNs to model nonlinear functions, such as precipitation. The function used to compute the error is the sum of squared errors (SSE). The ANN learning algorithm used was resilient backpropagation with weight backtracking (Rprop) by Riedmiller (1994); which has shown to yield faster convergence and higher accuracy than backpropagation (Prasad et al. 2013), due to its resilient weight update-values routine. This algorithm is not based solely on a learning rate, to avoid the possibility of skipping a local minimum, including increase and decrease parameters.

Scaling the input dataset to a common range reduces the convergence time of the learning algorithm. Numerous normalization methods are employed to optimize the ANNs, such as min-max, standardization, median, MAD (median absolute deviation), and tanh-estimators (Hampel et al. 1981). The tanh-estimators are reported to be robust, highly efficient and not sensible to outliers (Jain et al. 2005).

The normalization known as modified tanh-normalization was applied (Latha and Thangasamy 2011). It does not calculate the Hampel estimators, instead the mean and standard deviation of the variable are used, which is a simpler and faster method. This normalization does not oblige the maximum value to be equal to 1, which provides advantages when extrapolation is required by the TF, contributing to satisfy the third assumption of PP. Equation (1) represents the estimator applied to the input layers, where $x$ is each individual value, $\mu$ the mean and $\sigma$ the standard deviation of each layer.

\[
\text{Estim}_{\text{tanh}}(x) = \frac{1}{2} \left[ \tanh \left( \frac{x-\mu}{100\sigma} \right) + 1 \right] \tag{1}
\]

ANNs have proved to possess good predictive power compared to traditional models. Yet, their interpretation remains challenging, requiring an understanding of the independent effects of each predictor on the predictand. Intrator and Intrator (2001) developed an approach to interpret the ANNs, coined as generalized weights (GWs), which are, in a simplified manner, comparable to the weights of a multiple linear regression. Thus, the GWs were calculated and utilized to explain the influence of the predictors on each model.

### 3.2 Model evaluation

To evaluate the performance of the ANNs, several metrics were calculated to assist the model selection process. The selected metrics are: BIAS, mean absolute error (MAE), mean squared error (MSE), Pearson correlation coefficient ($R$), normalized standard deviation ($\hat{\sigma}$), pierce skill score (PSS, Pierce et al., 2009, Equation (2)), which was conceived to measure the similarity between two climate patterns, and the Taylor skill score (TSS, Taylor, 2001, Equation (3)). In the following equation $m$ and $o$ stand for the mean of the modelled and the observed values, respectively, while $\sigma_o$ represents the standard deviation of the observed data.

\[
\text{PSS} = R^2 - (R - \hat{\sigma})^2 - \left( \frac{m - \bar{o}}{\sigma_o} \right)^2 \tag{2}
\]

\[
\text{TSS} = \frac{4(1+R)^4}{(\hat{\sigma} + \frac{3}{4})(1+R_o)^4} \tag{3}
\]

Taylor diagrams (TDs) were also employed to evaluate the behaviour of the calibrated models. TDs supply a concise statistical summary of the performance of the modelled values against the observed data, based on their correlation, root-mean-squared error and ratio of their standard deviations. For further details see Taylor (2001).

### 3.3 Climate extreme indices (CEI)

The expert team on climate change detection and indices (ETCCDI) defined a set of climate indices that provide a detailed statistical summary of precipitation and
temperature, centred mainly on their extremes (Karl and Easterling 1999). Table 3 presents the complete set of 27 indices, although some indices are irrelevant for the study region, such as frost and ice days. The calculations were done using the R package “climdex.pcic” (Bronaugh 2015).

3.4 | Workflow

The datasets were split by seasons according to the regional climate: the dry season from November to April (NA), the first peak of the rainy season from May to June (MJ), the MSD between July and August (JA), and the highest peak of the rainy season from September to October (SO). The models were also trained annually to evaluate the influence of the proposed predictors on both temporal aggregations. The divided datasets of observed data were then merged with the three PSs from the ERA-INTERIM reanalysis described in Section 3.1.1, independently, to train the ANNs under 10 random training datasets and six different architectures (as represented in Figure 3) for each variable and station.

Subsequently, employing the methods to evaluate the models (Section 3.2), a group of top performing models was selected. The TFs embedded in these models were coupled with the predictors extracted from CORDEX to project the climate at a daily level until the end of the 21st century. Then, further evaluation of the hindcast period, using the CORDEX datasets, was done to discern the most suitable models per predictand to calculate the CEI.

4 | RESULTS AND DISCUSSION

4.1 | Model calibration and selection

The final selection of models was done in a two-step approach, for both seasonal and yearly temporal aggregations. First, the models were ranked according to their PSS. Then, a revision using the methods described in Section 3.2 was done, to ensure the selection of the best model. The latter was modified only in the cases where a sensible difference in the results was expected, without sacrificing performance. The PSS ranking based selection predominantly found the overall best skilled TF. However, three model selections were changed after the inspection, being BIAS the main reason to change the selection. Usually, a decrease in the BIAS was coupled with an improvement of TSS, MAE and/or $\bar{\sigma}$. This was done to remove any persistent differences from the future projections, hence, misleading results.

Table 4 presents the metrics for the case of precipitation at station 84003 during MJ. In this example, the model first ranked as #2 was selected: PSS for #2 is slightly lower, 0.169 versus 0.166, which renders no practical difference, but its BIAS is considerably lower, 1.13 versus 0.18 mm day$^{-1}$, which can affect greatly the projected results.

Higher reliability is expected from unbiased models, therefore, the aforementioned model preselection was modified. Also, the TSS rendered a slightly better result, which was calculated using $R_0$ (Equation (3)) as the mean of all correlation values for each temporal aggregation per predictand. Observe how PS1, which includes the TSI and the proposed tailored predictors in the Pacific and the Caribbean, was present eight out of 10 times in the

| Label  | Index name               | Units          |
|--------|--------------------------|----------------|
| CDD    | Consecutive dry days     | days           |
| CSDI   | Cold spell duration      | days           |
| CWD    | Consecutive wet days     | days           |
| DTR    | Diurnal temperature range| °C             |
| FD     | Frost days               | days           |
| GSL    | Growing season length    | days           |
| ID     | Ice days                 | days           |
| PRCPTOT| Total wet-day precipitation| mm            |
| R1mm   | Number of wet days       | days           |
| R10mm  | Heavy precipitation days | days           |
| R20mm  | Very heavy precipitation days | days |
| R95p   | Very wet days            | mm             |
| R99p   | Extremely wet days       | mm             |
| RX1day | Max 1-day precipitation  | mm             |
| RX5day | Max 5-day precipitation  | mm             |
| SDII   | Simple daily intensity   | mm             |
| SU     | Summer days              | days           |
| TN10p  | Cold nights              | %              |
| TN90p  | Warm nights              | %              |
| TNn    | Min Tmin                 | °C             |
| TNx    | Max Tmin                 | °C             |
| TR     | Tropical nights          | days           |
| TX10p  | Cold days                | %              |
| TX90p  | Warm days                | %              |
| TXn    | Min Tmax                 | °C             |
| TXx    | Max Tmax                 | °C             |
| WSDI   | Warm spell duration      | days           |

*ETCCDI set it as an user-defined threshold, which was established as 1 mm in this study.
best-performing models for this example, which can be interpreted as an overall improvement of the skill when using the customized predictors. The latter applies for most of the seasonal calibrated models, where all three PSs were calibrated. The remaining model selection changes were done during SO for precipitation for station 84003 and for Tmax at station 73018. Additionally, it is worth pointing out that an averaging of the ANNs weights for the different runs under the same HLA was also evaluated. Nonetheless, in most cases this approach rendered noisy results with remarkably low skill when compared to the presented results.

Table 5 comprises the characteristics and metrics of all the selected models. From the 36 seasonal models, 23 were trained with PS1, 10 with PS2 and 3 with PS3. Because of the latter and the aforesaid general improved performance with PS1, coupled with the fact that the training time of the yearly models was significantly longer, due to their substantial larger datasets, only PS1 was calibrated for these models. Furthermore, 29 of the selected seasonal models were trained using a single hidden layer, probably due to the simplification induced by grouping the predictands into similar climate patterns (seasons). Thus, the learning algorithm did not require intricate functions to accurately reproduce the data. In contrast, for six out of nine final yearly models the ANN architecture had two hidden layers, which could be interpreted as an effort of the algorithm to replicate the data by adding complexity to the TF, since the complete dataset was used for the calibration of these models.

Additionally, Figure 4 shows the TDs of the final selection of models by temporal aggregation and by station, simplifying the comparison between the models. In general, the models performed better for NA, for example, the PSS during NA was significantly higher than for the remaining seasons, specially for precipitation at station 84003 where its PSS was 0.516 and from 0.102 to 0.185 for the other seasons (see Table 5). This improved skill for precipitation during NA can also be appreciated in Figure 4c, as a shorter distance between the symbols

TABLE 4  Metrics and characteristics of the top 10 models for precipitation at station 84003 for May–July

| PS | Run | HLA   | MSE  | MAE  | BIAS | R    | σ    | TSS  | PSS  | Rank |
|----|-----|-------|------|------|------|------|------|------|------|------|
| 1  | 9   | 32–16 | 167.29 | 9.17 | 1.125 | 0.424 | 0.36 | 0.684 | 0.169 | 1    |
| 1  | 2   | 34    | 166.17 | 8.71 | 0.176 | 0.408 | 0.40 | 0.764 | 0.166 | 2    |
| 2  | 9   | 20    | 168.21 | 9.29 | 1.099 | 0.429 | 0.31 | 0.546 | 0.164 | 3    |
| 1  | 6   | 34    | 176.93 | 8.81 | 0.647 | 0.422 | 0.31 | 0.536 | 0.164 | 4    |
| 1  | 6   | 26    | 177.22 | 8.67 | 0.822 | 0.422 | 0.31 | 0.536 | 0.162 | 5    |
| 1  | 6   | 30    | 177.40 | 8.76 | 0.740 | 0.418 | 0.31 | 0.530 | 0.161 | 6    |
| 1  | 9   | 28–14 | 168.97 | 9.22 | 1.155 | 0.414 | 0.35 | 0.636 | 0.161 | 7    |
| 1  | 2   | 32–16 | 168.00 | 8.95 | 0.079 | 0.399 | 0.45 | 0.878 | 0.157 | 8    |
| 1  | 6   | 32–16 | 179.00 | 8.82 | 0.817 | 0.412 | 0.30 | 0.493 | 0.154 | 9    |
| 3  | 9   | 8–4   | 170.84 | 9.35 | 1.101 | 0.427 | 0.27 | 0.430 | 0.151 | 10   |

Note: Run is related to the combination of initial random weights and Cal/Val set employed to train the ANN. HLA stands for hidden layer architecture, for example, HLA = 32–16 means that the model was calibrated with two hidden layers, the first had 32 neurons and the second 16. Otherwise, a single number connotes one hidden layer with the stated amount of neurons.
| Var | Sta | TA | PS | Run | HLA | MSE | MAE | BIAS | R | $\sigma$ | TSS | PSS |
|-----|-----|----|----|-----|-----|-----|-----|------|---|-----|-----|-----|
| Tx  | 73018 | MJ  | 1  | 3  | 28–14 | 2.60 | 1.23 | 0.049 | 0.579 | 0.51 | 1.064 | 0.329 |
|     | JA  | 1   | 1  | 28–14 | 1.78 | 1.02 | 0.188 | 0.622 | 0.55 | 1.012 | 0.369 |
|     | SO  | 1   | 2  | 28–14 | 1.38 | 0.90 | $-0.007$ | 0.599 | 0.63 | 1.091 | 0.359 |
|     | NA  | 1   | 7  | 32–16 | 2.42 | 1.22 | $-0.062$ | 0.738 | 0.70 | 1.010 | 0.543 |
|     | Y   | 1   | 9  | 32–16 | 2.24 | 1.13 | 0.029 | 0.715 | 0.68 | 0.968 | 0.509 |
| 84003 | MJ  | 3   | 8  | 8   | 1.52 | 0.90 | 0.033 | 0.748 | 0.66 | 1.022 | 0.551 |
|     | JA  | 2   | 10 | 14  | 1.24 | 0.88 | $-0.039$ | 0.629 | 0.56 | 0.886 | 0.390 |
|     | SO  | 1   | 1  | 32–16 | 1.16 | 0.87 | $-0.155$ | 0.743 | 0.72 | 1.072 | 0.543 |
|     | NA  | 1   | 2  | 34  | 1.51 | 0.91 | $-0.021$ | 0.761 | 0.75 | 1.044 | 0.578 |
|     | Y   | 1   | 2  | 28–14 | 1.35 | 0.89 | $-0.034$ | 0.780 | 0.76 | 0.981 | 0.608 |
| Tn  | 73018 | MJ  | 1   | 10 | 26  | 1.84 | 1.09 | $-0.183$ | 0.588 | 0.53 | 1.002 | 0.330 |
|     | JA  | 1   | 7  | 34  | 1.83 | 1.04 | 0.013 | 0.488 | 0.49 | 0.759 | 0.238 |
|     | SO  | 1   | 9  | 30  | 1.65 | 1.01 | $-0.055$ | 0.502 | 0.45 | 0.789 | 0.248 |
|     | NA  | 1   | 2  | 30  | 2.41 | 1.17 | 0.012 | 0.628 | 0.56 | 0.878 | 0.390 |
|     | Y   | 1   | 5  | 28–14 | 2.09 | 1.10 | $-0.011$ | 0.614 | 0.64 | 0.921 | 0.376 |
| 84003 | MJ  | 1   | 7  | 34  | 0.93 | 0.76 | 0.098 | 0.653 | 0.52 | 1.106 | 0.403 |
|     | JA  | 1   | 8  | 30  | 0.69 | 0.67 | 0.019 | 0.551 | 0.58 | 0.944 | 0.303 |
|     | SO  | 1   | 4  | 30  | 1.06 | 0.75 | $-0.043$ | 0.467 | 0.41 | 0.726 | 0.214 |
|     | NA  | 2   | 5  | 12  | 1.19 | 0.83 | 0.066 | 0.743 | 0.73 | 1.063 | 0.551 |
|     | Y   | 1   | 2  | 26  | 1.17 | 0.81 | $-0.001$ | 0.726 | 0.71 | 0.944 | 0.527 |
| P   | 73018 | MJ  | 2   | 7  | 16  | 116.35 | 7.17 | $-0.637$ | 0.490 | 0.32 | 0.581 | 0.208 |
|     | JA  | 2   | 1  | 16–8 | 57.01 | 4.62 | $-0.088$ | 0.519 | 0.47 | 0.993 | 0.266 |
|     | SO  | 2   | 8  | 12  | 118.45 | 7.94 | 0.754 | 0.412 | 0.41 | 0.727 | 0.165 |
|     | NA  | 2   | 1  | 14  | 23.18 | 2.11 | 0.037 | 0.585 | 0.47 | 1.126 | 0.329 |
|     | Y   | 1   | 9  | 30  | 75.76 | 4.60 | $-0.031$ | 0.497 | 0.43 | 0.739 | 0.243 |
| 84003 | MJ  | 1   | 2  | 34  | 166.17 | 8.71 | $-0.176$ | 0.408 | 0.40 | 0.764 | 0.166 |
|     | JA  | 1   | 6  | 34  | 137.96 | 7.76 | $-1.036$ | 0.439 | 0.40 | 0.790 | 0.185 |
|     | SO  | 1   | 7  | 30  | 219.68 | 10.73 | $-0.044$ | 0.320 | 0.31 | 0.495 | 0.102 |
|     | NA  | 2   | 9  | 16  | 27.16 | 1.79 | 0.211 | 0.760 | 0.51 | 1.569 | 0.516 |
|     | Y   | 1   | 9  | 32–16 | 132.54 | 5.95 | $-0.695$ | 0.503 | 0.44 | 0.770 | 0.246 |
| 84018 | MJ  | 2   | 9  | 14  | 169.79 | 9.03 | $-0.667$ | 0.558 | 0.40 | 0.888 | 0.284 |
|     | JA  | 1   | 2  | 30  | 159.21 | 7.87 | $-0.323$ | 0.573 | 0.43 | 0.789 | 0.309 |
|     | SO  | 3   | 2  | 6   | 272.57 | 12.15 | 1.060 | 0.375 | 0.31 | 0.478 | 0.132 |
|     | NA  | 2   | 1  | 12  | 33.53 | 2.52 | 0.120 | 0.693 | 0.56 | 1.460 | 0.462 |
|     | Y   | 1   | 4  | 28–14 | 133.77 | 6.44 | $-0.035$ | 0.538 | 0.53 | 0.945 | 0.289 |
representing these models and the reference value ($\pm$), when compared to the other seasons. The latter applies also for Tmax (Figure 4a) and Tmin (Figure 4b), which is also reflected in the values of $R$ and $\sigma$. These results suggest that the relative higher atmospheric stability of this period facilitates its modelling. Nevertheless, it is essential to note how for all models the variability of the predicted data was lower than that of its observed counterpart (all $\sigma < 1$). This perceptible deficiency of the models could affect the results of the CPs and the CEI.

Typically in the cases of Tmax and Tmin the seasons related to the rainy period had similar PSS, $R$ and $\sigma$, except for Tmax during JA at 84003 and 84018, and Tmin during SO at 84003, where a relative lower skill was found. For precipitation, SO-related models had considerably lower skill. SO overlaps with enhanced tropical cyclone activity in the Atlantic, which increases atmospheric instability. The latter translates to greater variability of the observed data, thus, challenging modelling conditions for the ANNs.

Some models selected for precipitation presented substantially high BIAS, for example, station 73018 during MJ and SO ($-0.637$ and $0.754$ mm-day$^{-1}$, respectively). Nevertheless, their selection was maintained because the models with significant improvement in BIAS had serious deficiencies in other metrics, for example, ranked #28 and #44 (BIAS = 0.093 and $-0.173$ mm-day$^{-1}$, respectively). Ribalaygua et al. (2018) obtained also for CAM similar values of MAE and BIAS for the seasonal precipitation models, being MAE = 6.2 and BIAS = $-1.80$ mm-day$^{-1}$ the worst case scenarios. These metrics were only available for precipitation. However, the measurements presented in Table 5 for Tmax and Tmin revealed significant improvements compared to precipitation.

Despite their inaccuracy, the yearly models had a fair performance (see Figure 4d) compared to the analogous study in the region. Although CR is not within the studied zones, Stennett-Brown et al. (2017) achieved, for yearly models, a maximum $R^2$ of 53, 60 and 11% while the equivalent values obtained in the present study would be 61, 73 and 29% (Tmax, Tmin, and precipitation, respectively), as derived from Table 5. The mentioned paper presents only the ranges of BIAS, $\sigma$ and $R^2$, therefore, is difficult to make additional comparisons. Nevertheless, a perceptible improvement was obtained, especially for precipitation, with the proposed predictors and methodology employed. Moreover, the metrics of the yearly models were mostly better than for seasonal models.
| Var | Sta | TA | Top predictors |
|-----|-----|----|----------------|
| Tx  | 73018 | MJ | Tx2m* z500* U925-C* TT* q500* SLP-C* z850* V850* q850* T2m* |
|     |       | JA | SLP-P* z850* Tx2m* z200* rh* th8* SLP-C* T2m* q850* q* |
|     |       | SO | Tx2m* th8* q700* SLP-P* SLP-C* q* z200* V700* Tn2m* U925-C* |
|     |       | NA | Tx2m* U850* T2m* q* U925-C* Pr* SLP-P* rh* V850* SLP-C* |
|     | 84003 | Y  | SLP-P* z850* Tx2m* rh* T2m* q* z700* SLP-C* TT* V850* th8* |
|     |       | MJ | Tx2m* Tn2m* q500* T2m* th8* rh* V850* U-P* U850* SST-P* |
|     |       | JA | z500* th8* U700* q850* TCC* TT* SLP-P* U850* T2m* U700* |
|     |       | SO | z500* T2m* q850* T2m* Pr* SLP-C* z850* q700* Tx2m* |
|     |       | NA | z850* U850* T2m* th8* T2m* q* T2m* q850* T2m* |
|     | 84018 | Y  | z850* T2m* Tn2m* q* T2m* T2m* Tn2m* |
|     |       | MJ | th8* z500* SLP-P* z700* T2m* th8* q* SLP-P* |
|     |       | JA | TT* z200* q850* SLP-P* T2m* th8* U700* U850* VT* |
|     |       | SO | z500* T2m* q850* T2m* Th8* q700* T2m* |
|     |       | NA | z850* SLP-P* q* th8* T2m* q700* T2m* U850* T2m* |
|     |       | Y  | z850* SLP-P* th8* z700* rh* T2m* U850* T2m* |
| Tn  | 73018 | MJ | SLP-P* z850* Tn2m* rh* q* U-P* T2m* U850* |
|     |       | JA | th8* z500* SLP-P* Tn2m* z200* q* rh* T2m* z850* T2m* |
|     |       | SO | T2m* z700* z850* T2m* z500* VT* q* SLP-P* |
|     |       | NA | q* th8* T2m* T2m* TT* th8* q850* SLP-P* T2m* |
|     | 84003 | Y  | Tn2m* SLP-P* z850* T2m* q* q850* Z200* Pr* U850* |
|     |       | MJ | th8* q* T2m* z700* z850* T2m* th8* SLP-P* |
|     |       | JA | z500* z850* th8* q* SLP-P* T2m* th8* Pr* U850* |
|     |       | SO | z700* z850* VT* TT* th8* Tn2m* SST-C* V700* SLP-P* V500* |
|     |       | NA | th8* z500* T2m* q700* T2m* q* SLP-P* U850* |
|     |       | Y  | rh* T2m* q* q700* z500* Pr* th8* TT* U850* Tn2m* |
|     | 84018 | MJ | q* Pr* q500* T2m* z200* SLP-P* U850* |
|     |       | JA | T2m* q* z700* rh* SLP-P* z200* z850* U-P* T2m* |
|     |       | SO | q* rh* z850* SLP-P* T2m* th8* z200* z700* V850* |
|     |       | NA | T2m* q700* SST-P* SLP-C* q* T2m* q500* |
|     |       | Y  | rh* q700* q* U850* T2m* SLP-C* SST-P*
| Var  | Sta | TA  | Top predictors |
|------|-----|-----|----------------|
| P    | 73018 | MJ  | V* q500* U700* q* z500* q850* V850* z200* U850* rh |
|      | JA  | th8† | z700* U* z200* q500* q500* z500* T2m* Pr* V500* rh |
|      | SO  | SLP* | z700* U700* z850* U850* V* V500* q500* q* z200* |
|      | NA  | z500* | th8* rh* q* Pr* z850* T2m* SLP* q500* U850* |
|      | Y   | z850† | th8† z500† SLP-P* TT* VT* U850* q500* |
|      | 84003 | MJ  | z500² th8* z850* SLP-P* z700* TT* q850* VT* T2m* rh |
|      | JA  | SLP-P* | T2m* z850* rh* q* U850* th8* Tx2m* V500* q850* |
|      | SO  | z700* | SLP-P* rh* T2m* z500* V850* VT* q* |
|      | NA  | th8* | z500* rh* Pr* T2m* q* q850* U850* |
|      | Y   | SLP-P* | TT* z850* VT* rh* z500* VT* q850* U850* |
|      | 84018 | MJ  | th8* V* q850* z200* q500* q* rh* z500* SLP* Tx2m|
|      | JA  | z850* | T2m* z500* th8* q* rh* SLP-P* z700* z200* SLP-C* |
|      | SO  | q500* | U850* T2m* V700* TCC-P* U-P* V850* q700* U700* |
|      | NA  | rh† | q* T2m* SLP* th8* z500* U850* z700* q500* |
|      | Y   | SLP-P* | z700* VT* q500* Pr* TD2m-P* rh* U850* TT* q850* |

Note: The short names used for the predictors correspond to the ones in Table 2. Note that the pressure levels and/or region of the predictors (C and P) were concatenated to the short names. GWs symbols based on the explained variance per predictor: × < 5%; 5% ≤ * ≤ 10%; 10% ≤ A < 15%; 15% ≤ V < 20%; 20% ≤ † < 25%; 25% ≤ ‡ < 26%.
However, it is crucial to further evaluate their ability to mimic their variability, extremes and temporal allocation, as discussed in Section 4.4.

4.2 Predictors

Following the selection of the final TFs for the CPs, a ranking of the predictors with higher influence on each model was made using the GWs as shown in Table 6, which is complementary to Table 5. One of the most outstanding results of the seasonal precipitation models was the case of station 84018, during SO. This model was obtained with PS3 (only the 12 highest correlated predictors) and a rather simple ANN architecture consisting of one hidden layer with six neurons (see Table 5). This means that the precipitation process is, compared to other models, easily explained by few predictors and a straight-forward TF for this specific season and location.

The top six predictors for this example are specific humidity at 500 hPa (q500 with 16.7% of averaged explained variance), zonal wind at 850 hPa (U850 → 16.2%), near-surface temperature (T2m → 12.7%), and meridional wind at 700 hPa (V700 → 12.3%), all at the local level, and total cloud cover and zonal wind, at the proposed region in the Pacific Ocean in front of the coast of CR (TCC-P → 10.5% and U-P → 8.7%, respectively), accounting for ~77.1% of the total variance. The last two predictors reveal the synoptic correlation between the humidity conveyed by clouds in the Pacific, their movement towards CR and precipitation events. Both specific humidity at the mid-troposphere and near-surface temperature are related to vertical motion and convective processes that ease precipitation (Cavazos and Hewitson 2005). Also, the most influential predictor, q500, gives a measure of the amount of precipitation for wet days (Holton and Hakim 2012), which is crucial in the rainy season. The influence of U850 can be interpreted as a conveyor of humidity from both oceans to mainland.

During NA it was observed for all variables that specific (q) and relative humidity at the surface (rh), together with tropospheric thickness (th8), play a major role in the models. The latter intensifies in the case of precipitation, or its absence during this season, where these three predictors are among the seven more influential for all stations. Relative humidity is suited to conclude if precipitation will occur on a given day (Ribalaygua et al. 2018). This was appropriately estimated by the ANNs with rh found within the top predictors in most TFs for precipitation. Among the common top predictors for the seasonal precipitation models were: SLP-P, and local: rh, q500, th8, and geopotential (z500, z700, and z850).

SLP at the Pacific Ocean (SLP-P) was among the top four predictors during the rainy season for station 84003, which is the closest station to the Pacific coast. This finding confirms the relevance of including tailored regional predictors. Low-pressure systems tend to form in this region during this time of the year due to the shifting of the ITCZ (Hidalgo et al. 2015), hence the major role of SLP-P during the rainy season, as detected by the ANNs.

In the case of seasonal Tmax, specifically at station 73018, a great influence of the predictors calculated for the Caribbean region was found during all seasons. This station is situated in a watershed that flows into the Caribbean Sea and is also the closest to it. The latter ratifies the skill gain with the proposed predictors, specially the CLLJ index (U925-C) and SLP-C. In general, among the top predictors for seasonal Tmax were found: near surface maximum daily temperature (Tx2m), th8, SLP-P, geopotential height at different pressure levels and TT. Similar results for the most influential predictors were obtained for seasonal Tmin, exchanging Tx2m for its minimum counterpart (Tn2m). Moreover, another model belonging to PS3 was observed (MJ at 84003 for Tmin), where an HLA of only eight neurons was used, suggesting a simpler TF for this specific predictand.

The main reason for not using the principal components of grouped variables as predictors to train the ANNs is that in this study it was essential to be able to determine the added value, skill and/or influence of the proposed predictors. This traceability becomes blurred by the principal components analysis, since it summarizes the information in terms of explained variance per predictor, which turns fuzzier when coupled with ANNs. Nevertheless, for further research, we encourage to employ them and the hereby proposed predictors that demonstrated valuable skill.

4.3 Climate projections

The CPs $y_{sim}$ are obtained after applying the simulated PS $x_{sim}$ to the TFs calibrated for each predictand. Two models were selected per each predictand: a seasonally (which contains four models) and a yearly calibrated one. There is a limited number of CORDEX RCMs available for CAM, therefore, a selection was made from the obtainable models according to Hidalgo and Alfaro (2014). The simulated outputs from two RCP8.5 RCMs were used to downscale the historical and future climate, until the end of the 21st century. Thus, a total of four projections per predictand were obtained. The projections were not averaged to represent the variability of the different
combinations of TFs and RCMs. Figure 5 shows the boxplots of all the daily projected predictands grouped by period.

Most of the simulations possess an appreciable BIAS during the historical run period between observed and hindcast data. The selected ANNs models were chosen procuring an almost negligible BIAS. Therefore, this difference suggests a BIAS between the predictors of the RCM model and the reanalysis data and/or a variation of their ranges. Thus, the predictors were analysed during the historical period and sensible differences in the distribution of the datasets were found among them. Approximately only 15% of the predictors achieved a \( p \) value \( > 0.05 \) for both Student's \( t \) test and \( F \) test. This feature could be explained by the parametrizations of the RCMs, which could have produced data in a slightly different range than the reanalysis employed to train the ANNs. The latter could cause the noisy behaviour of some predictions, where the predictand range was extended in comparison to the observed dataset, specially in the yearly models, such as the for precipitation at stations 84003 and 84018, Tmax at station 84003 and Tmin at 73018 (see Figure 5). Hence, a delta-change approach was used to compare the results of the projections.

Generally, the average temperature projections agree with coarser scale estimates. For example, Knutti and Sedláček (2013) found a mean intermodal surface temperature change of \( \approx 4{\degree}C \) in the region at the end of the century, with high robustness for CMIP5, which agrees with the present results. However, we analysed only two different RCMs. Nevertheless, a similar warming trend, as the observed particularly for Tmin, is expected for an ensemble analysis, which in order to add reliability to the CPs, is encouraged for further research in CR and CAM. Additionally, the locally projected changes in temperature agree with the only antecedent SD studies found for the region (Stennett-Brown et al. 2017; Ribalaygua et al. 2018).

Precipitation is a much more complex variable to reproduce reliably. Knutti and Sedláček (2013) found intermodal inconsistencies in the late-century projections for precipitation during the boreal winter near the study area. CR is precisely in the confluence zone of several precipitation change ratios without intermodal robustness. Nevertheless, the study region is projected to have a yearly relative change in precipitation approximately between \( -10\% \) and \( -20\% \). The mentioned study does not show the variability during the local peak of the rainy

---

**Figure 5** Boxplots of the daily projected predictands, grouped by period, station and variable. The observed dataset (Obs) is added for comparison. S.EC stands for seasonal and Y.EC for yearly modelled projections done with the EC-EARTH model. Analogously, S.MPI and Y.MPI represent the projections made with the MPI-ESM-LR model. The diamonds (♦) represent the mean values. 

**H** correspond to the historical run period, 1979–2005, **M** to 2020–2050 and **L** to 2070–2100. In case of precipitation, a square root transformation was applied to the y axis.
season, therefore it is not possible to compare the annual change in precipitation. Yet, employing the seasonal models, which had a better performance and a less noisy behaviour than the yearly ones (Figure 5), for stations 84003 and 84018, an overall change in precipitation of the same magnitude as in Knutti and Sedláček (2013) was found.

At the regional level, Ribalaygua et al. (2018) found a considerable increase in precipitation during the dry season of approximately 60%, almost no difference during both peaks of the rainy season, and a slight decrease (around −10%) during the MSD at the Gulf of Fonseca. Although the distance and climatic differences between this area and the focus region are not substantial, the results for precipitation seem to be considerably divergent. Such different behaviour for precipitation in these adjacent areas is also seen in Knutti and Sedláček (2013) due to intermodal inconsistencies and low robustness, particularly during the dry season.

In the case of the yearly model-based projections for precipitation, a larger amount of extreme rainfall values is seen than in the observed data, especially with the EC-Earth model, which confirms the extrapolation capacity of the methodology. Seasonal precipitation models tend to model average daily precipitation well but lack the skill to mimic the amount of extreme events. All of the aforementioned statements should be taken into account when calculating the CEI and deeply scrutinized for future model improvements.

4.4 | Extreme climate

In order to provide reliable projected extreme climate and CEI, a detailed analysis of the performance of the ANN models for extremes was carried out for the hindcast period at station 84003, which more accurately represents the conditions where a large portion of the country's population lives. It is essential to distinguish between the noise or misrepresented past conditions and actual CC signal embedded in the projected predictors. Therefore, a set of statistical plots was arranged to summarize the model behaviour under extreme climate, as described in Figure 6. Even though the results presented there refer to the best-performing models, some deficiencies were detected, especially for the yearly models. According to their PDFs, the days with precipitation ranging from 1 to 10 mm day$^{-1}$ and extreme values (>100 mm day$^{-1}$), were greatly overestimated, unlike days without precipitation, in accordance with Dai (2006). Thus, the projections with these TFs will also lack the necessary skill to produce reliable CEI results.

On the other hand, the seasonal models yielded better skill modelling both the PDFs and the daily behaviour of

![Figure 6](image-url)
climate. Especially, S.MPI resembles better the observed data, despite the underestimation during MJ. S.EC does not reproduce the MSD at all, while setting the peak of precipitation during these months (JA). Therefore, the model that reproduced better the statistics for this predictand is S.MPI, which was used to calculate the CEI. Since the seasonal models were calibrated separately, there were large jumps at the boundaries between seasons, which should be addressed in succeeding research under different machine learning algorithms, in order to account for these transitions.

An analogous summary is shown in Figure 7 for Tmax. Seasonal runs have higher BIAS and variability than yearly models, while the highest values are considerably lower (~32°C) for both compared to a maximum recorded value of ~36°C. The yearly models yielded an overall better performance, yet, overestimate the range of Tmax. All models tend to underestimate the values during SO with a particularly noisy behaviour, which is reflected by a larger gap between the 5 and 95% quantile lines. Thus, the indices minimum and maximum Tmax (TXx and TXn) were discarded, since their results will not provide reliable information. The model that best replicates the observed statistics for Tmax is Y.EC. Nevertheless, to calculate the CEI with RCM output correspondence, Y.MPI was selected. This meant not using the best available model.

Subsequently, the same summary was done for Tmin (not shown). The values obtained with the yearly models have a higher BIAS and a smaller variability than the observed data. The seasonal models were able to reproduce better the variability of the data. Abrupt changes or mismatches in the estimates nearby MJ were observed, especially in the EC-EARTH model, but also for MPI. Except for this trait, S.MPI successfully reproduces the observed dataset throughout the year. Thus, this model was used to calculate the CEI.

Hence, the CEI for station 84003 were calculated using the aforementioned models and a base-period spanning between 1979 and 2005. Figure 8 presents the boxplots of a selection of indices for the different periods analysed. There, the influence of certain traits of the selected models for the CEI calculation can be noticed, for example, the S.MPI model yielded a higher than observed average precipitation during JA and SO (see Figure 6). This is reflected in the average yearly precipitation (PRCPTOT), where the mean observed rate is 1941 mm·year⁻¹ and the modelled one is 3,038 mm·year⁻¹. Please note: there was missing data in the observed dataset, which could have affected the indices.

Furthermore, the 5% quantile during SO (Figure 6) suggests that the amount of days without precipitation decreases, which affected indices such as the number of wet days per year (R1mm with an average of 135.8 for the observed data and 166.4 for the historical run data) and simple daily intensity (SDII, 14.2 vs. 18.3 mm·year⁻¹). Likewise, the 75% quantile of the observed data is below 10 mm·day⁻¹, while the modelled one is ~20 mm·day⁻¹. Thus, the heavy (R10mm) and very heavy precipitation days (R20mm) per year indices were greatly overestimated. For instance, the observed R10mm has a yearly mean of 64.6 days and the modelled for the historical run 108 days, while the observed R20mm is 33.8 days and the modelled one is 63.7 days. Similarly, maximum 1 and 5 days precipitation are overestimated. For more details on the dispersion of the CEIs, see Figure 8.
Although the mentioned precipitation indices in the historical run are mostly overestimated compared to those observed, the changes within the modelled indices among the analysed periods are noteworthy (delta-change approach). Overall, as mentioned above, lower amounts of yearly precipitation are expected. Almost all the precipitation related indices are expected to decrease over time, for example, an average 31% less PRCPTOT by the period 2070–2100. These differences agree with Sillmann et al. (2013b) for CAM on a coarser scale at the end of the century for an ensemble under the RCP8.5 scenario.

As suggested by the projected CEI, at the end of the century there will be a significantly warmer climate in the region. The cold nights index (TN10p), which presents the ratio of days below the temperature related to the 10% quantile per year calculated from the historical run period, will be almost zero (0.05%). The cold days index (TX10p) will also decrease greatly to 1.6%, on average. While the warm nights (TN90p) and warm days indices (TX90p) will increase greatly to 90.1 and 43%, respectively.

Moreover, the amount of tropical nights (TR) reproduced by the models projects an extremely large change from an average value of 9–275.9 days-year\(^{-1}\) at the end of the century. Also, the diurnal temperature range will decrease, due to greater warming of Tmin than Tmax, which leads to warmer overall climatic conditions. In summary, both daily Tmin and Tmax will increase, yet Tmin will have a more intense change than Tmax. These results agree with the values presented by other authors (e.g., Morak et al., 2013; Sillmann et al., 2013a,b), derived on a larger scale, both for historical and projected scenarios.

### 5 | SUMMARY AND OUTLOOK

The SD methodology under a PP approach was applied, to our knowledge, for the first time to project daily local-scale climate in CR until the end of the 21st century, by means of dynamically downscaled CMIP5 models (CORDEX CAM domain) for the RCP8.5 scenario. Three different PSs, extracted from the ERA-INTERIM reanalysis, were employed to build the TFs between scales, in order to emulate observed data of precipitation, Tmin and Tmax at three locations. These functions were optimized through a large number of ANNs, employing...
several combinations of architectures, temporal aggregations and random training subdatasets.

Considerable skill improvement was found with PS1, which included the proposed tailored predictors aiming to describe the regional climatic characteristics, such as ENSO, CLLJ, ITCZ, MSD and convective precipitation. Many of these predictors were calculated in sections of the oceans enveloping the study region, where synoptic and physical relationships with local-scale climate were expected and confirmed by the ANNs.

The performance of the TFs varied, although with satisfactory results. As expected, the ANNs trained for temperature had a better skill than those for precipitation. Accurate reproduction of the statistics, occurrence and amount of precipitation is a known issue, for both DD and SD, due to the parametrizations used for small-scale processes that favour convective precipitation and its associated BIAS, which affects the TFs. Nevertheless, the best-performing models provided valuable information on future climate.

Generally, drastic changes in the climate regime are expected by the end of the century, such as a considerable decrease in the yearly total precipitation and a significant increase in temperature, specially Tmin (~4°C), using a delta-change approach, which agrees with larger scale projections. In order to provide insightful data projected CEI were calculated at station 84003, located in La Argentina, Grecia, CR. The amount of cold days and nights per year at the end of the century will almost cease to exist while the warm days and nights will rise steeply by ~80% and ~33%, respectively. The mean number of tropical nights per year will increase by ~267 days and the yearly average precipitation will decrease ~31%. Since only one model at one station was analysed for the CEI, a multi-site local-scale ensemble approach is required to provide more robust and spatially detailed information for the region.

Further research employing vanguard machine learning algorithms is intended in order to enhance the performance of the TFs, particularly for extreme events of both precipitation and temperature, which are essential for CC impact evaluation and where there is potential for improvement. Moreover, additional DD efforts with higher impact evaluation and where there is potential for enhancing algorithms is intended in order to urgently implement adaptation schemes.

ACKNOWLEDGEMENTS
We want to thank: the National Meteorological Institute of Costa Rica (IMN), which facilitated the observed station data; the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-INTERIM reanalysis data; and the CORDEX initiative for the climate projections, particularly Dr. Grigory Nikulin of the Swedish Meteorological and Hydrological Institute (SMHI) for supplying additional variables not available in the online database. We appreciate the generous allocations of compute resources by the Center for Information Services and High Performance Computing (ZIH) of the Technische Universität Dresden and the support of the Competence Center for Scalable Data Services and Solutions Dresden/Leipzig (ScaDS). Additionally, we sincerely thank two anonymous reviewers whose suggestions and objective criticism helped to improve this paper.

ORCID
Dánnell Quesada-Chacón https://orcid.org/0000-0001-5700-1846

REFERENCES
Aguilar, E., Peterson, T.C., Ramirez Obando, P., Frutos, R., Retana, J.A., Solera, M., Soley, J., Gonzalez Garcia, I., Araujo, R.M., Rosa Santos, A., Valle, V.E., Brunet, M., Aguilar, L., Álvarez, L., Bautista, M., Castaño, C., Herrera, L., Ruano, E., Sinay, J.J., Sánchez, E., Hernández Oviedo, G.I., Obed, F., Salgado, J.E., Vázquez, J.L., Baca, M., Gutiérrez, M., Centella, C., Espinosa, J., Martínez, D., Olmedo, B., Ojeda Espinoza, C.E., Núñez, R., Haylock, M., Benavides, H. and Mayorga, R. (2005) Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. Journal of Geophysical Research: Atmospheres, 110, D23. https://doi.org/10.1029/2005JD006119.

Amador, J.A. (1998) A climatic feature of the tropical Americas: the trade wind easterly jet. Tópicos Meteorológicos y Oceanográficos, 5(2), 1–13.

Amador, J.A. (2008) The intra-Americas sea low-level jet. Annals of the New York Academy of Sciences, 1146(1), 153–188. https://doi.org/10.1196/annals.1446.012.

Barfus, K. and Bernhofer, C. (2015) Assessment of GCM capabilities to simulate tropospheric stability on the Arabian peninsula. International Journal of Climatology, 35(7), 1682–1696. https://doi.org/10.1002/joc.4092.

Barry, R.G. and Chorley, R.J. (2009) Atmosphere, Weather and Climate. London: Routledge.
Bjerknes, J. (1966) A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18(4), 820–829. https://doi.org/10.3402/tellusa.v18i4.9712.

Bronaugh, D. (2015). *For the Pacific Climate Impacts Consortium*, *climdex.pcc: PCIC Implementation of Climdex Routines*. R package version 1.1-6. https://CRAN.R-project.org/package=climdex.pcc.

Cavazos, T. (2000) Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the balkans. *Journal of Climate*, 13(10), 1718–1732. https://doi.org/10.1175/1520-0442(2000)013<1718:USOMTI>2.0.CO;2.

Cavazos, T. and Hewitson, B.C. (2005) Performance of ncep–ncar reanalysis variables in statistical downscaling of daily precipitation. *Climate Research*, 28(2), 95–107.

Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., Ananthakrishnan, R., Miller, N., Denvil, S., Morgan, M., Pobre, Z., Bell, G.M., Doubtiaux, C., Drach, R., Williams, D., Kershaw, P., Pascoe, S., Gonzalez, E., Fiore, S. and Schweitzer, R. (2014) The earth system grid federation: an open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 36, 400–417. Special Section: Intelligent Big Data Processing Special Section: Behavior Data Security Issues in Network Information Network Propagation Special Section: Energy-efficiency in Large Distributed Computing Architectures Special Section: eScience Infrastructure and Applications. https://doi.org/10.1016/j.future.2013.07.002.

Coppola, E., Grimes, D.I.F., Verdeccia, M. and Visconti, G. (2006) Validation of improved TAMANN neural network for operational satellite-derived rainfall estimation in Africa. *Journal of Applied Meteorology and Climatology*, 45(11), 1557–1572. https://doi.org/10.1175/JAM2426.1.

CORDEX (2018). *Cordex – Esgf Data Availability Overview*.

Dai, A. (2006) Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, 19(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1.

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., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E. V., Isaksen, L., Källberg, 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(656), 553–597. https://doi.org/10.1002/qj.828.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guiyllardi, É., Jakob, C., Kattsov, V., Reason, C. and Rummukainen, M. (2013) Evaluation of climate models. In: *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, England: Cambridge University Press, pp. 741–866.

Fritsch, S., Guenther, F. (2016). *neuralnet: Training of Neural Networks*. R package version 1.33.

Giorgi, F. (2006) Climate change hot-spots. *Geophysical Research Letters*, 33, 8. https://doi.org/10.1029/2006GL025734.

Goly, A., Teegavarapu, R.S.V. and Mondal, A. (2014) Development and evaluation of statistical downscaling models for monthly precipitation. *Earth Interactions*, 18(18), 1–28. https://doi.org/10.1175/EFI-D-14-0024.1.

Hallett, J. (2001) Climate change 2001: the scientific basis. Edited by I.T. Houghton, Y. Ding, D.J. Griggs, N. Noguer, P.J. Van der Linden, D. Xiaosu, K. Maskell and C.A. Johnson. Contribution of working group I to the third assessment report of the inter-governmental panel on climate change, Cambridge University Press, Cambridge. 2001. 881 pp. ISBN 0521 01495 6. *Quarterly Journal of the Royal Meteorological Society*, 128(581), 1038–1039. https://doi.org/10.1002/qj.200212858119.

Hampel, F.R., Rousseeuw, P.J. and Ronchetti, E. (1981) The change-of-variance curve and optimal descending m-estimators. *Journal of the American Statistical Association*, 76(375), 643–648.

Hastenrath, S.L. (1967) Rainfall distribution and regime in central America. *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, 15(3), 201–241. https://doi.org/10.1007/BF02243853.

Hewitson, B.C. and R.G. Crane (1994). *Neural Nets: Applications in Geography: Applications for Geography*, Volume 29. Dordrecht: Springer Science & Business Media.

Hewitson, B.C. and Crane, R.G. (1996) Climate downscaling: techniques and application. *Climate Research*, 7(2), 85–95.

Hidalgo, H.G. and Alfaro, E.J. (2014) Skill of cmip5 climate models in reproducing 20th century basic climate features in central America. *International Journal of Climatology*, 35(12), 3397–3421. https://doi.org/10.1002/joc.4216.

Hidalgo, H.G., Amador, J.A., Alfaro, E.J. and Quesada, B. (2013) Hydrological climate change projections for central America. *Journal of Hydrology*, 495, 94–112. https://doi.org/10.1016/j.jhydrol.2013.05.004.

Hidalgo, H.G., Durán-Quesada, A.M., Amador, J.A. and Alfaro, E.J. (2015) The Caribbean low-level jet, the inter-tropical convergence zone and precipitation patterns in the intra-Americas sea: a proposed dynamical mechanism. *Geografiska Annaler: Series A, Physical Geography*, 97(1), 41–59.

Holton, J.R. and Hakim, G.J. (2012) *An Introduction to Dynamic Meteorology*, Volume 88. Oxford, England: Academic Press.

Huebner, H., Cubasch, U., Langematz, U., Spanghel, T., Niehörsler, F., Fast, I. and Kunze, M. (2007) Ensemble climate simulations using a fully coupled ocean–troposphere–stratosphere general circulation model. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2089–2101. https://doi.org/10.1098/rsta.2007.2078.

Imbach, P., Chou, S.C., Lyra, A., Rodrigues, D., Rodriguez, D., Latinovic, D., Siqueira, G., Silva, A., Garofolo, L. and Georgiou, S. (2018) Future climate change scenarios in central America at high spatial resolution. *PLoS One*, 13(4), 1–21. https://doi.org/10.1371/journal.pone.0193570.

Intergovernmental Panel of Climate Change (IPCC) (1997). *The Regional Impacts of Climate Change. An Assessment of Vulnerability*. Geneva: IPCC.

Intrator, O. and Intrator, N. (2001) Interpreting neural-network results: a simulation study. *Computational Statistics & Data Analysis*. 35(1), 43–63.
material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA, 27.

Wilby, R.L., Dawson, C.W. and Barrow, E.M. (2002) sdsr - a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, 17(2), 145–157. https://doi.org/10.1016/S1364-8152(01)00060-3.

Wilby, R.L. and Wigley, T.M.L. (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography: Earth and Environment*, 21(4), 530–548. https://doi.org/10.1177/030913339702100403.

Wolter, K. (1989) Modes of tropical circulation, southern oscillation, and sahel rainfall anomalies. *Journal of Climate*, 2(2), 149–172. https://doi.org/10.1175/1520-0442(1989)002¡0149:MOTCSO¿2.0. CO;2.

Wolter, K., Timlin, M.S. (1993). Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proc. of the 17th Climate Diagnostics Workshop*, 1993.

**How to cite this article:** Quesada-Chacón D, Barfus K, Bernhoefer C. Climate change projections and extremes for Costa Rica using tailored predictors from CORDEX model output through statistical downscaling with artificial neural networks. *Int J Climatol*. 2021;41:211–232. https://doi.org/10.1002/joc.6616