

Assessing statistical downscaling in Argentina: Daily maximum and minimum temperatures

Rocio Balmaceda-Huarte^{1,2,3} | María Laura Bettolli^{1,2,3}

¹Department of Atmospheric and Ocean Sciences, Faculty of Exact and Natural Sciences, University of Buenos Aires (DCAO-FCEN-UBA), Buenos Aires, Argentina

²National Council of Scientific and Technical Research (CONICET), Buenos Aires, Argentina

³Institut Franco-Argentin d'Etudes sur le Climat et ses Impacts, Unité Mixte Internationale (UMI-IFAEI/CNRS-CONICET-UBA), Buenos Aires, Argentina

Correspondence

Rocio Balmaceda-Huarte, Faculty of Exact and Natural Sciences, University of Buenos Aires, Intendente Güiraldes 2160, Ciudad Universitaria, C1428EGA Buenos Aires, Argentina.

Email: rbalmaceda@at.fcen.uba.ar

Funding information

Fondo para la Investigación Científica y Tecnológica, Grant/Award Numbers: PICT-2018-02496, PICT-2019-02933; Universidad de Buenos Aires, Grant/Award Numbers: 20020170100357BA, 2018-20020170100117BA

Abstract

Empirical statistical downscaling (ESD) under the perfect prognosis approach was carried out to simulate daily maximum (Tx) and minimum temperatures (Tn) in 101 meteorological stations over the different climatic regions of Argentina. To this end, three ESD families were evaluated: analogs (AN), generalized linear models (GLM) and artificial neural networks (ANN) considering a variety of predictor sets with multiple configurations driven by three different reanalyses (ERA, JRA, NCEP). ESD models were cross-validated using folds of nonconsecutive years (1979–2014) and then evaluated in a warmer set of years (independent warm period, 2015–2018) to assess their extrapolation capability. Depending on the aspect analysed, AN, GLM or ANN models were more/less skilful, but no method fulfilled all the features of both predictand variables. In this sense, the predictor set and model configuration were key factors. For each ESD method, the different predictor structures (point-wise, spatial-wise and combinations of them) introduced the main differences, regardless of the predictand variable, region and reanalysis choice. However, some specific results could be highlighted. ERA (NCEP)-driven ESD models were the most (least) skilful in representing Tx and Tn. In the case of Tn, models' skills considerably increased when humidity information was included in the predictor set. Our results showed that downscaling models were able to capture the general characteristics of Tx and Tn in all regions, with better performance in the latter variable. However, regions with complex topography (Argentinian Patagonia and the subtropical Andes) pose a further challenge for capturing the local variability of daily extreme temperatures. The performance of the ESD models in the atypical warm conditions was similar to the one during the cross-validated period, showing some extrapolation skill. The results of this work set a reference for future ESD developments and comparisons in Argentina.

KEY WORDS

analog, artificial neural networks, generalized linear models, maximum and minimum temperature, perfect prognosis, reanalysis, regional climate downscaling, southern South America

1 | INTRODUCTION

Daily maximum and minimum temperatures (T_x and T_n , respectively) are very important variables for the assessment of climate change impacts (IPCC, 2021). They influence hydrologic regimes, condition crop yield responses and are decisive climate stressors for population health and energy management (Penalba *et al.*, 2007; Almeira *et al.*, 2016; García *et al.*, 2018; Montroull *et al.*, 2018). Moreover, the definition of temperature-related climate hazards such as heat waves and frost events are based on these daily extreme temperatures (Müller, 2007; Perkins, 2015). Many of the climate change indices relevant to climate change monitoring and detection (ETCCDI; Klein Tank *et al.*, 2009) and other heat-related indicators targeted to impact assessment (for instance, heat stress Schwingshackl *et al.*, 2021) are also built on daily T_x and T_n .

Global mean temperature has increased over the last decades but with uneven warming levels over the different regions of the planet (IPCC, 2021). Together with the increase of mean temperature, extreme temperature events have been identified as one of the main concerns related to climate change impacts since a worldwide increase in their frequency and/or intensity are expected in future climate projections (Coppola *et al.*, 2021; IPCC, 2021). A variety of literature has recorded an increase in the temperatures in different parts of Argentina, mostly through the study of extreme indices (e.g., Skansi *et al.*, 2013; Barros *et al.*, 2015; Rusticucci *et al.*, 2016; Lovino *et al.*, 2018; Olmo *et al.*, 2020). These studies commonly identified significant increases (decreases) in the frequency of warm (cold) nights/days over central and northeastern Argentina. Also, part of this warming signal was found in Argentinian Patagonia in a more recent period (Olmo *et al.*, 2020; Balmaceda-Huarte *et al.*, 2021). Only a few works addressed long-term changes regarding mean T_x and T_n . Ceccherini *et al.* (2016) studied their trends on an annual basis for all South America and identified that maximum temperature was generally increasing faster than the minimum temperature over the extratropics in the period 1980–2014. In the period 1979–2017, Balmaceda-Huarte *et al.* (2021) analysed the annual mean of maximum and minimum temperatures using station data and multiple reanalyses and found significant upward trends for both variables in different regions over southern South America. Using reanalyses data only, Cogliati *et al.* (2021) evaluated the trends in the seasonal cycle of mean minimum temperature over La Plata Basin in the period 1980–2015 and detected a general shift towards warmer conditions. In this sense, both reliable observations and simulations of daily maximum and minimum temperatures are

required to improve our understanding of climate variability impacts on socio-economic activities and the response of extreme events to global warming.

The continental territory of Argentina extends from approximately 21°S to 55°S covering a variety of orography, extensive coastlines, and low continental lands (Figure 1a). As a consequence of these features, diverse regional climates characterize the country (Beck *et al.*, 2018) and therefore, socio-economic activities are diverse, such as rainfed agriculture in the Pampas region, cattle raising in the Patagonian flat areas, viniculture in subtropical central Andes, energy generation and tourism. In particular, temperature increases would have impacts on agriculture productions such as crop and wheat (Bettolli *et al.*, 2009; Rolla *et al.*, 2018), on livestock (Rolla *et al.*, 2019) and in wine-growing regions (Cabré and Nuñez, 2020). A warming scenario will also affect snowpack melting over the subtropical central Andes, causing severe restrictions in water availability for human consumption and hydro-electric power generation (Zazulie *et al.*, 2017). Climate change is then expected to differently impact over these regions due to regional feedbacks and responses to changes in large-scale forcings. In this context, providing climate information at regional or local scale is a key element for designing and implementing efficient strategies for adaptation to climate change and its extremes. The projection of the future climate relies on the simulations by global climate models (GCMs), which describe large-scale climate characteristics and the possible evolution of future climate in response to increases in greenhouse gases. Nowadays, the spatial resolution of the GCMs is not enough to generate information on a regional/local scale with the quality required by the application models used for decision making due to subgrid scale processes not explicitly represented. Downscaling then arises as a bridge, trying to reduce the gap between the coarse-resolution output from GCMs and the local-scale climate information (Ambrizzi *et al.*, 2019). In particular, empirical statistical downscaling (ESD) techniques consist of generating climate information at local scale through empirical relationships between large-scale and smaller scale variables (Maraun *et al.*, 2010).

A variety of statistical methods—with different complexity—have been employed for statistical downscaling, varying from traditional and well known techniques such as linear regression, analogs, weather generators, weather typing and bias adjustment methods (D'onofrio *et al.*, 2010; Asong *et al.*, 2016; Bettolli and Penalba, 2018; Gutiérrez *et al.*, 2019; Casanueva *et al.*, 2020; Fan *et al.*, 2021; Olmo and Bettolli, 2021) to more complex machine learning methods including artificial neural networks, support vector machines,

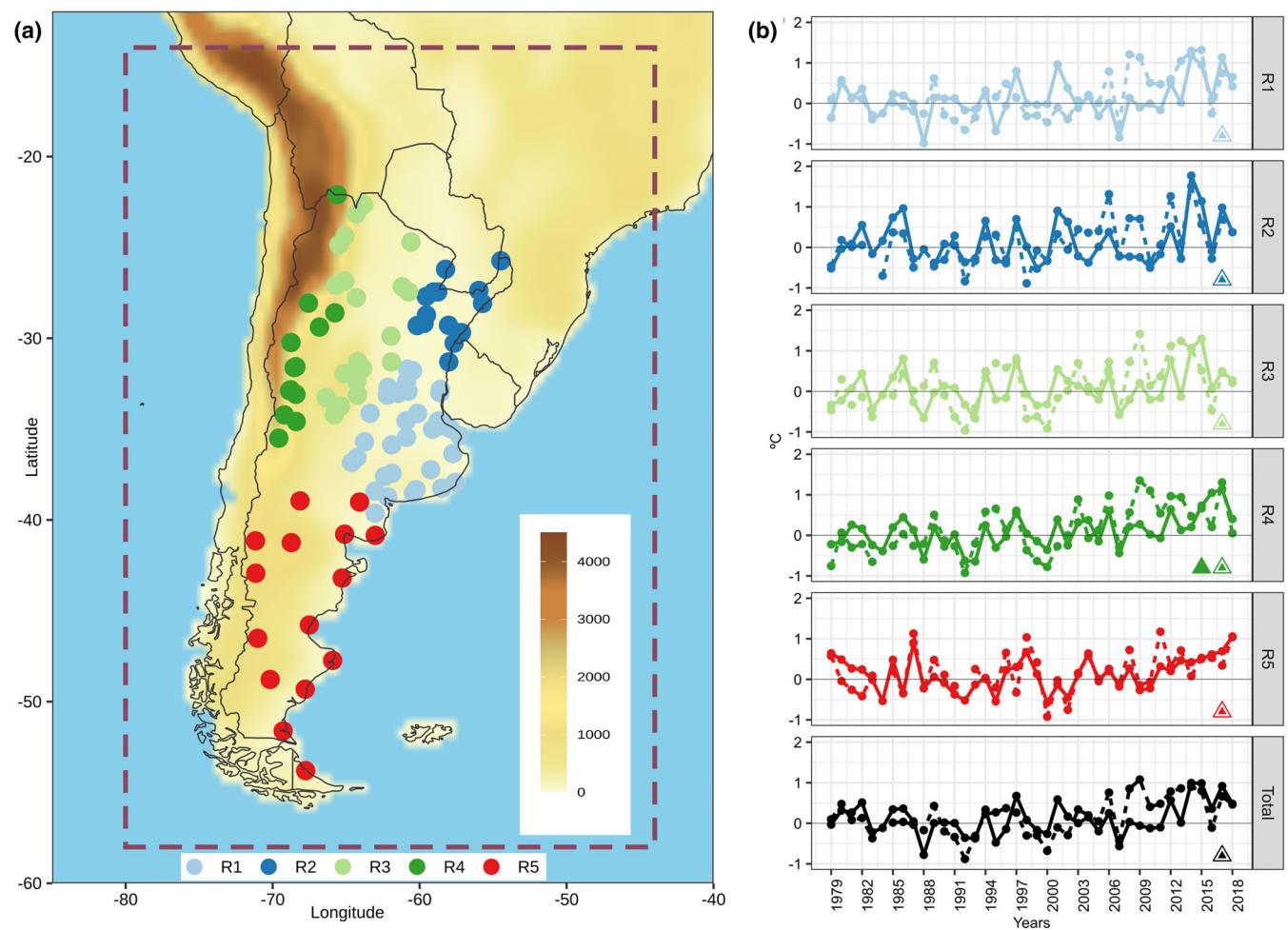


FIGURE 1 (a) Meteorological stations used during the period 1979–2018 regionalized by colours and numbered R1–R5. The purple box shows the domain considered for the predictor sets. (b) Annual time series of Tx (dashed line) and Tn (solid line) anomalies with respect to the period 1981–2010 averaged for the regions shown in (a). Triangles indicate significant linear trends using *t*-student with 95% level of significance

random forest, wavelet-based methods, among others (Sachindra *et al.*, 2013; MoradiKhaneghahi *et al.*, 2019; Kumar *et al.*, 2021; Polasky *et al.*, 2021; Quesada-Chacón *et al.*, 2021; Sun and Lan, 2021; Hernanz *et al.*, 2021a). In the case of temperatures, promising new machine learning techniques can be found in the recent literature. For instance, Vaughan *et al.* (2022) presented a novel model to downscale daily maximum temperature for multiple sites using convolutional conditional neural processes in Europe; Wang *et al.* (2021) analysed the Super Resolution Deep Residual Network algorithm for downscaling daily maximum and minimum temperature in China based on an advanced deep convolutional neural network with residual blocks and batch normalizations; and Sachindra and Kanae (2019) used parallel multipopulation genetic programming algorithm to downscale maximum and minimum temperatures in Japan. However, one of the disadvantages of the machine-learning-based methods is their interpretability, due to their black-box nature,

where the relationships between predictand and predictors remains hidden (Sachindra and Kanae, 2019; Baño-Medina *et al.*, 2020; Busuioc, 2022). Also, these methods usually require a large amount of good quality information to accurately reflect the dynamics of a target variable and thus become computationally expensive and difficult to implement under scarce resources (Jeong *et al.*, 2012). Some recent studies pointed out that, under certain conditions, machine learning methods provide limited improvement over the traditional ones like linear regression models (Gaitan *et al.*, 2014; Huth *et al.*, 2015; Baño-Medina *et al.*, 2020) while others showed their outperformance (Sun and Lan, 2021; Wang *et al.*, 2021; Vaughan *et al.*, 2022). Therefore, there is not sufficient consensus that one method is the most accurate to simulate a specific variable since the predictive skill of the techniques often vary depending on the region and assessed aspect.

The development of any ESD model—no matter how simple or complex the statistical method is—requires

making several decisions along the entire process that may constrain the final results. In the perfect prognosis approach, observed large-scale atmospheric predictor variables are used to estimate the local predictand variable of interest (Maraun *et al.*, 2010). Reanalyses are then used as pseudo-observations of large-scale variables to construct the models. In the ESD literature, the perfect assumption is not usually tested and the sensitivity of downscaling performance to the reanalysis choice is not considered (Horton, 2021). In regions such as southern South America, where some areas have sparse surface observations and poor temporal and spatial coverage of radio-sounding data, the data assimilation process may be affected and, in consequence, the quality of the reanalysis product. Added to this, the presence of the Andes mountain range as a large orographic barrier presents an additional challenge to the reanalysis performance (Balmaceda-Huarte *et al.*, 2021). In this regard, just a few studies have addressed the sensitivity of downscaling results to the reanalysis choice and most of them were focused on precipitation over Canada, India, France, Philippines and Switzerland (Koukidis and Berg, 2009; Kannan *et al.*, 2014; Dayon *et al.*, 2015; Manzanas *et al.*, 2015; Horton and Brönnimann, 2019; Horton, 2021), showing that the disagreement among reanalyses could introduce uncertainties in the downscaled results.

Furthermore, the selection of informative predictor variables and their area of influence depend on the region, season and predictand variable (Huth, 1999; 2004; Cavazos and Hewitson, 2005; Sachindra and Kanae, 2019; Araya-Osses *et al.*, 2020) in such a way that they need to account for a large fraction of local variability on all temporal scales under study (Maraun and Widmann, 2018). In addition to the selection of suitable predictors, the way they are incorporated into the model, what we call model configuration, is not a straightforward issue (Sehgal *et al.*, 2018). The statistical techniques and their configurations are also key elements for successful local climate simulations (Maraun *et al.*, 2019). Atmospheric predictors, and the physical processes they represent, can generally be incorporated into the model using raw data at different grid points (pointwise predictors; Baño-Medina *et al.*, 2020), by performing reduction techniques (such as empirical orthogonal functions; Gutiérrez *et al.*, 2019), by using indices that characterize specific phenomena or processes (for instance, indices representing low-level jets or indices representing instability; Quesada-Chacón *et al.*, 2021) or that provide information about the large and medium scale climatic forcings (such as ENSO; Sulca *et al.*, 2021). Also, combinations of these approaches are found in the literature (Olmo *et al.*, 2022). In recent years, new approaches have emerged as an alternative to efficiently handle high-dimensional data and generate predictions by making use

of machine learning algorithms that optimally extract the main features, thus overcoming the loss of information (Sun and Lan, 2021). On this subject, Panda *et al.* (2022) considered an alternative procedure to minimize the dimension of the predictor variables when downscaling monthly precipitation in India. Authors analysed the Multivariate Adaptive Regression Spline downscaling model using the representative grid location approach (Sachindra and Kanae, 2019) and found better results than when applying principal component analysis.

In addition, the statistical relationship between predictors and predictands should hold for the future (stationarity assumption), especially when used in climate change applications (Dixon *et al.*, 2016; Maraun and Widmann, 2018; Hernanz *et al.*, 2021b). In this sense, the assessment of the ESD model's extrapolation capability becomes a necessary task to justify their potential use for climate change projections (Baño-Medina *et al.*, 2020). Different approaches were proposed to provide insight into the ESD models transferability, such as training and evaluation on contrasting climate conditions (cold vs. warm, dry vs. wet or ENSO vs. non-ENSO years) (Liu *et al.*, 2015; Baño-Medina *et al.*, 2020; Olmo and Bettolli, 2021; Busuioc, 2022); or through pseudo-reality experiments in which ESD models are calibrated using GCMs large-scale predictors and surface variables from regional climate models (RCMs) as "observations" (Vrac *et al.*, 2007; Hernanz *et al.*, 2021b). However, the ESD models' transferability is still a challenging research gap.

Despite the considerable progress in the statistical downscaling research field around the world, only few studies have addressed the development of ESD models to downscale daily maximum and minimum temperatures in Argentina. Verdin *et al.* (2018, 2019) evaluated stochastic weather generators to simulate daily precipitation and daily maximum and minimum temperatures in the Salado river basin (in the Pampas region) for seasonal to multidecadal simulations and Bettolli and Penalba (2018) tested the analog method to downscale these variables over southern La Plata basin. The lack of studies that address a comprehensive assessment of statistical downscaling for simulating daily temperatures over the region is clear.

In this context, the main objective of the present work is to evaluate the capability of a suite of ESD models to simulate daily maximum and minimum temperatures in the different climatic regions of Argentina with special focus put on key issues of the perfect prognosis approach. This assessment will address the sensitivity to the reanalysis choice by comparing the ESD models driven by three different reanalysis datasets; the selection of suitable predictors and the way they are integrated in the models; and the extrapolation capability of the ESD

TABLE 1 Reanalyses considered in this study

Labels	Dataset	Horizontal resolution	Reference
ERA	ERA-Interim	~0.75°	Dee <i>et al.</i> (2011)
JRA	JRA-55	~0.56°	Kobayashi <i>et al.</i> (2015)
NCEP	NCEP-R1	2.5°	Kalnay <i>et al.</i> (1996) and Kistler <i>et al.</i> (2001)

models for simulating daily temperatures under unprecedented warm conditions. Three traditional statistical techniques (analogue method, multiple linear regression and artificial neural networks) are used as benchmark methods to set a reference for future comparisons with the latest cutting-edge techniques.

2 | DATA

2.1 | Predictands: Observations

Daily maximum (T_x) and minimum (T_n) temperatures from 101 stations of Argentina in the period 1979–2018 were used as predictands. The location of these stations is presented in Figure 1a. Data were provided by the National Weather Service of Argentina and quality controlled following the recommendations of Penalba *et al.* (2004) and Rusticucci and Barrucand (2001). The selected stations presented less than 10% of missing information in the total period.

Considering the extension of Argentina, its orography and the variety of its climates, station data were divided into five sub-regions (Figure 1a) in order to better analyse the local characteristics. The regions considered were the Pampas region (R1), northeast Argentina (R2), central Argentina (R3), northwest Argentina (R4) and Argentinian Patagonia (R5) (Figure 1a) based on Beck *et al.* (2018).

2.2 | Predictors: Reanalysis

Empirical statistical downscaling (ESD) under the perfect prognosis (PP) approach was carried out in this study, using predictand data from meteorological stations and large-scale predictors from reanalysis. In order to take into account the uncertainties in the predictor data associated with the reanalysis choice, three different reanalyses were considered for this study (see Table 1).

Daily reanalysis fields were used as large-scale predictors in the period 1979–2018. The variables considered as potential predictors for the different ESD models were meridional wind at 700, 850 and 1,000 hPa (v_{700} , v_{850} , v_{1000}), zonal wind at 700 and 1,000 hPa (u_{700} , u_{1000}), air temperature and specific humidity at 850 hPa (T_{850}

and q_{850} , respectively) and sea level pressure (slp). The choice of the potential predictor variables was made based on the main atmospheric features and properties associated with surface temperatures over the region (Barros *et al.*, 2002; Penalba *et al.*, 2013) and also on the results of other statistical downscaling applications over the region (Bettolli and Penalba, 2018; Araya-Osset *et al.*, 2020). In this sense, the large-scale variables used in this study considered circulation, temperature and moisture information at different levels trying to include signal-carrying predictors linked to climate change (Wilby *et al.*, 1998; Huth, 2004; Gutiérrez *et al.*, 2019).

The atmospheric domain extends between 14°–58°S and 44°–80°W (Figure 1a). This domain encloses Argentina, part of the Pacific and Atlantic oceans and the Andes mountain range which exert a strong influence in determining regional features of atmospheric circulation such as the low level jet of South America, the exchange of air masses and the propagation of synoptic-scale disturbances (Seluchi and Marengo, 2000; Barros *et al.*, 2002; Cavalcanti, 2012). Due to the different native resolution of the datasets (Table 1), all predictor data were regridded to a common 2° grid using bilinear interpolation. This resolution is a good compromise among the different reanalyses native resolutions and allows the extrapolation of this ESD experiment to coarse resolution GCMs in a future step.

3 | METHODOLOGY

3.1 | ESD methods

Three different ESD techniques were analysed: analogs (AN), generalized linear models (GLM) and artificial neural networks (ANN). For these techniques, a variety of configurations and predictors combinations were tested considering spatial-wise or point-wise predictors and combinations of them (see Table 2). In the case of the models with spatial predictors (S), empirical orthogonal functions (EOFs) were applied to the set of predictors over the whole domain to reduce their dimensionality and co-linearity. In this regard, principal component analysis (PCA) was carried out in S mode (Huth, 1996). For this procedure, an input matrix \mathbf{X} with $i \times v$

TABLE 2 Predictors and configurations considered for the downscaling methods

Predictors	Configuration	Label
P1: T850 and slp	S (slp) + L (T850)	P1.LS
P2: T850, q850 and slp	S (T850 + slp) + L (q850)	P2.LS.a
	S (slp) + L (T850 + q850)	P2.LS.b
P3: u1000, v1000, T850 and q850	S (u1000 + v1000 + T850) + L (q850)	P3.LS
	S	P3.S
	L	P3.L

Note: L(S) indicates the (spatial) local configuration of the predictor.

dimensions was constructed considering all predictors combined, that is, i time steps and v predictor variables ($v = \text{number of atmospheric variables} \times \text{number of grid points} \times \text{number of pressure levels}$) (Brands *et al.*, 2011). After standardizing the predictor variables by column (\mathbf{X}') and calculating \mathbf{L} —a square matrix with the eigenvectors (EOFs) of the covariance matrix as columns—the matrix of principal components $\mathbf{C} = \mathbf{X}'\mathbf{L}$ was obtained. In \mathbf{C} , the principal components (PCs) are arranged in columns and are ordered in descending order of their explained variance. For this work, only the first k PCs accounting for at least the 90% of the explained variance were retained and used to train the ESD models. When applying the ESD model in a new set of predictor data (\mathbf{A}), the new matrix of PCs (\mathbf{C}^*) was achieved by projecting the k EOFs retained from the training phase (\mathbf{L}_k) on the standardized matrix \mathbf{A}' , following the matrix multiplication $\mathbf{C}^* = \mathbf{A}'\mathbf{L}$. Then, \mathbf{C}^* was used as input data of the ESD model in the application phase.

In the case of point-wise predictors, the four nearest grid points to the target station point (L) were considered. Tests of sensitivity were performed regarding the number of grid-boxes closest to the target location (using 4 and 16 grid points). Four grid points resulted to be the most suitable number of points to cover a moderate area and at the same time restrict the information to local processes. Additionally, different combinations of local and spatial predictors were tested and are referred to as local-spatial (LS). In all cases, each predictor variable was previously standardized by gridbox.

The AN method (Zorita and von Storch, 1999) is a nonparametric technique that looks for similarities in large-scale predictors and assumes that these similarities are also present in the local surface variables. In this way, for each day, given a set of large-scale predictor variables, the method searches for an analog day and then the surface observations from the corresponding analog day are used as prediction. Various metrics can be used to assess the similarities and determine the analog (Matulla

et al., 2008), being the Euclidean norm the most commonly used. For this study, the closest analog was taken into account based on this distance metric.

The multiple linear regression models relate a set of predictor variables via some linear function to a predictand variable (Maraun and Widmann, 2018). GLM are an extension of linear regression methods that allow predictands to have non-normal probability distributions. In this work, GLM were implemented assuming a Gaussian distribution for the maximum and minimum temperatures, and the large-scale predictors were used to predict the expected values of the local observations (Huth, 1999; Asong *et al.*, 2016).

Artificial neural networks (ANN) are, essentially, nonlinear regression-based models that can model very complex relationships between predictands and predictors (Gardner and Dorling, 1998). An ANN is composed by an input layer (predictors information), an output layer (predictand information) and one or more hidden layers in between, which can be conformed with multiple connected neurons. These connections are characterized by different weights (that in this case, are optimized through a backpropagation learning algorithm) and activation functions (Rumelhart *et al.*, 1986; Huth *et al.*, 2008; Baño-Medina *et al.*, 2020; Wang *et al.*, 2021). For this study a two-layers ANN configuration with 25 and 15 neurons, respectively, was considered, following the experiment performed over southern South America by Olmo and Bettolli (2021). After a sensitivity analysis on the use of different activation functions, the sigmoidal transformation was employed in the neurons of the hidden layers, while a linear transformation was used in the output layer.

Finally, the ESD models tested in this study correspond to possible combinations of an ESD technique (AN, GLM or ANN) with a variety of predictor sets (P_i , $i = 1, \dots, 3$) (Table 2) and predictor configurations (S, L or LS). For instance, the model AN.P1.L corresponds to the analog method considering the predictors of P1 and the local (L) configuration as shown in Table 2. In order to explore the sensitivity to the chosen configurations, three different configurations were considered for P3 (S, L, LS), whereas for the rest of the predictors sets (P1, P2) only LS was evaluated, contemplating two different possibilities of predictors in point-wise: LS.a (only T850) and LS.b (T850 and Q850). The combination of predictor sets and configuration described in Table 2 were tested for AN, GLM and ANN, resulting in 18 ESD models for each extreme temperature.

All ESD models contemplated at least one circulation variable and temperature, in line with previous studies that found better results when the combination of these predictors was considered (Huth, 1999; Timbal and

McAvaney, 2001; Huth, 2004; Brands *et al.*, 2011; Bettolli and Penalba, 2018; Araya-Osset *et al.*, 2020). In addition, the set of predictor variables for each ESD model was built up considering different aspects:

- A predictor set that did not take into account humidity variables (P1).
- A simple predictor set that only considered one variable representative of each aspect: circulation, temperature and humidity (P2).
- A predictor set that did not include sea level pressure as a variable representative of the circulation (P3).

Other sets of predictors and configurations were also tested, but for the sake of conciseness only some of them are presented in this work.

3.2 | Calibration and validation

With the aim of designing the calibration and validation scheme, the spatially averaged annual time series of Tx and Tn anomalies with respect to 1981–2010 were explored (Figure 1b). In general, progressive warming conditions were observed for both variables in all regions where Tx showed significant linear trends. For Tn, the linear trends were not significant but for region R4. In the last years of the period (2015–2018) general positive anomalies were recorded except for the year 2016 where anomalies were near zero.

Taking these features into account, all ESD models were calibrated and validated in the period 1979–2014 (*cross-validated period*, hereafter). To avoid overfitting, a k -fold cross-validation approach was carried out in the calibration procedure with $k = 6$, following Gutiérrez *et al.* (2013). Each fold was arranged with 6 nonconsecutive years that were selected with a 6-year step among them. For example, fold 1 contained the years 1979, 1985, 1991, 1997, 2003, 2007; fold 2 follows the same path, but beginning in 1980; fold 3 begins in 1981, and so on. As discussed by Gutiérrez *et al.* (2013), constructing the folds in this way allows to circumvent the statistical artefacts that could emerge from the trends observed in Figure 1b. Moreover, the sample of years used for each fold considered years with both positive and negative anomalies, which leads to more robust models. For each fold, a calibration/validation process was performed, where in each case, one of the folds was treated as a validation set, and the remaining folds were used to calibrate the ESD model. This procedure was repeated six times (once for each calibration/validation set), and the model output of the six hold out folds were merged to construct a single 36-year cross-validation time

series (1979–2014) of independent data for validation. Further information about the cross-validation scheme can be found in Araya-Osset *et al.* (2020).

As an additional experiment, ESD models were built in the 1979–2014 period based on the configurations considered in the cross-validation procedure and applied in the period 2015–2018 (*warm period*, hereafter). As it was previously mentioned, this period was anomalously warm and allows us to evaluate the ESD models performance under atypical warm conditions which is particularly relevant for climate change studies (Gutiérrez *et al.*, 2013; Baño-Medina *et al.*, 2020). In this sense, this is a challenging period for the ESD models to reproduce as they were calibrated with colder years.

Different metrics were used to assess and compare ESD models to study different aspects of their performance. The daily correspondence and the predictor-predictand relationship was assessed using the Pearson correlation coefficient, which was calculated after removing the annual cycle from the daily series. The day-to-day accuracy of the simulations was quantified in terms of the root mean square error (RMSE). The evaluation metrics were analysed in the *cross-validated* period (1979–2014) for all stations of each region and considering the regional mean in the independent *warm* period 2015–2018.

The daily mean annual cycle Tx and Tn in the *cross-validated* period and in the independent *warm* period was also analysed for the ESD models. The cycles were calculated for each station and then regionally averaged.

The representation of the interannual variability of Tx and Tn by the ESD models was assessed during the completed period of the observations (1979–2018) and measured in terms of the Pearson correlation (of the linearly detrended series). In this analysis the evaluation of linear annual trends was also included. The trends were obtained by regressing the annual mean values against time (Huth *et al.*, 2015).

Special focus was made on the warmest and coldest temperatures in the *cross-validated* period. The 5th and 95th percentiles of the probability distribution (P5th and P95th) were evaluated for Tn and Tx, respectively.

The downscaling methods and the different calculations were implemented by using the R-based climate4R open framework following Bedia *et al.* (2020).

4 | RESULTS

4.1 | General aspects

As an initial inspection of the ESD models performance, the predictor-predictand relationship was analysed in

terms of the Pearson correlation. Tx and Tn daily correlation between the downscaled and observed series (after removing the daily annual cycle) for the different ESD models and for the three reanalyses are presented in Figure 2. Boxplots summarize the results of each ESD model across all stations of each region in the *cross-validated* period (1979–2014) and triangles indicate the regional mean considering the independent warm period (2015–2018). Generally, the AN method presented the lowest correlation values in both variables and exhibited noticeable differences among the models' configurations and sets of predictors. The GLM models performed similarly among the different configurations and exhibited high and mainly constant correlation values near 0.8 in all regions. Minor differences were also observed among the multiple ANN models regarding different configuration and predictor sets. For this method, the correlation values were similar in mostly all models and in general slightly inferior than in linear models. In this sense, the

strength of the predictor–predictand link was better captured by the regression-based models (ANN and GLM) than by the analog method for both Tx and Tn, in agreement with results from Huth *et al.* (2015) and Gutiérrez *et al.* (2019) for different regions in Europe. However, when comparing neural networks (ANN) and linear models (GLM) no method outperformed the other in this aspect, except for the ANN.P3.L that tended to show fairly higher correlation values.

Regional differences were also appreciable. In north and central Argentina (R1, R2 and R3) the ESD simulations presented more agreement with the observations than in northwest Argentina (R4) and Argentinian Patagonia (R5), where lower correlation values were observed, especially for Tn. In addition, Tn presented more dispersion among stations, clearly depicted by the larger boxes in R3, R4 and R5 (Figure 2).

When considering predictor data from the different reanalyses, ESD models showed differences among each

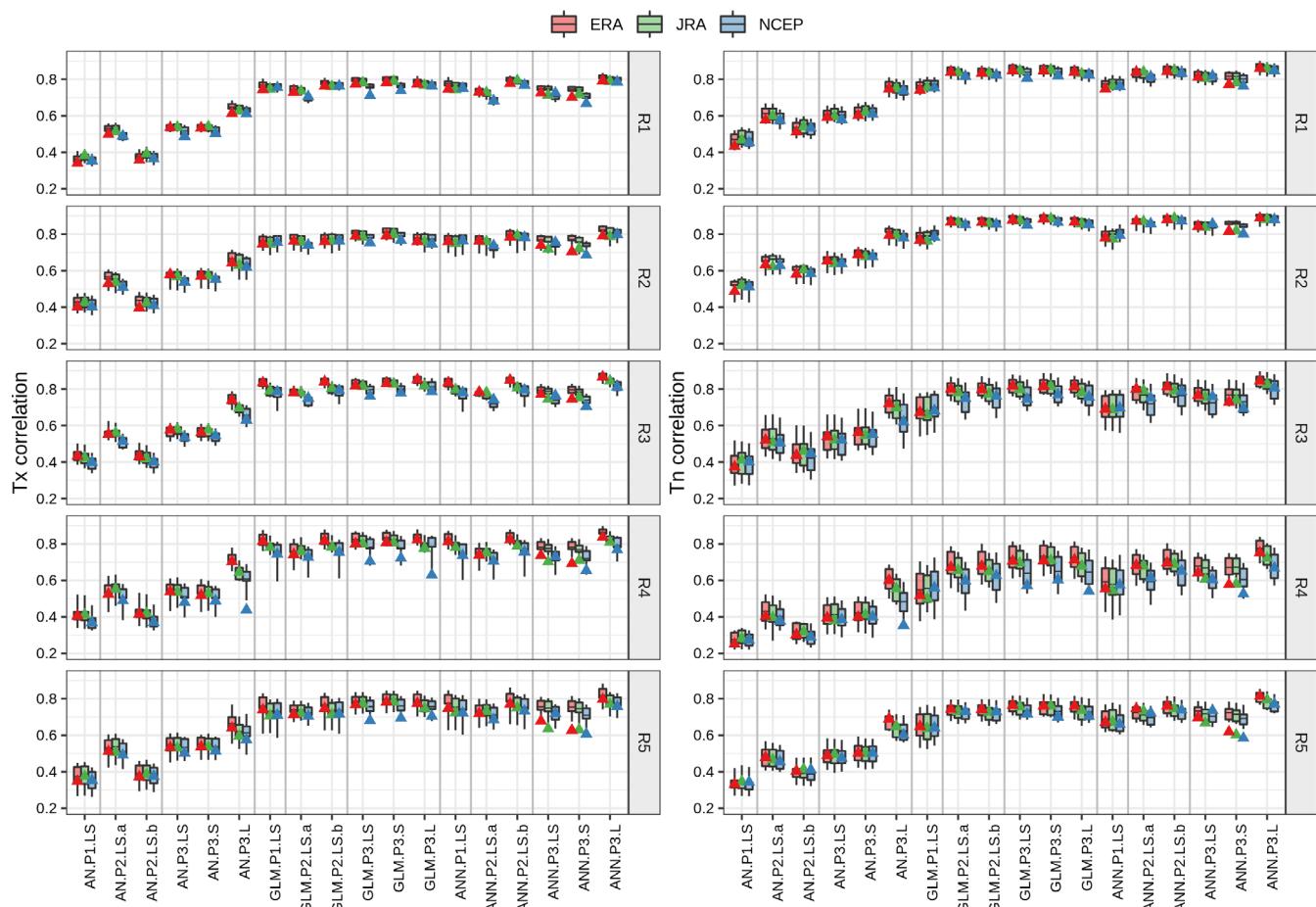


FIGURE 2 Box and whisker plots of daily Pearson correlation coefficient between the observed and downscaled data for Tx (left) and Tn (right) for each ESD model and reanalysis predictor data (box plot colour) in the cross-validation period 1979–2014. The box and whisker plot displays the results across all stations in each region. On each box, the central mark corresponds to the mean, and the bottom and top edges to the 25th and 75th percentiles. The whiskers extend to the 5th and 95th percentiles. Triangles indicate the station mean of each region considering the independent warm period 2015–2018

other depending on the configurations and regions. In general, it could be noticed that ERA-driven ESD models systematically presented the highest correlation values, while the lowest values were observed when considering NCEP predictor data. This was particularly evident in models with only local predictor configurations (L) for the AN and in the majority of GLM and ANN models for both Tx and Tn. In R4, almost all ESD models driven by ERA stood out from the rest with greater agreement with the observations. Note that R4 is a region with complex topography, where the Andes mountain range reaches heights greater than 6,000 m (Figure 1) and therefore orographic-derived differences among reanalyses may induce some discrepancies in downscaled temperatures.

Particularly for the AN models, the local (L) configuration highlights from the rest of the models with higher correlations and showing a stronger link with the predictor data in both Tx and Tn (Figure 2). This agrees with Timbal and McAvaney (2001) and Ribalaygua *et al.* (2013), who argued that raw data over small domains consider the full range of data variability and yield better results of the AN method than when using PCs. It should be stressed, however, that GCMs simulations of point predictor values (L) are not as reliable as field variables in larger domains (Huth, 1999; Ribalaygua *et al.*, 2013), which constitutes a particularly important consideration for climate change applications. The AN.P1.LS and AN.P2.LS.b models presented the lowest skills in all regions and for both variables. Note that the model AN.P2.LS.b considered the air temperature at 850 hPa as local predictor when compared with their version LS.a, indicating that the large-scale structure of T850 influences the determination of surface extreme temperatures when using the analog model.

No remarkable differences were observed for Tx among the variety of GLM model configurations (Figure 2). For Tn, GLM.P1.LS exhibited systematic lowest correlation values in all regions. The latter was also observed for the models AN.P1.LS and ANN.P1.LS for Tn and could be related with the fact that the P1 set did not include specific humidity as a predictor (see Table 2). Previous works already assessed the predictive skill of this variable at the low levels of the atmosphere and found that the accuracy of the analog models to downscale Tn is enhanced if humidity predictors are added to temperature and circulation predictors (Timbal *et al.*, 2003; Brands *et al.*, 2011; Bettolli and Penalba, 2018). In the case of Tx, AN.P1.LS also exhibited poor day-to-day agreement with the observations, whereas GLM.P1.LS and ANN.P1.LS presented similar skills to the rest of the models in their families. Particularly in the GLM models for Tx, no set of predictors was distinguished with a higher correlation value. This could be associated with

the strong link between Tx and T850 (a common predictor to all models), which seems not to be affected by the different configurations in the GLM models.

Similar to the analog models, the ANN method presented the highest correlation values when only local predictors were considered (Figure 2). For both variables, the model ANN.P3.L presented the best performance of all ANN models in terms of day-to-day agreement, being this more clear in the southern regions of Argentina (R4 and R5). The ANN.P2.LS.b model also distinguished from the rest of the ANN family showing higher skills representing Tx and Tn and differentiating from the LS.a configuration, which in comparison, presented lower correlation values.

Considering the *warm* period (triangles in Figure 2), the performance of the ESD models was comparable to the *cross-validated* period for both variables, thereby conserving the predictor–predictand link. However, some differences were observed for Tx when using NCEP predictor data. In general, NCEP-driven ESD models presented noticeable lower daily correlation values in R4 in this warm period when compared with the cross-validated period (blue triangles outside the blue boxes in Figure 2). This was also observed over the same region for Tn in some ESD models. Common to all regions, the GLM models fed with the predictor set P3 also presented low correlation values, for Tx. For both JRA and ERA predictor data, GLM and AN models preserved strength of the day-to-day agreement in the warm period for all regions and extreme temperatures. In the case of the ANN method, models which include the predictor set P3 with spatial configuration (ANN.P3.S) presented systematic lower correlation in the warm period. This was observed in all reanalyses and in both temperatures. This means that when only spatial data was considered, the ANN models showed higher difficulties to reproduce the predictor–predictand link in warm conditions.

In the following step, the daily RMSE were analysed for the different ESD models as shown in Figure 3. Overall, Tx simulations presented larger errors than Tn simulations in all regions. Higher RMSE values were observed in R3, R4 and R5 for both variables.

The AN models presented the greatest errors for Tx and Tn, especially in the models AN.P1.LS and AN.P2.LS.b in coincidence with the lowest correlations displayed in Figure 2. The AN with local predictors configurations (AN.P3.L) showed the lowest RMSE among the AN family over all regions, especially when driven by ERA. In this regard, the AN models appeared to be more accurate when local predictors are considered, in agreement with results found by Gutiérrez *et al.* (2013). It is important to bear in mind that this kind of setup of the AN might yield to a misrepresentation of the spatial

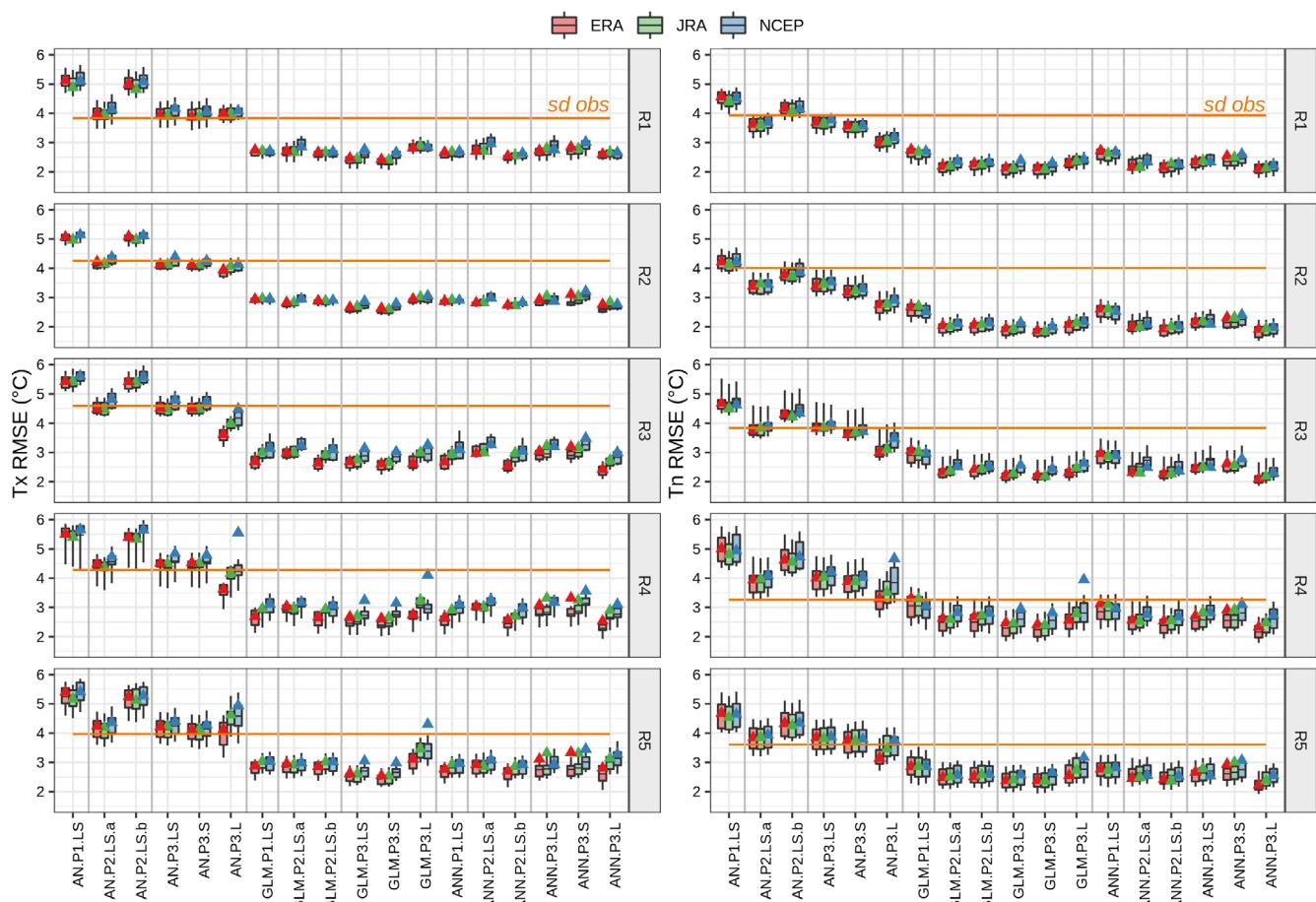


FIGURE 3 Similar to Figure 2 but for daily root mean square error (RMSE) in °C. The horizontal orange line (labelled sd obs) corresponds to the observed daily standard deviation (after removing the annual daily cycle) averaged across each region in the cross-validated period

structure of the predictand variables. Similarly, in the ANN models, the lowest errors in all regions and both variables were observed when only local predictors were considered (ANN.P3.L). The model ANN.P3.L exhibited an improvement over the rest of the configuration fed with P3, more appreciated in Tn. For the set of predictors P2, the model ANN.P2.LS.b overly presented a smaller RMSE compared to ANN.P2.LS.a for Tx, suggesting that in these models, the RMSE decreases when the near grid points of T850 predictor were considered.

Opposite to AN and ANN, in the case of the GLM family, GLM.P3.L generally exhibited larger errors than GLM.P3.S for Tn and Tx, although these differences with the other configurations resulted smaller than the ones detected in the variety of analogs or neural networks models. This was more evident in northwest Argentina (R4) and in Patagonia (R5). For Tx, when the configurations included the spatial structures of the predictor variables (LS or S), the errors tended to be slightly lower for predictor set P3 in central and northeast Argentina (R1 and R2) and Patagonia (R5).

For Tn, the ESD model that did not include humidity in the predictor set (GLM.P1.LS, AN.P1.LS, ANN.P1.LS) presented systematic high RMSE in all regions, reinforcing the need for humidity predictor variables to enhance the accuracy of simulations.

Differences among all ESD models were detected regarding the reanalysis choice. In general terms, down-scaled temperatures from ERA presented the best performance over the other reanalyses as depicted by their lowest errors, while ESD models driven by NCEP presented the largest RMSE. This was specially noticed in west-central Argentina (R3 and R4) and in Patagonia (R5) where ERA considerably reduced model errors.

In the *warm* period, ESD models using NCEP predictor data stranded out with higher errors than when considering the *cross-validated* period. This was particularly observed for the GLM models in regions R4 and R5 (Figure 3), in line with the lower correlation values observed in Figure 2.

RMSE values were smaller than the observed standard deviation (orange line in Figure 3) of Tx and Tn in

all regression-based models (GLM and ANN) in all regions and reanalysis choice. This indicates that the relative difference of the simulated and observed temperatures was smaller than their own natural daily variability, which could be considered as a benchmark measure in the models' capability. This was not the case of the AN models, that kept mostly around the observed standard deviation for Tx. For Tn instead, some improvements regarding this aspect were detected specially in R1 and R2.

The annual biases are shown in Figure S1, Supporting Information. During the *cross-validated* period, the biases of the AN models were all positive for both Tx and Tn over all Argentina, with the exception of AN.P3.L in R3 and R4, that exhibited values near zero. However, the maximum biases did not surpass 0.5°C. GLM models by construction preserve the observed means, and therefore, biases were around zero. Also, errors from ANN models were centred around zero for Tx and Tn, except in the Patagonia region (R5), where the models presented slightly positive biases. For the *warm* period instead, all ESD models performed similarly with mostly negative biases. In this regard, the ESD models improved sensibly the raw outputs from reanalyses, which showed strong biases in Argentina (from -3 to 5°C for Tn and from -6°C to -2°C for Tx; Balmaceda-Huarte *et al.*, 2021), indicating a clear added value of downscaling in this aspect.

The previous analyses showed that, depending on the ESD method, some configurations and predictor sets presented potential skills and may be more appropriate for downscaling daily extreme temperatures over Argentina. Taking this into account, the ESD models with the poorest performance will not be considered for future analysis: AN.P1.LS, AN.P2.LS.b, GLM.P1.LS, GLM.P3.L, ANN.P1.LS, ANN.P2.LS.a. Note that this could vary depending on the region, but to preserve the physical consistency along the domain, the same models are retained for all the regions. Furthermore, although differences in model performances were observed regarding the reanalysis choice, these were not as large as those due to the choice of the statistical family and model structure. Therefore, models with ERA predictor data will only be shown in the following results.

4.2 | Annual cycle, interannual variability and long-term changes

An additional temporal aspect to evaluate is the reproduction of the annual cycle, which allows us to explore the performance of the ESD models in the different seasons of the year. Recall that the ESD methods were not

calibrated by season, and therefore this systematic variation is implicitly included in the models via the chosen predictors. The spatially averaged daily annual cycle for each region and variable are presented in Figure 4 for the *cross-validated* period (solid lines) and for the independent warm period (dashed lines). For clarity, the ensemble mean and spread of the ESD models for each family are shown (in green for the AN, violet for the GLM and orange for ANN). Considering the *cross-validated* period, in general ESD models well represented the time evolution of Tn and Tx in all regions. In this period, the major differences with observations were exhibited in Tx (Figure 4), where all the ensembles presented a common feature of overestimations of temperature values during the austral winter in all regions. This feature was particularly evident for the analog models which also presented a wider dispersion when compared with the GLM and ANN. Whereas during summer, the ESD model ensembles tended to show underestimations of Tx over most of the country, except for central and northwest Argentina (R3 and R4) where mean maximum temperatures were fairly well reproduced. Over these two regions and in R5, the ESD model ensembles tended to show slight cold biases during spring months (September–November) and warm biases during autumn (March–May).

In the case of Tn, the ESD model ensembles could simulate the intra-annual variability although they showed overestimations in the winter season in all regions as for Tx. However, these overestimations were less pronounced than those of the maximum temperature in all cases, indicating that ESD models were able to better reproduce the coldest mean extreme temperatures of the year. Underestimations of the summer Tn were only observed in the Argentinian Patagonia region (R5) and to a lesser extent in northwest Argentina (R4).

In the *warm* period 2015–2018 (dashed lines in Figure 4), the observations showed a general characteristic of higher Tx and Tn compared to the mean annual cycle of the *cross-validated* period (1979–2014) in almost all months of the year. Depending on the region and variable, this warming was more pronounced in different seasons. For Tx, warmer temperatures were mainly observed in summer and spring in all regions (except for R5). While for Tn, higher temperatures were extended throughout the year in northeast and central Argentina (R1, R2 and R3), more exclusive to winter and summer in R4 and R5.

All these features of the warm period were satisfactorily represented by the ESD models but some of them misrepresented in intensity. Again, a better performance was observed in Tn when compared with Tx. In this period, the ESD models generally better captured the winter Tn, while in summer, Tn was underestimated in

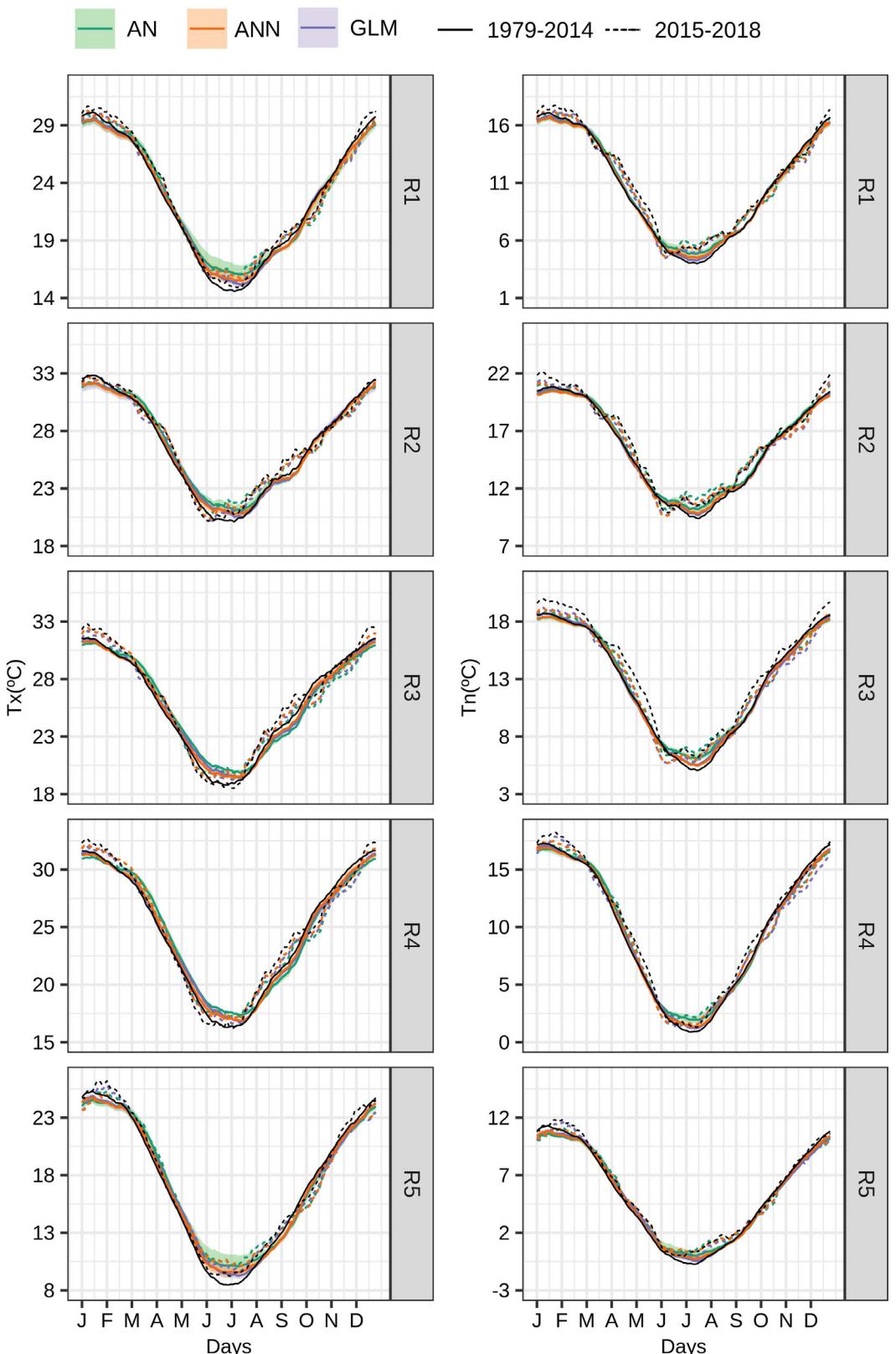


FIGURE 4 Daily annual cycle of T_x (left) and T_n (right). The ESD simulations are exhibited in green for the AN family, violet for the GLM family and orange for the ANN family. For the period 1979–2014, the solid lines indicate the ensemble mean and the shading indicates the range of the ensemble members in each case. In the same colours dashed lines indicate the ensemble mean of each model family corresponding to each averaged in the period 2015–2018. The observations are exhibited in black in solid (1979–2014) and dashed lines (2015–2018). A moving window of 20 days was previously applied in each cycle for smoothness

all regions and this was also extended to autumn in the Pampas region and central-northeast Argentina (R1, R2 and R3). In the case of Tx, all ESD models presented the same difficulties as in the *cross-validated* period, however the underestimations of summer Tx seemed to be slightly higher in this period and common to all regions.

The time evolution of Tx and Tn annual means are shown in Figure 5 for observations and ESD models. A general good agreement between the ensemble mean of simulations and observations was observed, indicating that ESD models were able to reproduce the year-to-year variations but misrepresented their intensity in some particular years. Pearson correlation values between the linearly detrended time series quantified these agreements being higher than 0.73 in all cases and regions except for the GLM ensemble in R4 for Tn where the correlation reached 0.63. In the latter case, the spread of the Tn simulations was also wider than in the other regions. Whereas the Tn interannual variability over R2 seemed to be the best captured not only in terms of correlation values but also in terms of the narrow spread of simulations. When comparing the model families, ANN model ensembles tended to be more skilful in reproducing the interannual correspondence than AN and GLM model ensembles for both predictand variables as depicted by the higher correlation values in all regions.

The long-term changes in temperature that characterized the different regions of the country, positive and significant in all regions for Tx and in R4 for Tn (Figure 1b), were not completely captured by the ESD models, judging from the visual inspection of the series of Figure 5. In these cases, overestimations of temperatures during the first part of the period and the opposite behaviour from approximately year 2000 onwards were observed. However, it is worth noticing that ensembles of ANN and GLM families tended to reproduce the positive sign of the long-term changes though not their intensity. Although not shown, similar results were found when analysing ESD models driven by JRA and NCEP predictor data. As discussed in Huth *et al.* (2015) and Maraun and Widmann (2018), interannual and long-term changes are inherited from reanalyses through the chosen predictors. Therefore, ESD models may be able to capture and reproduce these temporal features if they are correctly simulated by the predictor data and the construction of the ESD model manages to imprint them. In a previous work, Balmaceda-Huarte *et al.* (2021) found that ERA-Interim, JRA and NCEP showed difficulties in representing mean Tx and Tn trends over Argentina, which is in line with the results found here where only the regression-based models were able to reflect to some extent the long-term changes in Tx and Tn.

4.3 | Percentiles

In this section the focus is put on the tails of the distributions and their reproduction by ESD models. The warmest and coldest temperatures of the year were analysed through the summer 95th (P95th) and winter 5th (P5th) percentiles of Tx and Tn, respectively. The spatial distribution of the differences between the downscaled and observed values (bias) are presented in Figure 6a. Very similar performances were found among the different predictor sets of each family (AN, GLM or ANN), hence, for the sake of conciseness, only selected models are exhibited.

Overall, ESD models underestimated the highest Tx in summer (P95th) and overestimated the lowest Tn in winter (P5th) in almost all stations. In agreement with the results of Figure 4, the largest differences were observed for Tx and for the GLM and ANN family. For the AN models, the differences with both observed percentiles were very similar throughout the domain, where most stations exhibited biases lower than 1°C (in absolute value). Particularly, some stations in Argentinian Patagonia (R5) presented slightly higher underestimations of the summer Tx P95th than the rest. This was in accordance with the large dispersion observed in general in this region as displayed in Figures 2 and 3 for the AN models. Unlike the AN, the GLM showed some spatial variations in the percentile biases and exhibited the largest errors for the summer P95th, with differences down to -2.5°C with the observations. Similarly, the ANN models exhibited the largest biases in the 95th percentile of Tx, presenting slightly lower values than linear models, with the exception of the stations near the Andes mountain range (R4). For the winter Tn P5th, the ANN family well-reproduced the intensity of the cold temperatures and showed better results than the GLM models, in most of the stations. Notwithstanding, compared to the ANN models, in the stations near the Andes mountain range (R4) the linear models tended to better capture the Tx P95th and Tn P5.

A deeper inspection of the capacity of the ESD models in representing extreme temperatures as depicted by the annual cycle of the daily percentiles (Figure 6b), revealed seasonal and intrapersonal different behaviours from those observed in summer and winter (Figure 6a). Generally, AN models (green lines) simulated very well the Tx P95th during summer in the regions of northern Argentina (R1–R4) as seen in Figure 6a, satisfactorily capturing the warmest temperatures, although overestimations were exhibited in autumn and winter. This feature was intensified in the west and south of Argentina (R3, R4 and R5). Conversely, the GLM family (violet lines) was not able to satisfactorily capture the intensity of the Tx P95th during summer, but it presented a better performance than the AN models in autumn and winter

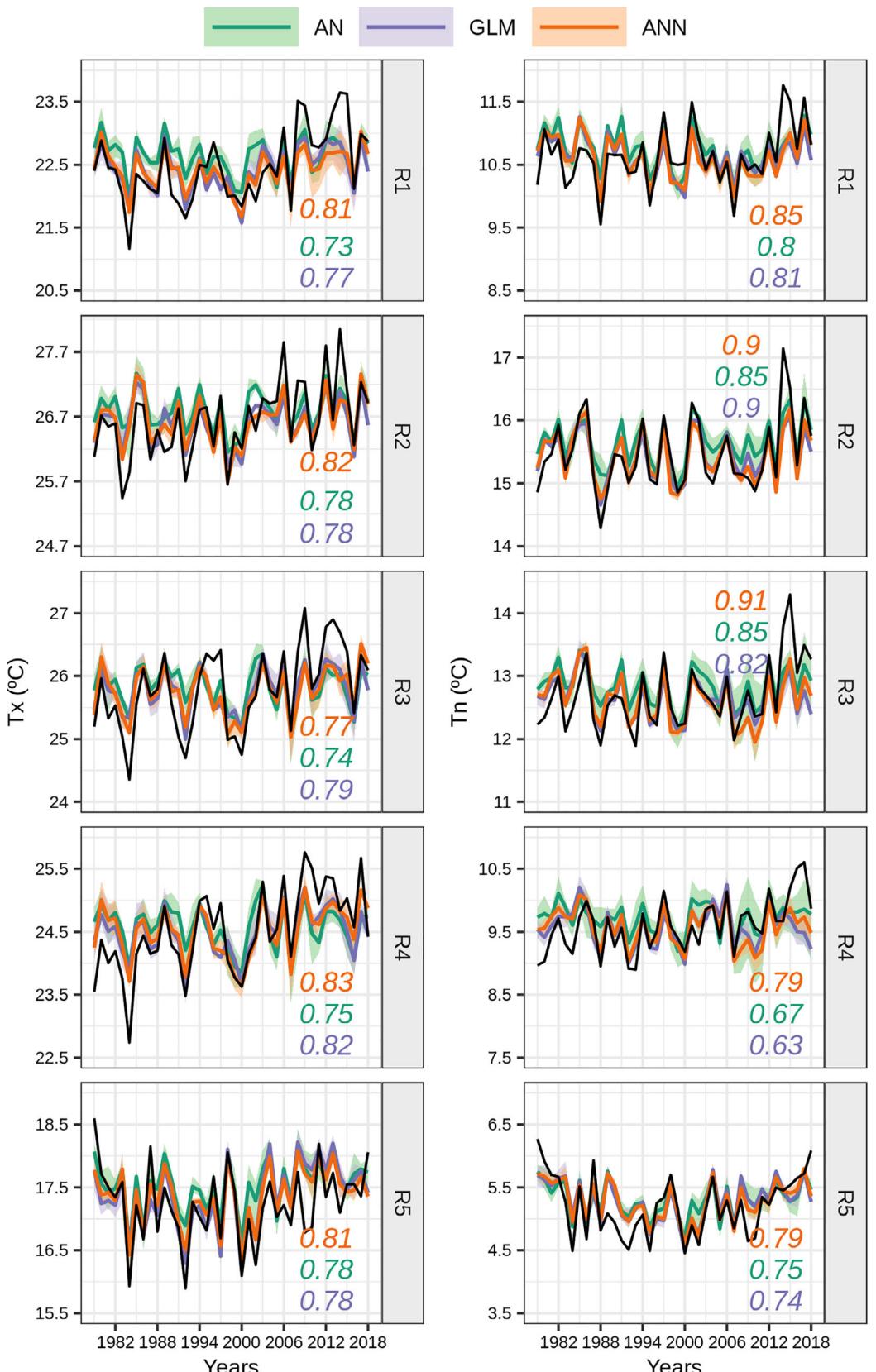


FIGURE 5 Time series of T_x (left) and T_n (right) in the period 1979–2018 for each region (rows) and predictand variable (columns). The ESD simulations are exhibited in green for the AN family, violet for the GLM family and orange for the ANN family: The solid lines indicate the ensemble mean and the shading indicates the range of the ensemble members in each case. Observations are presented with black solid lines. Pearson correlations values between observed temporal series and each family ensemble means are indicated

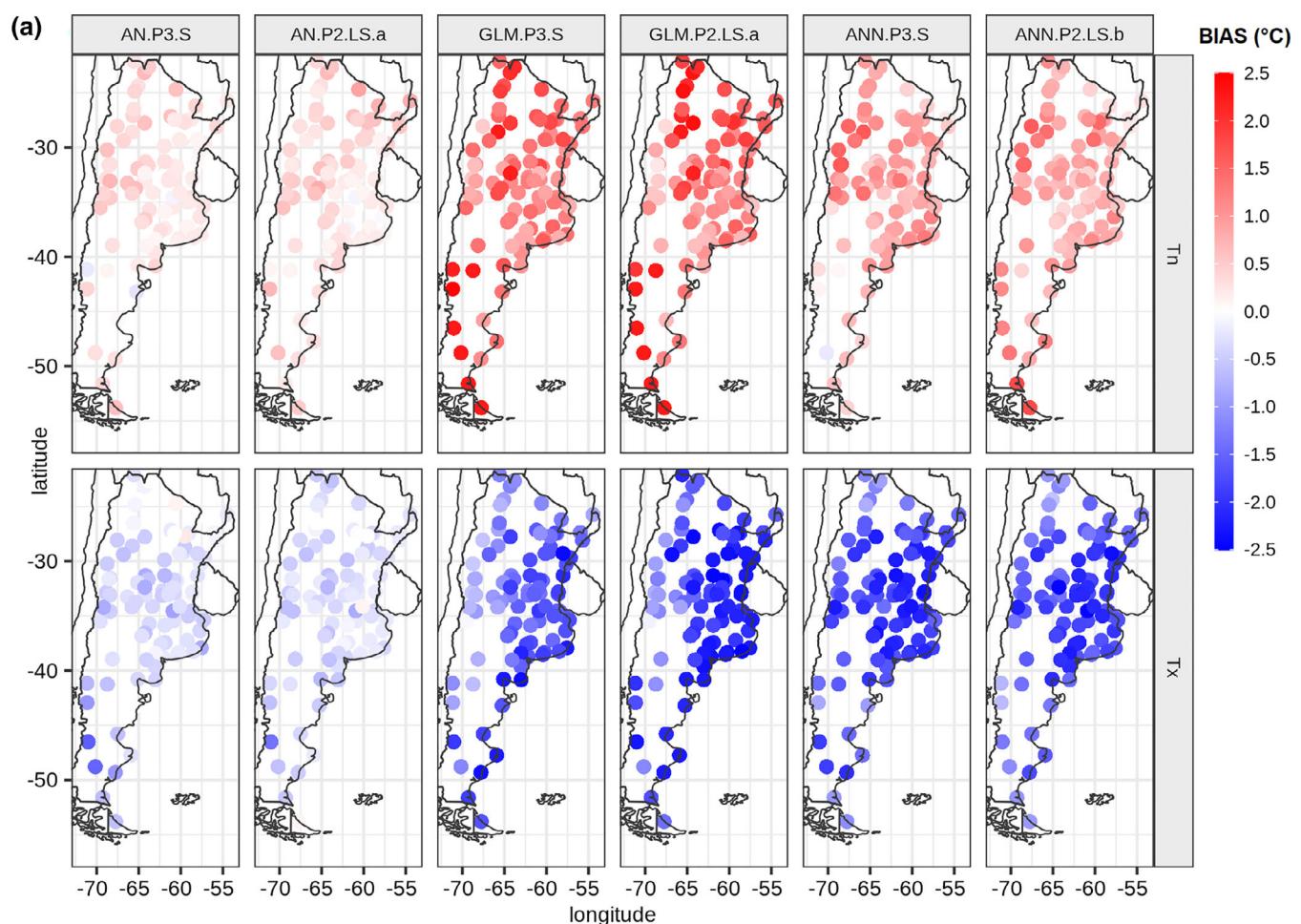


FIGURE 6 (a) Differences between downscaled and observed values of the 95th percentile of Tx (DJF) and 5th percentile of Tn (JJA) for each ESD model considering the *cross-validated* period (1979–2014) in °C. (b) Daily annual cycle of the 95th percentile of Tx and 5th percentile of Tn. The results are shown for AN (green), GLM (violet) and ANN (orange) with predictor sets P2.LS (in light colours) and P3.PC (in dark colours). Observations are exhibited in black

in most regions with the exception of R2. In this region, GLM underestimated P95th over all the year, distinguishing from the other methods (AN and ANN) and observations. Similarly to R2, in R3 GLM models underestimated P95th almost all year as well, although to a lesser extent during autumn and with slight overestimations in some winter months (June and July). Furthermore, the dispersion among the GLM models considerably increased during spring and the beginning of summer, especially over western and central Argentina (R3 and R4) where GLM.P3.S presented values closer to observations (Figure 6b). The ANN models tended to perform similarly to linear models during summer, underestimating the warmest temperatures of the year, in all regions. In contrast, along the rest of the year, neural network models closely followed the observations showing a good performance.

It is worth noting that all ESD models were able to capture the local minimum of Tx P95th during

September (in R1, R2, R3 and R4) although misrepresenting its intensity in the case of the GLM models. This local minimum could be related to the onset of the rainy season over most tropical and subtropical Argentina (north of 40°S). By the end of winter, Tx starts increasing rapidly due to radiative effects. This temperature rise is then interrupted by cloudiness associated with the wet season of the year (roughly between September and April) (Rusticucci and Penalba, 2000). This behaviour is not observed in R5 where its mid-latitude climate regime is generally associated with transient activity throughout the year. It is therefore relevant that the predictor variables were able to contain these control processes of Tx. For instance, although it is not shown for brevity, the AN model that does not include humidity at 850 hPa (AN.P1.LS) was not able to capture this remarkable behaviour in P95th, whereas GLM.P1.LS poorly reproduced it.

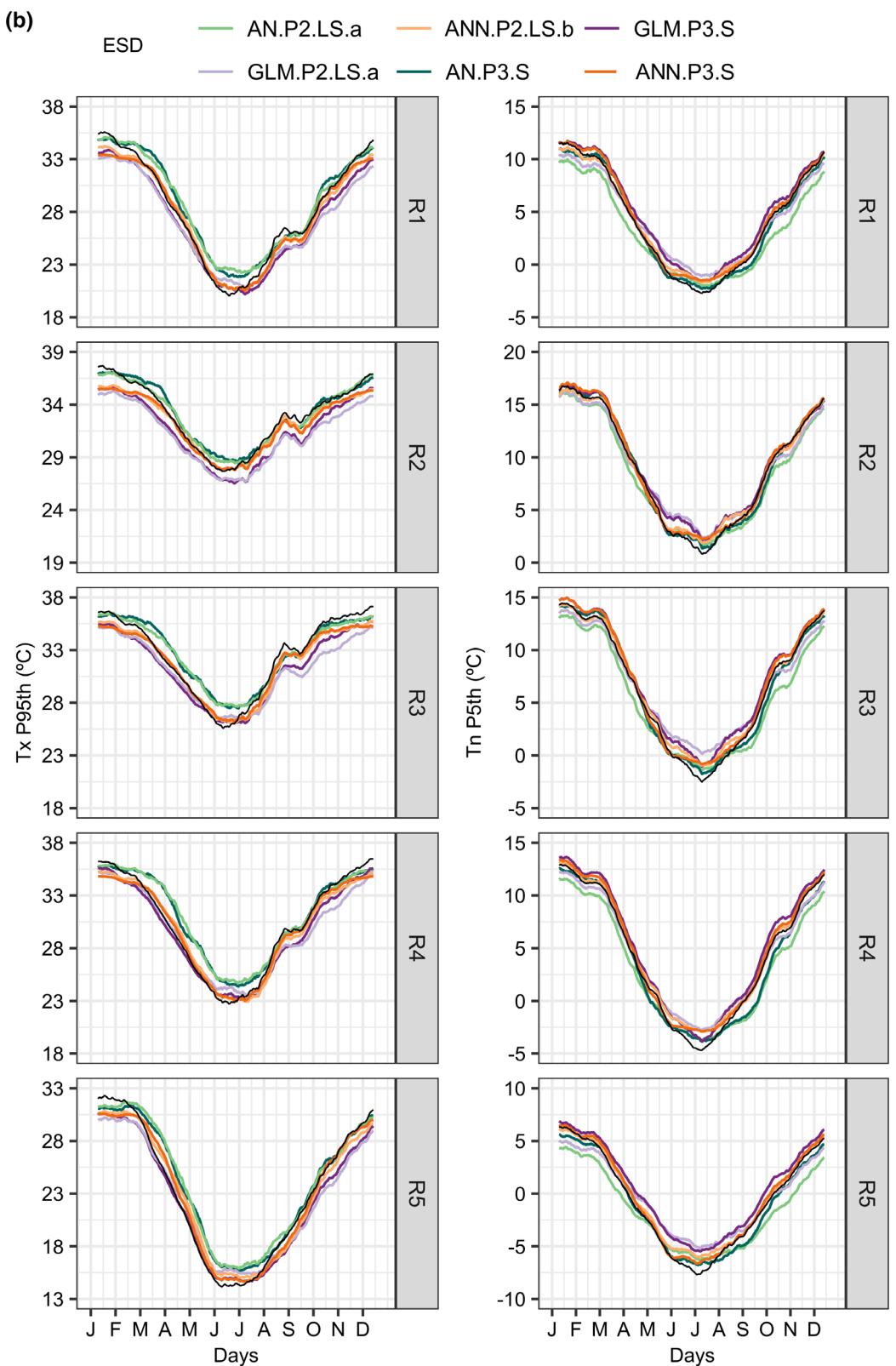


FIGURE 6 (Continued)

Compared to P95th, the annual cycle of the Tn P5th showed less dispersion among ESD models (Figure 6b) in R2, R3 and R4. Overall, models from the ANN family

well captured the annual evolution of the P5th Tn, although they overestimated the Tn P5th in winter. This winter misrepresentation was also observed for the GLM

models, being more evident in the Patagonia region (R5). These overestimations of the linear models were stronger and extended to the end/beginning of autumn/spring. Notwithstanding, in R4 the GLM models were more skilful than the ANN family during winter as observed in Figure 6a. In the case of AN, underestimations were detected during spring over R4 and R5. These underestimations were generalized to all regions in the case of the AN models with P2.LS.a predictor set. It is interesting to highlight that for the case of the GLM and AN models, the ability in reproducing the Tn P5th during summer seemed to be more related to the choice of predictors set than to the model family itself. In all regions, both GLM.P2.LS.a and AN.P2.LS.a exhibited underestimations of the observed P5th while GLM.P3.S and AN.P3.S presented closer values to observations. This behaviour was not observed in the ANN family. Both ANN models were almost indistinguishable and close to the observations during summer in all regions, regardless of the predictor setting.

5 | DISCUSSION AND CONCLUSIONS

Statistical downscaling is a powerful tool that bridges the gap between the global climate model information and what decision makers require (Wilby *et al.*, 2002). However, developing efficient and appropriate empirical statistical downscaling (ESD) models is challenging.

In Argentina, limited studies explored statistical techniques to downscale maximum and minimum temperatures with concern on climate change. In light of this, in the present study advances have been made in the assessment of ESD models to simulate daily maximum and minimum temperatures in the different climatic regions of Argentina. Three ESD techniques were evaluated: analogs (AN), generalized linear models (GLM) and neural networks (ANN) with focus on key aspects of the perfect prognosis approach. A variety of predictor sets with multiple configurations driven by three different reanalyses (Table 1) were considered. ESD models were cross-validated using folds of nonconsecutive years (1979–2014) and then evaluated in an independent warm period (2015–2018). In this regard, ESD models were challenged to represent atypical warm conditions, assessing in this way their extrapolation capability.

Different aspects of the predictand variable are of interest depending on the specific requirements of the sector where the simulations of the regional climate may be applied (Huth, 2002; Maraun and Widmann, 2018; Sun *et al.*, 2020). Taking into account that no model is expected to succeed in reproducing all regional climate

aspects, the focus of the performance assessment of the ESD models was put on some key marginal aspects (mean values and extreme percentiles) and temporal aspects (day-to-day, intra-annual and inter-annual variability and long-term changes) of Tx and Tn, leaving evaluations of other aspects for future studies.

From these assessments, it became clear that, for each ESD method, the predictor set and model configuration were decisive in the downscaling performance, in line with the results discussed in Huth (2002, 2004) and Maraun *et al.* (2019). It was found that the different predictor structures (point-wise, spatial-wise and combinations of them) and model families introduced the main differences among the ESD models, regardless of the predictand variable, region and reanalysis choice. The difficulties to reproduce the characteristics of Tx rather than the ones of Tn was a common feature of all the ESD models, observed in general in all the aspects studied to a greater or lesser extent. In particular, the AN and ANN models seemed to be more accurate (in terms of the RMSE and day-to-day correlations) when using only local predictors in both Tx and Tn. This was in accordance with results of Gutiérrez *et al.* (2013) and Huth *et al.* (2008) when evaluating analogs and neural networks in Europe, respectively. Differently, the GLM family was more skilful when including spatial predictors structures. On this subject, Baño-Medina *et al.* (2020) found that GLM with predictor data in several grid boxes yield better predictions for daily mean temperature.

The principal component analysis (PCA) is one of the most widely used algorithms to reduce predictor dimensionality in statistical downscaling (Huth *et al.*, 2008; Gutiérrez *et al.*, 2019; Satya Sai *et al.*, 2021). PCA is able to well reduce overfitting and has the advantage of being a noniterative and therefore less time-consuming method (Anowar *et al.*, 2021). This reduction in computational demand is valuable when the downscaling is applied to large ensembles of global climate models (GCMs), which is necessary to account for natural variations and model differences (Benestad *et al.*, 2015). However, some caution must be taken when using PCA since they struggle handling nonlinear data. Moreover, if the selection of the retained PCs in the reduction process is not done carefully, valuable information can be lost. In addition, some studies showed that the use of PCs as inputs can be detrimental for the downscaling model performance (Sachindra *et al.*, 2013; Panda *et al.*, 2022), since PCs coefficients remain as fixed components of the downscaling model and may distort the PCs in model prediction (Sachindra *et al.*, 2013; Sehgal *et al.*, 2018). This strengthens the decision of using both spatial and local predictors configurations in the present work, whereas special attention should be taken in follow-up

applications when extrapolating the statistical models for climate change studies.

When focusing on the predictor variables, commonly in all ESD models, predictor sets constructed with humidity information at the low-level atmosphere reduced models' errors to simulate Tn in agreement with previous discussions of other authors (Timbal *et al.*, 2003; Brands *et al.*, 2011; Bettolli and Penalba, 2018). For Tx instead, the added value of including humidity at low levels was noted mainly in the AN models.

A sensitivity analysis on the reanalysis choice showed that ERA-driven ESD models were the most skilful in representing Tx and Tn, while the models with NCEP predictor data exhibited the lowest correlation values and highest errors, being clearer for the local predictor configuration. The performance of ESD models coupled with JRA depended on the configuration, predictor set and region, but they commonly showed better or similar abilities than NCEP-driven ESD models to reproduce Tx and Tn. Despite this, the differences observed in ESD models due to the reanalysis choice were notably lower than the ones obtained due to changes in the statistical family (AN, GLM or ANN) and model structure. These results were generally detected in all regions and for both variables. Manzanas *et al.* (2015) and Horton and Brönnimann (2019) addressed this topic when downscaling precipitation using GLM and AN, respectively, and found that the models' performance was sensitive to the reanalysis choice differing from the results exposed here for temperature. Recall that the small differences detected in this work correspond to the present climate conditions and these could be considerably amplified in the future when projections are obtained by means of statistical downscaling calibrated with different reanalyses (Manzanas *et al.*, 2015). Moreover, from the perfect prognosis perspective, the predictors are assumed to be perfectly simulated by the reanalyses and in this sense the reanalysis choice plays a decisive role (Brands *et al.*, 2012); notwithstanding, this assumption should be tested for GCMs in the case of a climate change study.

Regarding temporal aspects, the day-to-day variability (measured with the Pearson correlation) was better captured by both linear regression and neural networks models than by the analog method for both Tx and Tn, in line with results of Gutiérrez *et al.* (2019) and Huth *et al.* (2015) when intercomparing ESD methods. In the case of the annual cycle of Tx and Tn daily mean, GLM and ANN models represented better their seasonal variations than the AN family, especially during winter. In spite of this, when analysing 95th and 5th percentile of Tx and Tn, respectively, AN models showed skills to represent the warmest (summer Tx P95th) and coldest temperatures (winter Tn P5th) of the year, while GLM and ANN

failed to capture their intensity. Notwithstanding, the GLM and ANN family better reproduced P95th (P5th) of Tx (Tn) in autumn and winter (spring and summer). Even though ANN and GLM models performed similarly representing Tx P95th, ANN models outperformed the linear models in the case of Tn P5th. All ESD models satisfactorily captured the characteristic intra-seasonal variations of the P95th in the northern regions of Argentina (R1–R4). The ANN tended to be more skilful than the rest of the statistical families in reproducing the interannual variability of Tx and Tn, presenting higher agreement with observations (reflected as higher year-to-year correlation values) in all regions for both variables. Furthermore, even though the long-term changes were not completely captured by the ESD models, the regression-based models (AN and GLM families) ensembles tended to reproduce the sign of the long term changes and in this regard outperformed the AN models.

Regional differences in the ESD models' performance were also appreciable. In general, ESD models presented higher skills to simulate Tx and Tn in north and central Argentina (R1, R2 and R3) than in northwest Argentina (R4) and Argentinian Patagonia (R5). Over these two regions, the complex topography of the Andes range poses a further challenge to reanalyses for representing predictors and to ESD models for capturing the local variability of daily extreme temperatures.

Any ESD model should be able to show some extrapolation ability when driven by conditions that were unknown during the calibration phase to analyse their potential use for climate change projections (Baño-Medina *et al.*, 2020). In this regard, the evaluation during the warm period (2015–2018) showed that ESD models performed similarly to the cross-validated period (1979–2014), thereby showing skill to represent maximum and minimum temperatures in a more challenging set of years. Hence, the ESD models exhibited capability to simulate warmer conditions (in line with possible future climate conditions) and are therefore a promising tool for climate change applications, in agreement with the literature that addressed this approach (Gutiérrez *et al.*, 2013; Baño-Medina *et al.*, 2020). Note that this strategy is presented here as a first attempt to address the stationarity assumption issue when constructing the ESD models for the region. Certainly, additional strategies should be considered in future analyses.

All described above stresses the relevance of identifying the advantages and shortcomings of the ESD models in simulating daily Tx and Tn over the different regions of Argentina, which are the basis to leverage the potential benefits of downscaling in the multiple applications over the region. The distinctive characteristics of each studied region, as evidenced by the ESD models' different

performance, reinforces the importance of providing tailored regional climate information. Therefore, the findings of this work encourage us to continue exploring some aspects not addressed here such as extreme temperature events, multisite and multivariable dependencies. The use of other brand-new and promising techniques (such as machine learning techniques) would also be of great value to improve our understanding of ESD models' capabilities over the region as well as to develop multimodel ensembles for uncertainty assessment in climate projections.

ACKNOWLEDGEMENTS

This work was supported by the University of Buenos Aires (2018-20020170100117BA, 20020170100357BA) and the ANPCyT (PICT-2019-02933 and PICT-2018-02496) projects.

ORCID

Rocio Balmaceda-Huarte  <https://orcid.org/0000-0003-2188-2797>

Maria Laura Bettolli  <https://orcid.org/0000-0001-7423-0544>

REFERENCES

- Almeira, G., Rusticucci, M. and Suaya, M. (2016) Relationship between mortality and extreme temperatures in Buenos Aires and Rosario. *Meteorológica*, 41(2), 65–79.
- Ambrizzi, T., Reboita, M.S., da Rocha, R.P. and Llopis, M. (2019) The state of the art and fundamental aspects of regional climate modeling in South America. *Annals of the New York Academy of Sciences*, 1436, 98–120. <https://doi.org/10.1111/nyas.13932>.
- Anowar, F., Sadaoui, S. and Selim, B. (2021) Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, 40, 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>.
- Araya-Osses, D., Casanueva, A., Román-Figueroa, C., Uribe, J.M. and Panque, M. (2020) Climate change projections of temperature and precipitation in Chile based on statistical downscaling. *Climate Dynamics*, 54, 4309–4330. <https://doi.org/10.1007/s00382-020-05231-4>.
- Asong, Z.E., Khaliq, M.N. and Wheater, H.S. (2016) Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models. *Climate Dynamics*, 47, 2901–2921. <https://doi.org/10.1007/s00382-016-3004-z>.
- Balmaceda-Huarte, R., Olmo, M.E., Bettolli, M.L. and Poggi, M.M. (2021) Evaluation of multiple reanalyses in reproducing the spatio-temporal variability of temperature and precipitation indices over southern South America. *International Journal of Climatology*, 2021, 1–24. <https://doi.org/10.1002/joc.7142>.
- Baño-Medina, J., Manzanas, R. and Gutiérrez, J.M. (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13, 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>.
- Barros, V.R., Boninsegna, J.A., Camilloni, I.A., Chidiak, M., Magrín, G.O. and Rusticucci, M. (2015) Climate change in Argentina: trends, projections, impacts and adaptation. *Wiley Interdisciplinary Reviews: Climate Change*, 6(2), 151–169.
- Barros, V.R., Grimm, A. and Doyle, M.E. (2002) Relationship between temperature and circulation in southeastern South America and its influence from El Niño and La Niña events. *Journal of the Meteorological Society of Japan*, 80, 21–32. <https://doi.org/10.2151/jmsj.80.21>.
- Beck, H., Zimmermann, N., McVicar, T., Vergopolan, N., Berg, A. and Wood, E.F. (2018) Present and future Köppen–Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5, 180214. <https://doi.org/10.1038/sdata.2018.214>.
- Bedia, J., Baño-Medina, J., Legasa, M.N., Iturbide, M., Manzanas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A.S. and Gutiérrez, J.M. (2020) Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, 13, 1711–1735. <https://doi.org/10.5194/gmd-13-1711-2020>.
- Benestad, R.E., Chen, D., Mezghani, A., Fan, L. and Parding, K. (2015) On using principal components to represent stations in empirical–statistical downscaling. *Tellus A: Dynamic Meteorology and Oceanography*, 67, 1. <https://doi.org/10.3402/tellusa.v67.28326>.
- Bettolli, M.L. and Penalba, O.C. (2018) Statistical downscaling of daily precipitation and temperatures in southern La Plata Basin. *International Journal of Climatology*, 38, 3705–3722. <https://doi.org/10.1002/joc.5531>.
- Bettolli, M.L., Vargas, W.M. and Penalba, O.C. (2009) Soya bean yield variability in the Argentine Pampas in relation to synoptic weather types: monitoring implications. *Meteorological Applications*, 16(4), 501–511.
- Brands, S., Gutiérrez, J.M., Herrera, S. and Cofiño, A.S. (2012) On the use of reanalysis data for downscaling. *Journal of Climate*, 25(7), 2517–2526.
- Brands, S., Taboada, J.J., Cofiño, A.S., Sauter, T. and Schneider, C. (2011) Statistical downscaling of daily temperatures in the NW Iberian Peninsula from global climate models: validation and future scenarios. *Climate Research*, 48, 163–176. <https://doi.org/10.3354/cr00906>.
- Busuioc, A. (2022) Empirical-statistical downscaling: nonlinear statistical downscaling. *Oxford Research Encyclopedia of Climate Science*. <https://doi.org/10.1093/acrefore/9780190228620.001.0001/acrefore-9780190228620-e-770>.
- Cabré, F. and Nuñez, M. (2020) Impacts of climate change on viticulture in Argentina. *Regional Environmental Change*, 20, 12. <https://doi.org/10.1007/s10113-020-01607-8>.
- Casanueva, A., Herrera, S., Iturbide, M., Lange, L., Jury, M., Dosio, A., Maraun, D. and Gutiérrez, J.M. (2020) Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmospheric Science Letters*, 2020(21), e978. <https://doi.org/10.1002/asl.978>.
- Cavalcanti, I.F.A. (2012) Large scale and synoptic features associated with extreme precipitation over South America: a review and case studies for the first decade of the 21st century. *Atmospheric Research*, 118(2012), 27–40. <https://doi.org/10.1016/j.atmosres.2012.06.012>.

- Cavazos, T. and Hewitson, B. (2005) Performance of NCEP–NCAR reanalysis variables in statistical downscaling of daily precipitation. *Climate Research*, 28(2), 95–107.
- Ceccherini, G., Russo, S., Ameztoy, I., Romero, C.P. and Carmona-Moreno, C. (2016) Magnitude and frequency of heat and cold waves in recent decades: the case of South America. *Natural Hazards and Earth System Sciences*, 16, 821–831. <https://doi.org/10.5194/nhess-16-821-2016>.
- Cogliati, M.G., Müller, G.V. and Lovino, M.A. (2021) Seasonal trend analysis of minimum air temperature in La Plata river basin. *Theoretical and Applied Climatology*, 144, 25–37. <https://doi.org/10.1007/s00704-020-03512-w>.
- Coppola, E., Raffaele, F., Giorgi, F., Giuliani, G., Xuejie, G., Ciarlo, J.M., Sines, T.R., Torres-Alavez, J.A., Das, S., di Sante, F., Pichelli, E., Glazer, R., Müller, S.K., Abba Omar, S., Ashfaq, M., Bukovsky, M., Im, E.S., Jacob, D., Teichmann, C., Remedio, A., Remke, T., Kriegsmann, A., Bülow, K., Weber, T., Buntermeyer, L., Sieck, K. and Rechid, D. (2021) Climate hazard indices projections based on CORDEX-CORE, CMIP5 and CMIP6 ensemble. *Climate Dynamics*, 57, 1293–1383. <https://doi.org/10.1007/s00382-021-05640-z>.
- D'onofrio, A., Boulanger, J.P. and Segura, E.C. (2010) CHAC: a weather pattern classification system for regional climate downscaling of daily precipitation. *Climatic Change*, 98, 405–427.
- Dayon, G., Boé, J. and Martin, E. (2015) Transferability in the future climate of a statistical downscaling method for precipitation in France. *Journal of Geophysical Research: Atmospheres*, 120, 1023–1043. <https://doi.org/10.1002/2014JD022236>.
- 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., Hölm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>.
- Dixon, K.W., Lanzante, J.R., Nath, M.J., Hayhoe, K., Stoner, A., Radhakrishnan, A., Balaji, V. and Gaitán, C.F. (2016) Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, 135, 395–408. <https://doi.org/10.1007/s10584-016-1598-0>.
- Fan, X., Jiang, L. and Gou, J. (2021) Statistical downscaling and projection of future temperatures across the Loess Plateau, China. *Weather and Climate Extremes*, 32, 100328. <https://doi.org/10.1016/j.wace.2021.100328>.
- Gaitan, C.F., Hsieh, W.W., Cannon, A.J. and Gachon, P. (2014) Evaluation of linear and non-linear downscaling methods in terms of daily variability and climate indices: surface temperature in southern Ontario and Quebec, Canada. *Atmosphere-Ocean*, 52(3), 211–221. <https://doi.org/10.1080/07055900.2013.857639>.
- García, G.A., Miralles, D.J., Serrago, R.A., Alzueta, I., Huth, N. and Drecer, M.F. (2018) Warm nights in the Argentine Pampas: modelling its impact on wheat and barley shows yield reductions. *Agricultural Systems*, 162, 259–268. <https://doi.org/10.1016/j.agrosy.2017.12.009>.
- Gardner, M.W. and Dorling, S.R. (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32, 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0).
- Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M.J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A.M., Cardoso, R.M., Soares, P.M.M., Czernecki, B. and Pagé, C. (2019) An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*, 39, 3750–3785. <https://doi.org/10.1002/joc.5462>.
- Gutiérrez, J.M., Maraun, D., Brands, S., Manzanas, R. and Herrera, S. (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, 26(1), 171–188.
- Hernanz, A., García-Valero, J.A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M.A. and Rodríguez-Camino, E. (2021a) Evaluation of statistical downscaling methods for climate change projections over Spain: present conditions with perfect predictors. *International Journal of Climatology*, 42, 762–776. <https://doi.org/10.1002/joc.7271>.
- Hernanz, A., García-Valero, J.A., Domínguez, M. and Rodríguez-Camino, E. (2021b) Evaluation of statistical downscaling methods for climate change projections over Spain: future conditions with pseudo reality (transferability experiment). *International Journal of Climatology*, 1–14. <https://doi.org/10.1002/joc.7464>.
- Horton, P. (2021) Analogue methods and ERA5: benefits and pitfalls. *International Journal of Climatology*, 1–19. <https://doi.org/10.1002/joc.7484>.
- Horton, P. and Brönnimann, S. (2019) Impact of global atmospheric reanalyses on statistical precipitation downscaling. *Climate Dynamics*, 52, 5189–5211. <https://doi.org/10.1007/s00382-018-4442-6>.
- Huth, R. (1996) An intercomparison of computer-assisted circulation classification methods. *International Journal of Climatology*, 16, 893–922.
- Huth, R. (1999) Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, 13, 91–101.
- Huth, R. (2002) Statistical downscaling of daily temperature in central Europe. *Journal of Climate*, 15, 1731–1742.
- Huth, R. (2004) Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate*, 17(3), 640–652.
- Huth, R., Kliegrová, S. and Metelka, L. (2008) Nonlinearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology*, 28, 465–477.
- Huth, R., Mikšovský, J., Štěpánek, P., Belda, M., Farda, A., Chládová, Z. and Pišot, P. (2015) Comparative validation of statistical and dynamical downscaling models on a dense grid

- in central Europe: temperature. *Theoretical and Applied Climatology*, 120, 533–553. <https://doi.org/10.1007/s00704-014-1190-3>.
- IPCC. (2021) In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R. and Zhou, B. (Eds.) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. (in press).
- Jeong, D.I., St-Hilaire, A., Ouarda, T.B.M.J. and Gachon, P. (2012) Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. *Stochastic Environmental Research and Risk Assessment*, 26, 633–653. <https://doi.org/10.1007/s00477-011-0523-3>.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–472.
- Kannan, S., Ghosh, S., Mishra, V. and Salvi, K. (2014) Uncertainty resulting from multiple data usage in statistical downscaling. *Geophysical Research Letters*, 41, 4013–4019. <https://doi.org/10.1002/2014GL060089>.
- Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R. and Fiorino, M. (2001) The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 82, 247–268. [https://doi.org/10.1175/1520-0477\(2001\)082.0247:TN50YR.1](https://doi.org/10.1175/1520-0477(2001)082.0247:TN50YR.1).
- Klein Tank, A.M.G., Zwiers, F.W. and Zhang, X. (2009) *Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation*. Geneva: WMO. WMO-TD No. 1500, WCDMP-No. 72, 52 pp.
- Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93, 5–48. <https://doi.org/10.2151/jmsj.2015-001>.
- Koukidis, E.N. and Berg, A.A. (2009) Sensitivity of the Statistical Downscaling Model (SDSM) to reanalysis products. *Atmosphere–Ocean*, 47, 1–18. <https://doi.org/10.3137/AO924.2009>.
- Kumar, Y.P., Maheswaran, R., Agarwal, A. and Sivakumar, B. (2021) Intercomparison of downscaling methods for daily precipitation with emphasis on wavelet-based hybrid models. *Journal of Hydrology*, 599, 126373. <https://doi.org/10.1016/j.jhydrol.2021.126373>.
- Liu, W., Zhang, A., Wang, L., Fu, G., Chen, D., Liu, C. and Cai, T. (2015) Projecting the future climate impacts on streamflow in Tangwang River basin (China) using a rainfall generator and two hydrological models. *Climate Research*, 62, 79–97.
- Lovino, M.A., Müller, O.V., Müller, G.V., Sgroi, L.C. and Baethgen, W.E. (2018) Interannual-to-multidecadal hydroclimate variability and its sectoral impacts in northeastern Argentina. *Hydrology and Earth System Sciences*, 22, 3155–3174. <https://doi.org/10.5194/hess-22-3155-2018>.
- Manzanas, R., Brands, S., San-Martín, D., Lucero, A., Limbo, C. and Gutiérrez, J.M. (2015) Statistical downscaling in the Tropics can be sensitive to reanalysis choice: a case study for precipitation in the Philippines. *Journal of Climate*, 28(10), 4171–4184.
- Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M. and Thiele-Eich, I. (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48, RG3003. <https://doi.org/10.1029/2009RG000314>.
- Maraun, D. and Widmann, M. (2018) *Statistical Downscaling And Bias Correction For Climate Research*. London: Cambridge University Press.
- Maraun, D., Widmann, M. and Gutierrez, J.M. (2019) Statistical downscaling skill under present climate conditions: a synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology*, 39, 3692–3703. <https://doi.org/10.1002/joc.5877>.
- Matulla, C., Zhang, X., Wang, X.L., Wang, J., Zorita, E., Wagner, S. and von Storch, H. (2008) Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. *Climate Dynamics*, 30, 133–144. <https://doi.org/10.1007/s00382-007-0277-2>.
- Montroull, N.B., Saurral, R.I. and Camilloni, I.A. (2018) Hydrological impacts in La Plata basin under 1.5, 2 and 3°C global warming above the pre-industrial level. *International Journal of Climatology*, 38, 3355–3368. <https://doi.org/10.1002/joc.5505>.
- MoradiKhaneghahi, M., Lee, T. and Singh, V.P. (2019) Stepwise extreme learning machine for statistical downscaling of daily maximum and minimum temperature. *Stochastic Environmental Research and Risk Assessment*, 33(4–6), 1035–1056. <https://doi.org/10.1007/s00477-019-01680-4>.
- Müller, G.V. (2007) Patterns leading to extreme events in Argentina: partial and generalized frosts. *International Journal of Climatology*, 27, 1373–1387. <https://doi.org/10.1002/joc.1471>.
- Olmo, M., Bettolli, M.L. and Rusticucci, M. (2020) Atmospheric circulation influence on temperature and precipitation individual and compound daily extreme events: spatial variability and trends over southern South America. *Weather and Climate Extremes*, 29, 100267.
- Olmo, M.E., Balmaceda-Huarte, R. and Bettolli, M.L. (2022) Multi-model ensemble of statistically downscaled GCMs over southeastern South America: historical evaluation and future projections of daily precipitation with focus on extremes. *Climate Dynamics*. <https://doi.org/10.1007/s00382-022-06236-x>.
- Olmo, M.E. and Bettolli, M.L. (2021) Statistical downscaling of daily precipitation over southeastern South America: assessing the performance in extreme events. *International Journal of Climatology*, 42, 1283–1302. <https://doi.org/10.1002/joc.7303>.
- Panda, K.C., Singh, R.M., Thakural, L.N. and Sahoo, D.P. (2022) Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall. *Journal of Hydrology*, 605, 127381. <https://doi.org/10.1016/j.jhydrol.2021.127381>.
- Penalba, O.C., Bettolli, M.L. and Krieger, P.A. (2013) Surface circulation types and daily maximum and minimum temperatures

- in southern La Plata Basin. *Journal of Applied Meteorology and Climatology*, 52(11), 2450–2459.
- Penalba, O.C., Bettolli, M.L. and Vargas, W.M. (2007) The impact of climate variability on soybean yields in Argentina. Multivariate regression. *Meteorological Applications*, 14, 3–14. <https://doi.org/10.1002/met.1>.
- Penalba, O.C., Rivera, J.A. and Pantano, V.C. (2004) The CLARIS LPB database: constructing a long-term daily hydro-meteorological dataset for La Plata Basin, southern South America. *Geoscience Data Journal*, 1, 20–29.
- Perkins, S.E. (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, 164–165 (2015), 242–267. <https://doi.org/10.1016/j.atmosres.2015.05.014>.
- Polasky, A.D., Evans, J.L., Fuentes, J.D. and Hamilton, H.L. (2021) Statistical climate model downscaling for impact projections in the Midwest United States. *International Journal of Climatology*, 42, 3038–3055. <https://doi.org/10.1002/joc.7406>.
- Quesada-Chacón, D., Barfus, K. and Bernhofer, C. (2021) Climate change projections and extremes for Costa Rica using tailored predictors from CORDEX model output through statistical downscaling with artificial neural networks. *International Journal of Climatology*, 41, 211–232. <https://doi.org/10.1002/joc.6616>.
- Ribalaygua, J., Torres, L., Pórtoles, J., Monjo, R., Gaitán, E. and Pino, M.R. (2013) Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, 114, 253–269.
- Rolla, A.L., Nuñez, M.N., Guevara, E.R., Meira, S.G., Rodriguez, G.R. and Ortiz de Zárate, M.I. (2018) Climate impacts on crop yields in Central Argentina. Adaptation strategies. *Agricultural Systems*, 160, 44–59. <https://doi.org/10.1016/j.agsy.2017.08.007>.
- Rolla, A.L., Nuñez, M.N., Ramajon, J.J. and Ramajon, M.E. (2019) Impacts of climate change on bovine livestock production in Argentina. *Climatic Change*, 153, 439–455. <https://doi.org/10.1007/s10584-019-02399-5>.
- Rumelhart, D., Hinton, G. and Williams, R. (1986) Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a0>.
- Rusticucci, M. and Barrucand, M. (2001) Climatología de temperaturas extremas en la Argentina. Consistencia de datos. Relación entre la temperatura media estacional y la ocurrencia de días extremos. *Meteorológica*, 26, 69–83.
- Rusticucci, M., Barrucand, M. and Collazo, S. (2016) Temperature extremes in the Argentina central region and their monthly relationship with the mean circulation and ENSO phases. *International Journal of Climatology*, 37(6), 3003–3017. <https://doi.org/10.1002/joc.4895>.
- Rusticucci, M. and Penalba, O. (2000) Interdecadal changes in the precipitation seasonal cycle over southern South America and their relationship with surface temperature. *Climate Research*, 16, 1–15.
- Sachindra, D.A., Huang, F., Barton, A.F. and Perera, B.J.C. (2013) Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology*, 33, 1087–1106. <https://doi.org/10.1002/joc.3493>.
- Sachindra, D.A. and Kanae, S. (2019) Machine learning for downscaling: the use of parallel multiple populations in genetic programming. *Stochastic Environmental Research and Risk Assessment*, 33, 1497–1533. <https://doi.org/10.1007/s00477-019-01721-y>.
- Satya Sai, K.V.R., Krishnaiah, S. and Manjunath, A. (2021) Statistical downscaling of rainfall under climate change in Krishna River sub-basin of Andhra Pradesh, India using artificial neural network (ANN). *Nature Environment and Pollution Technology*, 20, 805–818.
- Schwingshakl, C., Sillmann, J., Vicedo-Cabrera, A.M., Sandstad, M. and Aunan, K. (2021) Heat stress indicators in CMIP6: estimating future trends and exceedances of impact-relevant thresholds. *Earth's Future*, 9, e2020EF001885. <https://doi.org/10.1029/2020EF001885>.
- Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R. and Sridhar, V. (2018) Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling. *Journal of Hydrology*, 556, 1078–1095. <https://doi.org/10.1016/j.jhydrol.2016.10.048>.
- Seluchi, M.E. and Marengo, J.A. (2000) Tropical–midlatitude exchange of air masses during summer and winter in South America: climatic aspects and examples of intense events. *International Journal of Climatology*, 20, 1167–1190. [https://doi.org/10.1002/1097-0088\(200008\)20:10<1167::AID-JOC526>3.0.CO;2-T](https://doi.org/10.1002/1097-0088(200008)20:10<1167::AID-JOC526>3.0.CO;2-T).
- Skansi, M.M., Brunet, M., Sigró, J., Aguilar, E., Arevalo, J., Bentancur, O., Castellon, Y., Correa, R., Jacome, H., Ramon, A., Oria, C., Pasten, M., Sallons-Mitro, S., Villaroel, C., Martínez, R., Alexander, L. and Jones, P. (2013) Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global and Planetary Change*, 100, 295–307. <https://doi.org/10.1016/j.gloplacha.2012.11.004>.
- Sulca, J., Vuille, M., Timm, O.E., Dong, B. and Zubietta, R. (2021) Empirical–statistical downscaling of austral summer precipitation over South America, with a focus on the central Peruvian Andes and the equatorial Amazon Basin. *Journal of Applied Meteorology and Climatology*, 60(1), 65–85.
- Sun, F., Mejia, A., Sharma, S., Zeng, P. and Che, Y. (2020) Evaluating the credibility of downscaling: integrating scale, trend, extreme, and climate event into a diagnostic framework. *Journal of Applied Meteorology and Climatology*, 59, 1453–1467. <https://doi.org/10.1175/JAMC-D-20-0078.1>.
- Sun, L. and Lan, Y. (2021) Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: localization and comparison with other methods. *International Journal of Climatology*, 41, 1128–1147. <https://doi.org/10.1002/joc.6769>.
- Timbal, B., Dufour, A. and McAvaney, B. (2003) An estimate of future climate change for western France using a statistical downscaling technique. *Climate Dynamics*, 20, 807–823. <https://doi.org/10.1007/s00382-002-0298-9>.
- Timbal, B. and McAvaney, B.J. (2001) An analogue-based method to downscale surface air temperature: application for Australia. *Climate Dynamics*, 17, 947–963.
- Vaughan, A., Tebbutt, W., Hosking, J.S. and Turner, R.E. (2022) Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development*, 15, 251–268. <https://doi.org/10.5194/gmd-15-251-2022>.
- Verdin, A., Rajagopalan, B., Kleiber, W., Podestá, G. and Bert, F. (2018) A conditional stochastic weather generator for seasonal

- to multi-decadal simulations. *Journal of Hydrology*, 556, 835–846.
- Verdin, A., Rajagopalan, B., Kleiber, W., Podestá, G. and Bert, F. (2019) BayGEN: a Bayesian space-time stochastic weather generator. *Water Resources Research*, 55, 2900–2915. <https://doi.org/10.1029/2017WR022473>.
- Vrac, M., Stein, M.L., Hayhoe, K. and Liang, X.Z. (2007) A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, 34, L18701.
- Wang, F., Tian, D., Lowe, L., Kalin, L. and Lehrter, J. (2021) Deep learning for daily precipitation and temperature downscaling. *Water Resources Research*, 57, e2020WR029308. <https://doi.org/10.1029/2020WR029308>.
- Wilby, R.L., Dawson, C.W. and Barrow, E.M. (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling and Software*, 17, 147–159.
- Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B. C., Main, J. and Wilks, D.S. (1998) Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, 34(11), 2995–3008. <https://doi.org/10.1029/98WR02577>.
- Zazulie, N., Rusticucci, M. and Raga, G.B. (2017) Regional climate of the subtropical central Andes using high-resolution CMIP5 models—part I: past performance (1980–2005). *Climate Dynamics*, 49, 3937–3957. <https://doi.org/10.1007/s00382-017-3560-x>.
- Zorita, E. and von Storch, H. (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, 12(8), 2474–2489.

SUPPORTING INFORMATION

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

How to cite this article: Balmaceda-Huarte, R., & Bettolli, M. L. (2022). Assessing statistical downscaling in Argentina: Daily maximum and minimum temperatures. *International Journal of Climatology*, 42(16), 8423–8445. <https://doi.org/10.1002/joc.7733>