

Statistical downscaling of daily precipitation over southeastern South America: Assessing the performance in extreme events

Matías Ezequiel Olmo^{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-IFAECI/CNRS-CONICET-UBA), Buenos Aires, Argentina

Correspondence

Olmo Matías Ezequiel, Faculty of Exact and Natural Sciences, University of Buenos Aires, Intendente Güiraldes 2160, Ciudad Universitaria, Buenos Aires C1428EGA, Argentina.

Email: molmo@at.fcen.uba.ar

Funding information

Agencia Nacional de Promoción Científica y Tecnológica, Grant/Award Number: PICT-2018-02496; Universidad de Buenos Aires, Grant/Award Numbers: 20020170100357BA, 2018-20020170100117BA

Abstract

The performance of multiple empirical statistical downscaling (ESD) methods was assessed for simulating daily precipitation during 1979–2017 over southeastern South America (SESA), a region where extremes are remarkable. Meteorological stations were used as reference and three gridded precipitation products were included to account for observational uncertainty. The set of ESD models involved different configurations of the analogue method (ANs), deterministic and stochastic versions of neural networks (NNs) and generalized linear models (GLMs) and circulation-conditioned GLMs (GLM_WTs). The years with the largest number of extreme events (wet years) were calibrated separately. The spatio-temporal variability of extremes was assessed by analysing their intensity, spatial extent, frequency and interannual variability. An overall good performance of the ESD models was found for several aspects of daily precipitation. ESD performance dispersion was usually contained in the observational spread. No particular model configuration was found to perform best in all aspects, indicating the advantage of considering a multi-model ensemble. The ANs tended to follow the stations, satisfactorily simulating daily precipitation and its extremes. The deterministic GLMs strongly underestimated precipitation estimates and were not able to represent extreme frequencies and intensities, but this was alleviated by employing a stochastic version of the method. Furthermore, the use of weather types to condition the GLMs (GLM_WTs) considerably improved model performance, particularly for the annual cycle and the spatial structure of extreme precipitation. The NNs adequately reproduced the spatial behaviour and intra-annual variability of extreme precipitation, although they underestimated its intensity in their deterministic version. An analysis of the regional time series of extremes showed consistency among datasets and evidenced the influence of the ENSO teleconnection on the wet years, which were commonly well-simulated by the statistical models. ESD models presented good skills in simulating a wetter climate over SESA, which is of particular importance in a climate change scenario.

KEY WORDS

extremes, observational uncertainty, southeastern South America, statistical downscaling

1 | INTRODUCTION

Southeastern South America (SESA, roughly between 50–65°W and 20–40°S) is one of the most populated regions of South America, covering central-eastern Argentina, southern Brazil, southern Paraguay and Uruguay, and stands out as one of the雨iest portions of the continent. The region is highly exposed to extreme precipitation events affecting the socio-economic activities, energy demand and health systems. Furthermore, SESA is part of La Plata Basin, the second basin of South America in terms of river discharge and size, which plays a critical role in the regional economy (Berbery and Barros, 2002; Barros *et al.*, 2006). Cattle raising, rainfed agriculture and hydroelectric power are the most valuable assets, resulting especially susceptible to extreme precipitation, which often leads to overflows and floods in the region (Vörösmarty *et al.*, 2013; Cavalcanti *et al.*, 2015).

Extreme rainfall in SESA is one of the main contributors to the hydrological cycle. It is typically associated with extratropical synoptic systems during the cold season, cyclogenesis during the transition seasons and mesoscale convective systems particularly during the warm season (Cavalcanti, 2012). Precipitation variability and extremes in SESA are influenced by different forcings at multiple temporal and spatial scales, such as sea surface temperature (SST) anomalies, teleconnections patterns like El Niño–Southern Oscillation (ENSO), the position of the South American Convergence Zone and the South American Low Level Jet (SALLJ) (Salio *et al.*, 2007; Barreiro, 2010; Robledo *et al.*, 2016). Moreover, several studies have detected an increasing frequency and intensity of precipitation extremes in the region, mainly during the warm season (Penalba and Robledo, 2010; Cerón *et al.*, 2020; Olmo *et al.*, 2020). Notwithstanding, only few studies have addressed the analysis of future projections of extreme precipitation over SESA, indicating that these changes are expected to be more pronounced during the mid and late 21st century (Sörensson *et al.*, 2010; Sillmann *et al.*, 2013; Blázquez and Solman, 2020; Li *et al.*, 2020). For this purpose, it is necessary to first evaluate model performances during the historical period in order to reduce the uncertainty in the projected changes. In line with this, the SESA Flagship Pilot Study initiative endorsed by the Coordinated Regional Climate Downscaling Experiment (CORDEX) addresses the study of precipitation extremes over the region from a

collaborative intercomparison experiment of modelling strategies (Bettolli *et al.*, 2021). Above all, studying extreme precipitation in SESA implies a complex but notably essential piece of work.

The need for regional climate information for impact and adaptation planning in a climate change scenario is leading discussion among the scientific community and policy makers (Maraun *et al.*, 2015; Gutiérrez *et al.*, 2019). In particular, a key problem to solve is the gap between the coarse-resolution outputs of global climate models (GCMs) and local spatial scales, which can be assessed by two main approaches: dynamical downscaling techniques—based on regional climate models (RCMs)—and empirical statistical downscaling methods (ESD), by identifying the relationships between large-scale atmospheric variables and local climate. In spite of their strengths and weaknesses, both procedures have demonstrated skills in representing different aspects of the regional climate (Huth *et al.*, 2015; Casanueva *et al.*, 2016; Ambrizzi *et al.*, 2018; Gutiérrez *et al.*, 2019, among others).

A growing body of literature has validated and compared ESD methods for local climate modelling in different regions around the globe (Benestad, 2010; Frost *et al.*, 2011; Castellano and De Gaetano, 2016; Horton and Brönnimann, 2018; Maraun *et al.*, 2018; Araya-Osset *et al.*, 2020). Gutiérrez *et al.* (2019) presented a wide inter-comparison of statistical downscaling methods over Europe and found that most of the ESD methods greatly improved raw model biases, while no approach seemed to be superior in general, since there was large method-to-method variability. Similar outcomes were found by Frost *et al.* (2011) when testing multiple methods of downscaling GCM simulations for daily precipitation in southeastern Australia. Furthermore, for the European Alps, Schmidli *et al.* (2007) described that the performance of the downscaling models varies substantially among regions and the seasons of the year. In spite of this, in SESA the ESD potential to simulate regional climate characteristics has not been explored as thoroughly or systematically as in other parts of the world. Few studies pointed out the capability of ESD simulations in both probabilistic and deterministic estimates of different properties of daily and monthly precipitation such as the observed frequencies, mean values and internal model variability (D'onofrio *et al.*, 2010; Menéndez *et al.*, 2010). Bettolli and Penalba (2018) employed the analogue method for downscaling daily precipitation and extreme

temperatures and evidenced the analogues skills in reproducing extreme percentiles and their spatial distributions, as well as the probability of compound temperature and precipitation extreme events. More recently, Bettolli *et al.* (2021) and Solman *et al.* (2021) analysed the capability of statistical (analogues and generalized linear models) and dynamical regional simulations and multiple sources of observations in representing a set of daily extreme events in SESA, exhibiting large dispersion among them. Hence, there is still a need to evaluate different ESD techniques and quantify their errors over the SESA region, particularly for precipitation extremes.

All things considered, the main goals of the present study are the calibration and validation of multiple statistical downscaling techniques for daily precipitation in SESA and the assessment of their performance with respect to the spatial and temporal variability of extreme events.

2 | DATA AND METHODOLOGY

2.1 | Precipitation data

Daily precipitation data were considered during the period 1979–2017 from a meteorological stations network over SESA (Figure 1), which was employed and quality controlled in previous studies by the research group (Olmo *et al.*, 2020). Data were provided by the National Weather Services of Argentina, Brazil, Paraguay and Uruguay and the National Institute of Agricultural Technology of Argentina, resulting in 86 stations over SESA with less than 20% of missing data (see Table S1, Supporting Information for a description of the meteorological stations). Despite the growing availability of several observational databases from multiple sources (including radar, satellite, gauge observations and combinations of them), the uncertainties in precipitation extremes are still large, which should be especially considered for model calibrations and evaluations (Bettolli *et al.*, 2021). Thus, including as much information as possible is of key importance to assess the observational uncertainty when performing climate research and model evaluations. Therefore, three gridded daily precipitation products with different spatial resolutions were considered for comparative purposes: (a) the CHIRPS dataset, which provides blended gauge-satellite precipitation estimates; (b) the CPC gauge-based dataset, which is built using in-situ information and numeric models; (c) the MSWEP dataset derived from satellite data that also combines information from gauge and reanalysis sources. A more detailed description of the observational datasets used in this study is displayed in Table 1.

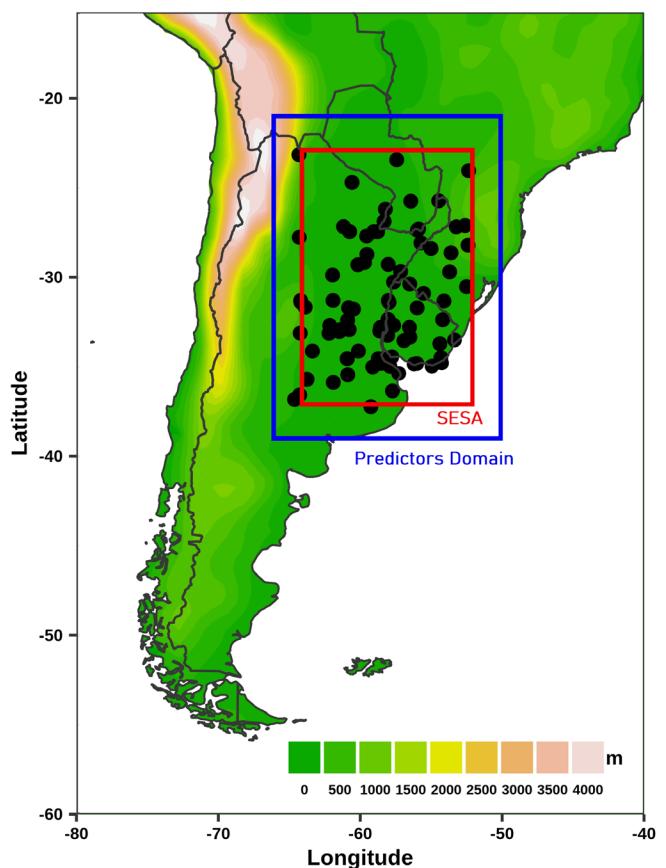


FIGURE 1 Location of the meteorological stations considered over southeastern South America (SESA, in the red box) and the ESD predictors variables domain (blue box) [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Empirical statistical downscaling

2.2.1 | Data

Daily precipitation from the station points dataset was used as predictand in the SESA domain (Figure 1). The ESD experiment followed the perfect prognosis approach, which relies on predictor variables well represented by reanalyses—considered as quasi-observed predictors—to train the statistical models (Maraun *et al.*, 2010). ERA-Interim daily mean fields over 50–67°W and 20–40°S were interpolated at 2° and employed for the following set of predictors including circulation, moisture and temperature variables (Dee *et al.*, 2011):

- Geopotential height at 500 and 1,000 hPa (Z500 and Z1000, respectively).
- Meridional wind at 850 hPa (V850).
- Specific humidity at 700 and 850 hPa (Q700 and Q850, respectively).
- Air temperature at 700 and 850 hPa (T700 and T850, respectively).

TABLE 1 Observational datasets of daily precipitation considered in this study

Dataset	Label	Grid resolution	Reference
CHC InfraRed Precipitation with Station data	CHIRPS	0.25°	Funk <i>et al.</i> (2014)
CPC Global Unified Gauge-Based Analysis of Daily Precipitation	CPC	0.5°	Xie <i>et al.</i> (2010)
Multiple-Source Weighted-Ensemble Precipitation	MSWEP	0.1°	Beck <i>et al.</i> (2017)
Meteorological Stations	STATIONS	—	—

TABLE 2 Predictor combinations employed for the different ESD models configurations

Label	Predictors configuration	AN	GLM	GLM_ST	GLM_WT	NN	NN_ST
L16	Z500, Q700, T700, U850, V850, Q850, Z1000: 16 nearest grid-points	x	x	x		x	x
L16qv	Z500, Q850, V850: 16 nearest grid-points		x	x	x		
L4	Z500, Q700, T700, U850, V850, Q850, Z1000: 4 nearest grid-points				x		
L4qv	Z500, Q850, V850: 4 nearest grid-points	x				x	x
LS	Z500, V850, Z1000: Principal components (95% explained variance) Q700, T700, Q850, T850: 4 nearest grid-points	x				x	x
PC	Z500, Q700, T700, U850, V850, Q850, Z1000: Principal components (95% explained variance)	x	x	x	x	x	x
qv	Z500, Q850, V850: Principal components (95% explained variance)		x	x	x		

Note: Each model will be referred to as *method_label* (for instance: AN_L16 corresponds to the analogue method when using the L16 configuration).

These variables were chosen based on a sensitivity analysis of different predictors, a predictor screening performed in La Plata basin by Bettolli and Penalba (2018) and the evaluation of the synoptic environment associated with extreme rainfall events over SESA (Lavín-Gullón *et al.*, 2021; Olmo and Bettolli, 2021) and mesoscale convection systems by Rasmussen and Houze Jr. (2016). These studies found a baroclinic configuration leaded by a dipolar structure of Z500 anomalies with an anticyclonic (cyclonic) centre over the Atlantic (Pacific) Ocean and an intensification and southern penetration of the SALLJ, strengthening the meridional advects of humid and warm air from the Amazonia to SESA, which significantly enhances the occurrence of extreme rainfall events over the region. Taking this into consideration, the large-scale atmospheric structures were considered with spatial-wide predictor variables by using the principal components (PCs), explaining 95% of the total variance. Additionally, pointwise predictors using the values from the four and 16 nearest grid points to the target station point were used to address local influences (see Table 2, predictors configuration). It is worth mentioning that several trials were performed using the

different combinations of predictors of Table 2. For the sake of conciseness, only four configurations in each ESD family are presented here.

2.2.2 | Cross-validation procedure

The 39-year period 1979–2017 was split into eight folds, seven 5-year sets and one 4-year set. All folds but one were randomly selected. For each fold, the ESD models were calibrated using the remaining seven folds. The full annual series were employed, that is, ESD models were not trained by individual season. Finally, every model output for the eight folds were merged into a single 39-year time series for validation. Further information about the cross-validation scheme can be found in Gutiérrez *et al.* (2013). Considering that SESA stands out as a region with a remarkable increment in both the frequency and intensity of extreme precipitation (Cavalcanti, 2012; Cerón *et al.*, 2020; Olmo *et al.*, 2020), the non-random fold mentioned above was formed by five selected years with extreme precipitation conditions to assess the performance of the ESD methods in possible

future climate conditions. This *wet fold* consisted of the years 1983, 1991, 1998, 2002 and 2014, which presented the greatest extreme rainfall events frequency over SESA. In this way, these wet years are especially challenging in this setup, since the models were calibrated with drier years. A comprehensive analysis of these wet years was introduced in sections 3.3 and 3.4.

2.2.3 | Methods

In order to move forward with studies on the ESD methods capability to simulate daily precipitation in SESA, different statistical downscaling techniques were chosen:

Analogues

The analogue method (AN) is based on the assumption that similar atmospheric configurations can lead to similar local occurrences (Zorita and von Storch, 1999; San Martín *et al.*, 2017). Thereby, for each daily record in a fold, the most similar large-scale situation (the nearest neighbour) can be found in the folds used for calibration according to the Euclidean distance (Timbal *et al.*, 2009). Commonly, one can select the first analogue, a random or the mean of the closest analogues. In this study, we selected the first analogue day, although other configurations were evaluated, resulting in poorer performances (not shown).

Generalized linear models

The generalized linear models (GLM) are an extension of linear regression, allowing for non-normal predictand distributions. Given that the focus of this study is on precipitation extremes, both deterministic (expected value) and stochastic (random sample from the predicted distribution at each time step) versions of GLMs are presented (GLM and GLM_ST, respectively). These models consisted in a two-stage implementation, with a Bernoulli distribution and a logit link (equivalent to a logistic regression) for precipitation occurrence, and a Gamma distribution and log link for precipitation amounts (San Martín *et al.*, 2017; Bettolli *et al.*, 2021). The climatological probability of rain occurrence (set above 1 mm) in each station was used as a reference threshold for determining precipitation occurrence in the deterministic simulations. In this way, two simulated time series were produced at each station point (one for precipitation amount and the other one of binary values for precipitation occurrence) so that the final rainfall time series was calculated by multiplying these two series. See Bedia *et al.* (2020) for further details about these procedures.

Generalized linear models + weather types

This family of ESD methods is made up of circulation-conditioned GLMs (GLM-WT) (San Martín *et al.*, 2017). In particular, we used the 16 weather types (WT) found by a Self-Organizing Maps (SOM) clustering of Z500 anomalies over an extended domain over southern South America by Olmo and Bettolli (2021). This classification spans the variety of synoptic states during the 1979–2017 period of study, and it was found to successfully represent the atmospheric configurations associated with extreme precipitation events over SESA. Then, a conditional GLM model to each WT was set for the different configurations as shown in Table 2.

Neural networks

The artificial neural network models (NN) are, essentially, nonlinear regression-based methods composed of neurons that are organized in layers of feed-forward networks. These neurons are connected between consecutive layers, from the input layer—in our case, daily predictors fields—through a set of hidden layers in the network, to the output layer, returning the daily precipitation simulations. These connections are characterized by different weights—which are learnt from data—and activation functions, which are nonlinear functions applied in each neuron (Cavazos, 1999; Haylock *et al.*, 2006; Baño-Medina *et al.*, 2020). In this study, a sensitivity analysis was performed by testing several NN configurations, varying the topology of the networks as well as the different parameters that control them (not shown). Finally, a two-layer NN configuration with 25 and 15 neurons, respectively, was selected. The learning rate parameter was set to 0.01 and the activation function in the hidden layers was the sigmoid function. As for the GLMs, both deterministic and stochastic (NN and NN_ST, respectively) setups were included in this study. The neural network models were implemented in two stages: one for precipitation occurrence, for which the output layer performed a linear function, and one for precipitation amounts, for which the output layer performed the sigmoid function. Following the approach employed by Baño-Medina *et al.* (2020), the NN optimizes the negative log-likelihood of a Bernoulli-Gamma distribution. Specifically, it estimates the probability of rain of the Bernoulli distribution and the shape and scale parameters of the Gamma distribution to obtain the rain occurrences (delimited by the climatological probability of rain occurrence in the deterministic simulations) and amounts, respectively. Analogously to the GLMs implementation, the final time series resulted from the product between the precipitation occurrence and amount.

The ESD methods were implemented using the R-based climate4R open framework (Iturbide *et al.*, 2019;

Bedia *et al.*, 2020). The different configurations and models discussed in this study are outlined in Table 2.

2.3 | Validation framework

Model evaluation was carried out considering the station points over SESA (used as reference) and, in the case of the gridded precipitation products, the nearest grid cell to each meteorological station was selected for the comparison.

On one hand, several metrics and indices considered for the evaluation of ESD precipitation models in the VALUE intercomparison experiment (Gutiérrez *et al.*, 2019; Bedia *et al.*, 2020) were estimated (Table 3). These metrics were performed for the wet fold and for the remaining folds (*normal folds* hereafter) separately, so model evaluation on extreme rainfall conditions could be assessed. On the other hand, focus was put on the assessment of the spatial and temporal variability of extreme precipitation by analysing their intensity, spatial extent, frequency and interannual variability. For the latter analyses, the full time series, obtained by joining the normal and wet folds to reconstruct the period 1979–2017, were used. For each observational dataset and ESD model output, extreme precipitation was defined at each grid cell or station point as the precipitation values in those days when the accumulated precipitation exceeded the 95th percentile (P95) of the empirical distribution of rainy days (accumulated precipitation above 1 mm). This daily percentile was calculated in the base-period 1981–

2010, based on a 29-day moving window centred on each calendar day, following the methodology employed in previous studies (Olmo *et al.*, 2020). Both the amounts and the frequency of exceedance of P95 in the period of analysis were evaluated. Additionally, Taylor diagrams (Taylor, 2001) were used for summarizing the representation of the spatial distribution of precipitation in SESA. These diagrams quantify the degree of statistical similarity between the reference observations (STATIONS) and the rest of the datasets, reporting the Pearson correlation coefficient, the standard deviation and the centred root-mean-squared error.

3 | RESULTS

3.1 | Main features

The different indices outlined in Table 3 were estimated in every station point over SESA for all ESD models and gridded precipitation datasets (the nearest grid cell to each station), considering the meteorological stations as reference. Boxplots in Figure 2 show results for the normal conditions (across all station points) and points represent the spatial median value of each index during the wet fold. The first three boxes present the comparison between the gridded and local observations (green colours). Spearman correlation of the daily time-series indicated general good correspondence for the ESD models with STATIONS. The lowest correlations were found for the ANs family, followed by the stochastic versions of the GLMs and NNs, being GLM_ST_L16qv and GLM_ST_qv and the ones exhibiting the lowest values. Note that these stochastic models are expected to lose part of the temporal correlation compared to their deterministic versions. The GLM_WTs typically presented lower correlations than the GLMs, similar to the NNs. In this sense, the simplest ESD models in terms of local predictors information and number of variables were the least correlated such as AN_L4qv, GLM_WT_L16qv and NN_ST_L4qv. Models with local predictors (16 nearest points) usually had better performance, even though the spatial predictors configuration showed some good results as well, like AN_PC and GLM_WT_PC. Broadly speaking, ESD dispersion was contained in the spread exposed by the observational datasets, where CHIRPS (MSWEP) evidenced the lowest (highest) daily correlations among the gridded precipitation products. In the case of the RMSE, the ANs and GLM_WTs presented the highest errors for the deterministic models, whereas the NNs and especially the GLMs exhibited RMSE values more comparable to the observational datasets. Both NN_STs and GLM_STs presented the highest RMSE values, larger than their deterministic

TABLE 3 Validation metrics and indices considered from the VALUE intercomparison experiment (Gutiérrez *et al.*, 2019)

Metric/index	Definition
Spearman correlation	Daily time series Spearman correlation
RMSE	Root mean squared error
R01 (ratio)	Relative frequency of wet days (≥ 1 mm)
R20 (ratio)	Relative frequency of days with precipitation ≥ 20 mm
R20p (ratio)	Precipitation amount falling in days with precipitation ≥ 20 mm
SDII (ratio)	Simple Daily Intensity Index
P98 (ratio)	98th percentile of wet days (≥ 1 mm)
WWprob (ratio)	Probability of wet-wet transition
DWprob (ratio)	Probability of dry-wet transition

Note: The indices are expressed as a ratio between the predicted and observed index.

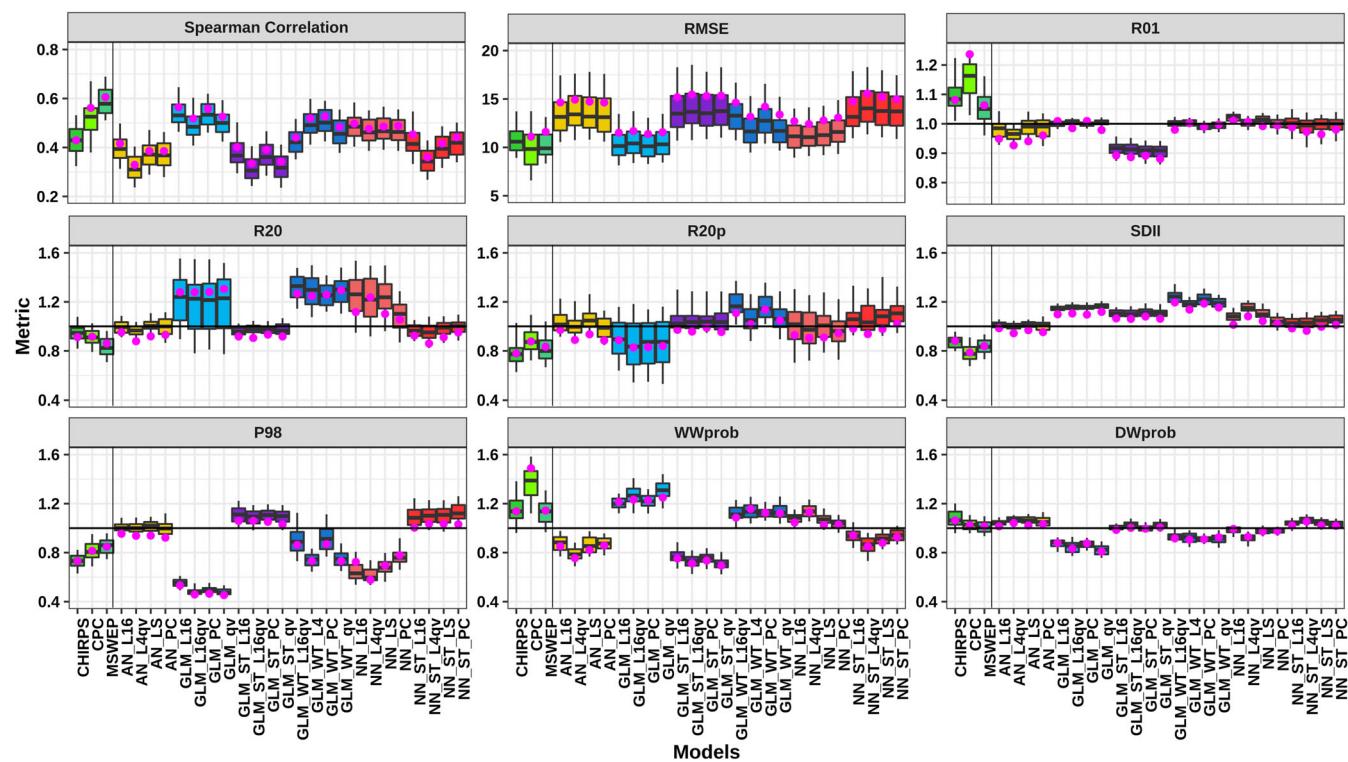


FIGURE 2 Results of the metrics and indices outlined in Table 3, estimated at every station point over SESA for all ESD models and the gridded precipitation datasets (the nearest grid cell to each station was selected) considering the meteorological stations (STATIONS) as reference. Boxplots show results for all stations under normal conditions, indicating the 25th, 50th and 75th percentiles in their boxes and the 5th and 95th percentiles in their whiskers, while points represent the spatial median value of each index during the wet fold. Colours indicate the different observational datasets and ESD model families [Colour figure can be viewed at wileyonlinelibrary.com]

versions but similar to the ANs. Turning now to the evaluation of the wet-day frequency (R01), all ESD models performed similarly to the stations reference, as seen by the small dispersion of the R01 ratios around one. Note that the deterministic regression-based GLMs and NNs models by construction reproduce the observed precipitation frequency. However, the stochastic version of GLMs consistently underestimated this index, probably because the random resampling configuration of the model causes the prediction to be slightly biased. The ANs models slightly underestimated R01 and presented the largest dispersion, although smaller than the one exhibited by the gridded precipitation products, which typically overestimated wet-day frequency, more pronounced in CPC. In terms of daily precipitation intensities (SDII), the ANs models adequately reproduced the observations, whereas the rest of the models tended to overestimate this index, although with small dispersion. Surprisingly, the deterministic regression-based models slightly overestimate SDII when, by construction, they are calibrated to optimize the conditional mean. This could be due to the theoretical fit introduced in the construction of these models and should be explored in future sensibility analysis. R20 was usually overestimated by the ESD models such as the

GLM_WTs and deterministic NNs and GLMs, with larger spatial variability in the latter. The exception was detected in the ANs, which successfully reproduced the observed frequencies of heavy precipitation days, particularly the configurations that included local predictors. The ANs performance was often better than the gridded precipitation products, which tended to underestimate the observed frequencies (R20) and rainfall amounts associated with R20 (R20p) compared to STATIONS. Note that the GLM_STs and NN_STs successfully reproduced this index, presenting ratios around one and reducing the spatial dispersion observed in the deterministic regression-based models. R20p was well represented by the ANs and the neural networks models like NN_LS and NN_L16, while the GLM_WTs (GLMs) tended to slightly overestimate (strongly underestimate) heavy precipitation intensities. The stochastic model families showed similar performances, tending to overestimate the index. When studying precipitation extremes, the P98 index exposed the really good skills of the analogues, with ratios around 1, while the other deterministic models tended to underestimate it: GLM_WTs and NNs presented similar values than the precipitation products, whereas the GLMs showed a more pronounced

underestimation of high-intensity rainfall values, which is a common limitation of these deterministic regression-based methods (San Martín *et al.*, 2017; Hertig *et al.*, 2018; Bettolli *et al.*, 2021). Regarding this, the weather-types conditioned GLM evidenced an improvement of the GLM models in representing the tail of daily precipitation distribution. The stochastic versions of the GLMs and NNs considerably reduced the differences in P98 with the STATIONS reference but exposed higher extreme intensities. Recall that the deterministic regression-based techniques tend to reproduce the mean conditions towards the observations and therefore the benefit of introducing a stochastic component in these methods arises when simulating extremes. In the case of the transition-probability indices WWprob and DWprob, opposite results were generally found, and smaller dispersion was detected for the latter: the ANs underestimated (overestimated) WWprob (DWprob) and the rest of the deterministic models presented overestimations (underestimations) of this index. The largest differences were found for the simplest models of each family, such as GLM_L16qv and NN_L4qv. In addition, the stochastic GLMs and NNs models tended to present a contrary behaviour compared to the deterministic versions.

Note that, in general, the performance during the wet fold (defined in section 2.2) was in good agreement with the results during the rest of the period, as depicted by the point representing spatial median wet fold conditions, typically contained in the boxplots of almost all metrics for each ESD model (Figure 2). Recall that these wet years were especially challenging in the models' setup since they were calibrated with drier years. For instance, the daily correlation values were in accordance with results from the normal folds. The mean errors (RMSE) were positioned in the upper limits of the boxplots for both the ESD models and the gridded precipitation products, indicating the higher difficulty of the models and the gridded precipitation datasets to capture the features of these years with more intense precipitation. Contrary to the normal folds, the analogues underestimated the heavy precipitation indices (for instance: R20, R20p and P98) and also exhibited more difficulty in reproducing the rain frequency during the wet years (R01), whereas results obtained for the regression-based models were generally within the normal folds values.

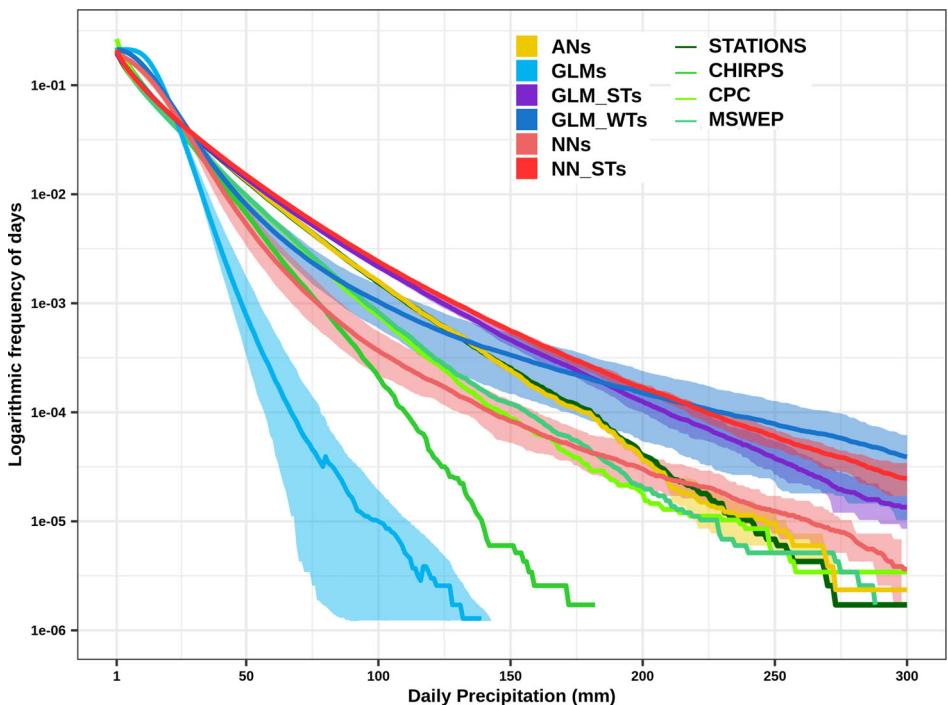
In a next step, the frequency distribution of daily precipitation in all days during 1979–2017 was calculated (Figure 3). The ensemble means (solid lines) and dispersion among members (shaded colours) were displayed for each ESD model family. The analogues closely followed the STATIONS distribution, with only small dispersion between the different ANs in daily precipitation intensities above 200 mm. The rest of the models—except for

the GLM_STs and NN_STs—tended to highly overestimate the frequencies of light precipitation up to 20 mm, as reflected by the misrepresentation of the shape of the frequency curve for these intensities. Both GLM_STs and NN_STs presented similar distributions and a reduced ensemble spread, overestimating the frequency distribution of daily precipitation above 50 mm. Nevertheless, both ESD families—but particularly the former one—appeared to improve the representation of the frequency distribution compared to their deterministic version. In the case of the GLM_WTs, the ensemble mean depicted an underestimation (overestimation) of the number of days with precipitation values less (greater) than 150 mm, with larger dispersion in the ensemble for higher precipitation amounts. As found for some precipitation-intensity indices in Figure 2, the circulation-conditioned GLM models outperformed the simple GLMs, which were found to clearly underestimate the frequencies of days not only for heavy rainfall but also for moderate intensities higher than 25 mm. Even more, their ensemble spread notably increased for the highest precipitation rates. In the case of the NNs, they tended to underestimate precipitation frequencies compared to STATIONS, in agreement with Haylock *et al.* (2006), and presented slight overestimations for high precipitation amounts. However, these models showed some similarities to the gridded precipitation datasets and therefore within the observational uncertainty. CHIRPS exhibited the largest underestimations, whereas MSWEP and CPC performed better. The foregoing discussion implies that the general underestimation by the gridded precipitation products when considering the nearest station point could be due to these datasets being composed of grid cells that represent areal-averages. Hence, they usually present lower rainfall values than the meteorological station points. Additionally, the same analysis was replicated only for the years contained in the wet fold and results did not sensibly change compared to the results discussed above (see Figure S1).

3.2 | Extreme events

Figure 4 displays the seasonal average values of P95 over SESA during both seasons of the year, separately. During the warm season (October–March), maximum values were found in the centre of the domain with large amounts of precipitation, which makes SESA a distinctive region of South America for precipitation extremes occurrence (Re and Barros, 2009; Solman and Blázquez, 2019; Olmo and Bettolli, 2021). The observational datasets exhibited some discrepancies in P95 intensities, which were typically underestimated (mainly

FIGURE 3 Daily precipitation frequency distribution over SESA considering all data together across the region. Colours indicate the different observational datasets and ESD model families. For each ESD family, the ensemble mean and the range of the ensemble members are shown in solid lines and in shading colours, respectively [Colour figure can be viewed at wileyonlinelibrary.com]



during the warm season) but the spatial distribution was relatively well captured. When addressing the evaluation of the ESD models, the analogues family—and especially AN_PC—stood out from the rest of the simulations as these models successfully simulated both the intensity and the spatial behaviour of P95 across SESA. The reproduction of the spatial structures is a natural feature of the analogue models as they sample the historical records, preserving the correlation patterns (Frost *et al.*, 2011). The stochastic GLMs and NNs also reproduced well the observed P95 pattern, exhibiting similar configurations among each model family. GLM_WTs often showed lower values than the STATIONS percentiles, although the configuration with spatial-wide predictors (GLM_WT_PC) presented better results. However, the circulation-conditioned GLMs indicated an improved representation than the simple GLMs, as the latter strongly underestimated both P95 intensities and spatial variability, in accordance with our findings in Figures 2 and 3. Deterministic NNs performed similarly to the GLM_WTs, and NN_PC was the configuration that seemed to best represent the spatial extent of the P95 maximum in this family. During the cold season, the set of ESD models exhibited comparable performances to the warm season, although the stochastic models slightly overestimated the P95 maxima. ESD models adequately reproduced the eastward shift of extreme precipitation detected in all the observational datasets, although CHIRPS and MSWEP depicted maximum values of greater magnitude than STATIONS. These seasonal analyses indicated that, in spite of a wider extent of P95

maximum intensities during the warm season, some areas of SESA are exposed to large amounts of rainfall during all the year and that ESD models were able to capture the seasonal behaviour of extreme precipitation across the region.

The annual cycle of P95 over SESA is illustrated in Figure 5. The comparison among observational datasets evidenced that the precipitation products underestimated the P95 throughout the year, which was consistent with the underestimation of precipitation values stated in section 3.1, while the shape of the annual cycle was well represented. The ANs intensities and intra-annual variability was congruent with the meteorological stations, although they tended to show higher P95 values during the austral winter and spring seasons. The GLM_WTs showed more underestimations of the percentile, particularly during the warm season. Again, these models outperformed the simple GLMs, which failed in reproducing P95 intensities and clearly underestimated the amplitude of the annual cycle though the shape of the cycle was fairly well represented. The stochastic version of the GLMs notably upgraded the performance despite overestimating the P95, mainly during winter. In the case of the deterministic NNs, they tended to present extreme intensities similar to the gridded precipitation products but underestimating the amplitude of the cycle. Again, following a stochastic approach in regression-based methods improved the models' performance in reproducing extreme values, such was the case of the NN_STs models. Overall, the largest differences with the STATIONS were detected during austral summer—when a relative P95 maximum

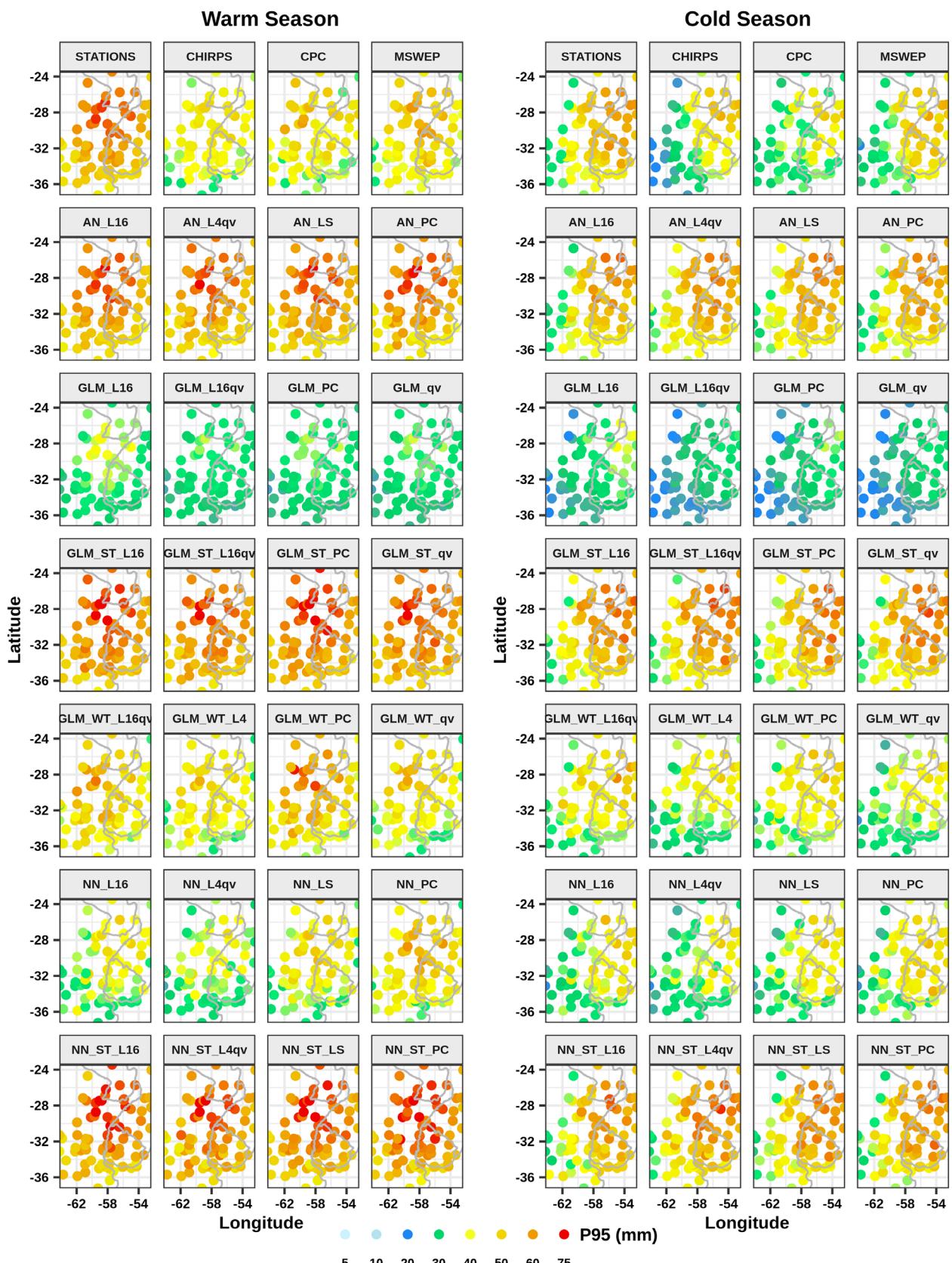


FIGURE 4 Extreme precipitation as depicted by the seasonal averaged 95th percentile (P95) during the reference period 1981–2010, expressed in millimetres for the warm and cold seasons (left and right panels, respectively). Results are presented for each observational dataset and ESD model [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 Annual cycle of the spatially averaged 95th percentile of daily precipitation (P95) over SESA during the reference period 1981–2010. A 7-day moving average was applied for illustrative purposes. Colours indicate the different observational datasets and ESD model families. For each ESD family, the ensemble mean and the range of the ensemble members are shown in solid lines and in shading colours, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

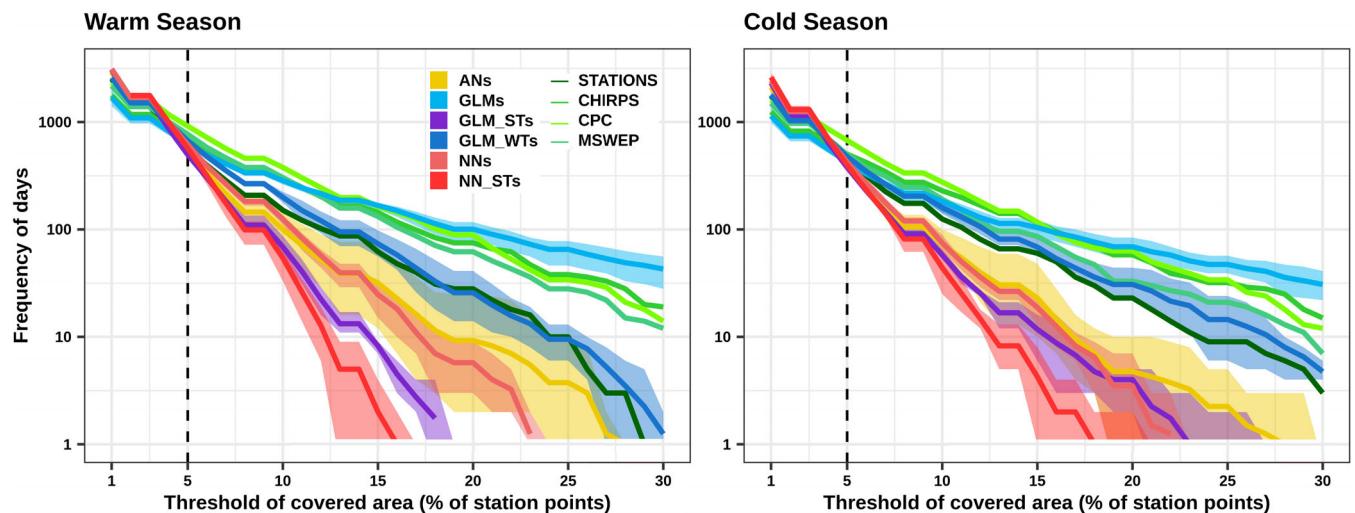
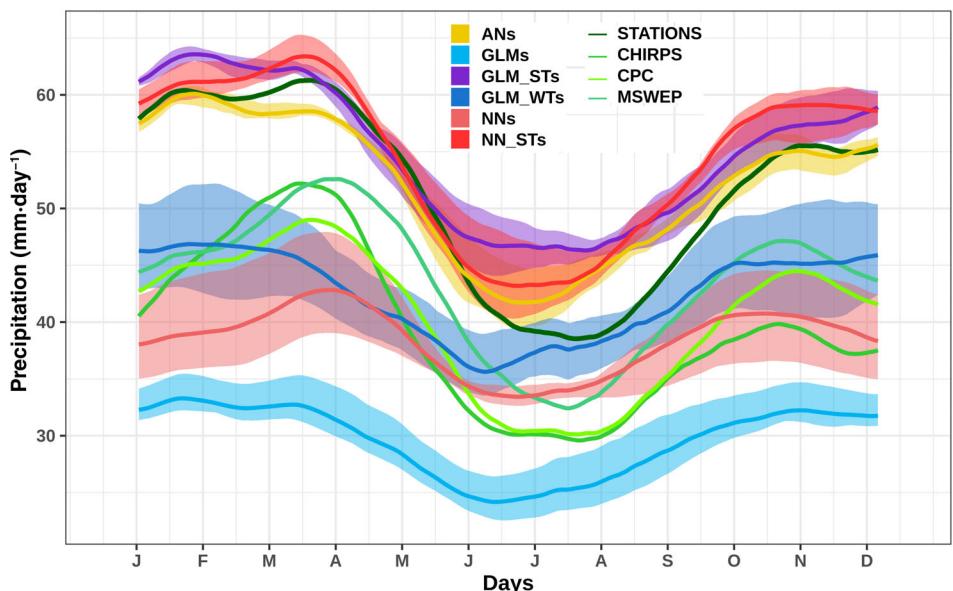


FIGURE 6 Total number of days versus the threshold of covered area, quantified as percentages of station points with daily precipitation above its P95 in SESA for the warm and cold seasons, separately. Colours indicate the different observational datasets and ESD model families. For each ESD family, the ensemble mean and the range of the ensemble members are shown in solid lines and in shading colours, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

occurred—although with large observational uncertainty not only in the intensities of P95 but also in the shape of the P95 evolution during the warm season. It is worthwhile noting that the models spread—not only among different ESD families but also within them—seemed smaller during winter, probably due to a good representation of the physical processes related to precipitation occurrence in the ERA-Interim set of predictors, such as the passage of cold fronts that contributes to the precipitation annual cycle at this time of the year (Espinoza *et al.*, 2013; Cavalcanti *et al.*, 2015).

In order to look into the spatial coverage of daily extreme precipitation in SESA, the total number of days during 1979–2017 according to different thresholds of covered area—given by the percentages of station points with precipitation above the P95—was calculated for the warm and cold seasons, separately (Figure 6). Both seasons showed comparable distributions, although the number of days with the largest number of stations with extreme rainfall was lower during the warm season than in the cold season. This result shows the smaller-scale precipitation systems that dominate the warm season, which agrees with the above discussion of Figure 5. Note

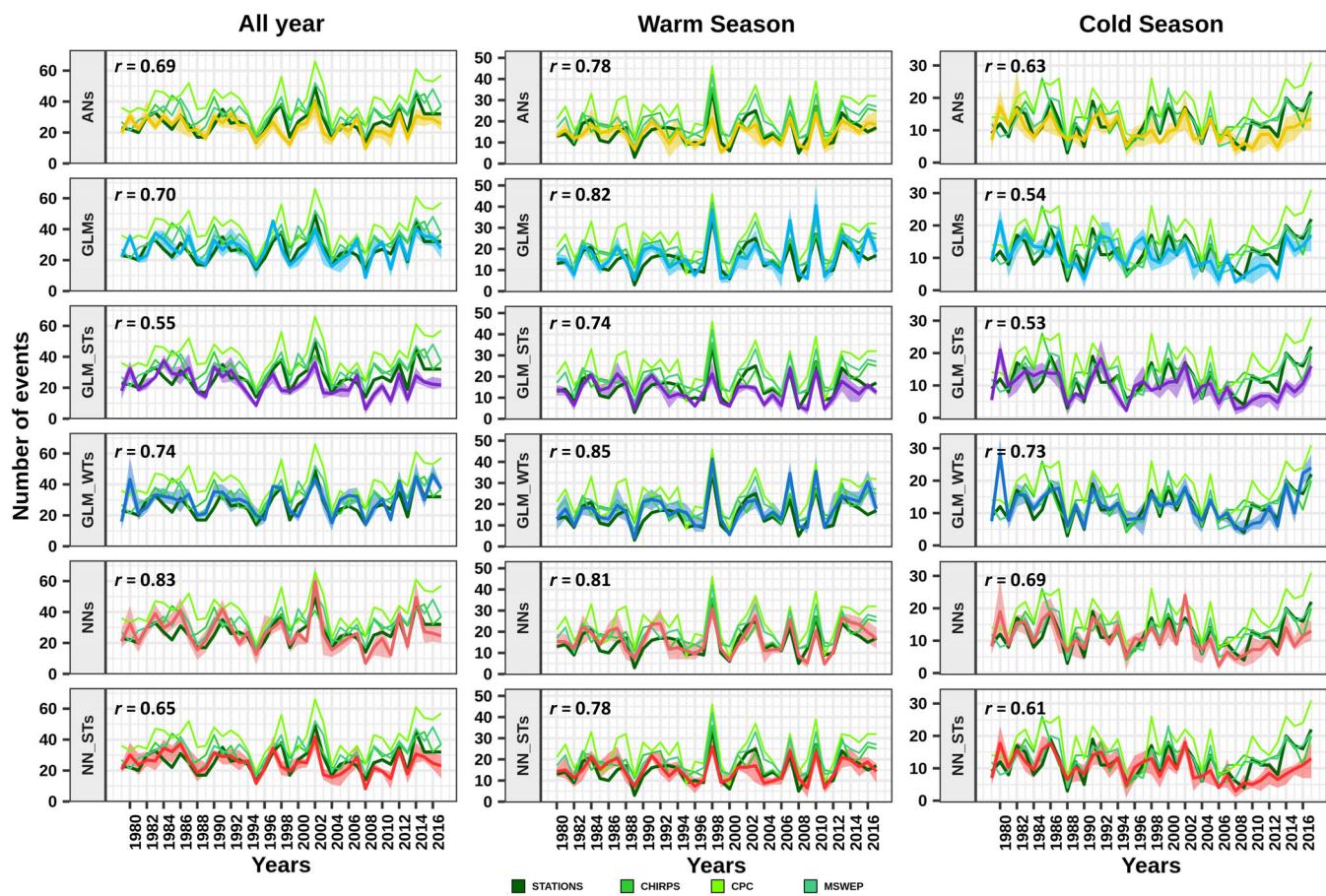


FIGURE 7 Time series of the frequency of extreme precipitation events in SESA (at least 5% of station points) considering the whole year, the warm and cold seasons, separately. Colours indicate the different observational datasets and ESD model families. For each ESD family, the ensemble mean and the range of the ensemble members are shown in solid lines and in shading colours, respectively. Pearson correlation values (r) between the STATIONS reference and each ESD family ensemble mean are included [Colour figure can be viewed at wileyonlinelibrary.com]

that most models—and gridded observations—did not capture this seasonality, except for the GLM_WTs and GLM_STs. This could be due to the model calibration procedure not considering the individual seasons separately. In the case of the gridded precipitation products, a general overestimation of the number of days with a large covered area was observed, probably associated with the comparison procedure with the meteorological stations. Given that the closest grid cell to each station was selected to compare with the meteorological stations, some values of the gridded products could be repeated in nearby stations and, therefore, producing an overestimation of the covered area. ESD models were able to simulate the spatial behaviour of rainfall extremes for low percentages of station points, while larger differences emerged when increasing the number of stations with extreme precipitation at the same time. Particularly, the ANs ensemble exhibited a large spread and the ensemble mean tended to underestimate the number of days with increasing spatial coverage. It is interesting to mention

that—although this cannot be seen in the ANs ensemble—the model AN_PC closely followed the STATIONS distribution even for spread wide spatial extents, which was probably related with the spatial-wide predictors configuration. The GLM_WTs family showed similarities among them and adequately reproduced the STATIONS, indicating as in the former analyses a clear upgrade of the simple GLMs. GLMs tended to overestimate the frequencies of days with increasing spatial extent. However, they were found to underestimate the frequencies of daily precipitation above different absolute thresholds (Figure 3). Recall that the definition of extreme precipitation was relative to each model or dataset since it was based on their own percentiles. The spatial behaviour of the deterministic GLMs has been pointed out by Solman *et al.* (2021), who also found that these models tended to smooth the intensity over the core precipitation area, broadening its pattern. The GLM_STs presented an opposite behaviour, underestimating the total number of days with increasing covered area.

In the case of the neural networks, their deterministic version presented comparable results to the analogues, underestimating the number of days with increasing numbers of stations with precipitation above their P95, which was more pronounced in the NN_STs.

Thereby, an extreme precipitation event was defined using a minimum threshold of 5% of station points exceeding their P95, detecting 577 (460) events during the warm (cold) season in the meteorological stations. This percentage allows the analysis of more structured heavy precipitation systems and also events with a greater spatial extension, in the same way as in Olmo and Bettolli (2021) when analysing these events as simulated by a set of RCMs. In addition, the intra-annual variability of these spatially extended extreme events showed good correspondence with the models' performance as depicted by the P95 annual cycle displayed in Figure 5 (not shown).

3.3 | Interannual variability

The frequencies of extreme precipitation events in SESA (at least 5% of station points) are illustrated in Figure 7 considering the whole year, the warm and cold seasons separately. ESD ensembles are shown by families and compared with the STATIONS reference and the other observational datasets. Additionally, the Pearson correlation (r , included in the panels) between the mean of each ESD ensemble and STATIONS was calculated as a measure of correspondence. A first insight into the observational datasets exhibited an overestimation of the observed frequencies of extreme events by the precipitation products when compared to STATIONS, although the interannual variability was generally well captured, especially in the warm season. The analogues tended to underestimate the frequencies of extreme events, which could be related to the underestimation of the rain frequency in the wet fold as exposed in R01 results from Figure 2. Particularly, they underestimated the maximum number of events during the 1998 warm season and exhibited less correspondence with the observations during the winter season. These outcomes were in line with Bettolli and Penalba (2018), who showed that the analogues had some difficulties in reproducing the interannual variability of seasonal precipitation over the region. In the case of the GLM_WTs and GLMs, they often overestimated the maximum frequencies during the warm season, like in the years 1998, 2007 and 2010. The temporal variability of the regional series was generally well represented by these models as some of the highest correlation values with STATIONS were found for these models during the different seasons, especially

in the circulation-conditioned GLMs. The GLM_STs presented more discrepancies in the number of events per year but managed to simulate the inter-annual variability of extreme events. The NNs models presented similar results than the GLMs and their ensemble mean exhibited the highest correlation for the series considering the whole year. The NN_STs showed lower correlation values than the NNs but always higher than the stochastic GLMs. This good performance of the neural networks in modelling the inter-annual variability was coincident with results by Haylock *et al.* (2006) for daily extreme precipitation indices.

It is worthwhile noticing that the years with the greatest frequency of extreme precipitation events—which were selected for the wet fold in the cross-validation procedure—are characterized by El Niño events. SST anomalies in the central-eastern tropical Pacific and western Indian Ocean typically excite Rossby wave trains extended along the South Pacific towards South America that induce a cyclonic (anticyclonic) circulation anomaly at extratropical (tropical) regions of the continent, enhancing moisture convergence and upward conditions in SESA (Grimm *et al.*, 2000; Penalba and Vargas, 2004; Robledo *et al.*, 2016). Particularly, the years 1982–1983 and 1997–1998 were some of the strongest El Niño events and 1991–1992 also exhibited a notable signal. Contrary, La Niña events—such as 1995–1996 and 2007–2008—were associated with lower numbers of extreme precipitation events over SESA. These maximum and minimum frequencies were well represented by most of the ESD models, which inherited from the reanalysis predictors chosen in this study the main signals present on the large-scale circulation controlled by this oscillation.

3.4 | Wet years

As it was observed in the previous section, the years included in the wet fold were characterized by maximum frequencies of extreme precipitation events over SESA. In the following analysis, focus was put in these years to validate the ESD models performance in terms of the spatial distribution and intensity of the total annual precipitation (PRCPTOT). Figure 8 displays, for each wet year, the spatial fields of PRCPTOT in the meteorological stations, Taylor diagrams presenting the individual performance of each dataset compared to the STATIONS reference and boxplots of the PRCPTOT intensities over SESA in the gridded precipitation products and the statistical models.

The ESD models exhibited an overall good performance in reproducing the spatial pattern of PRCPTOT

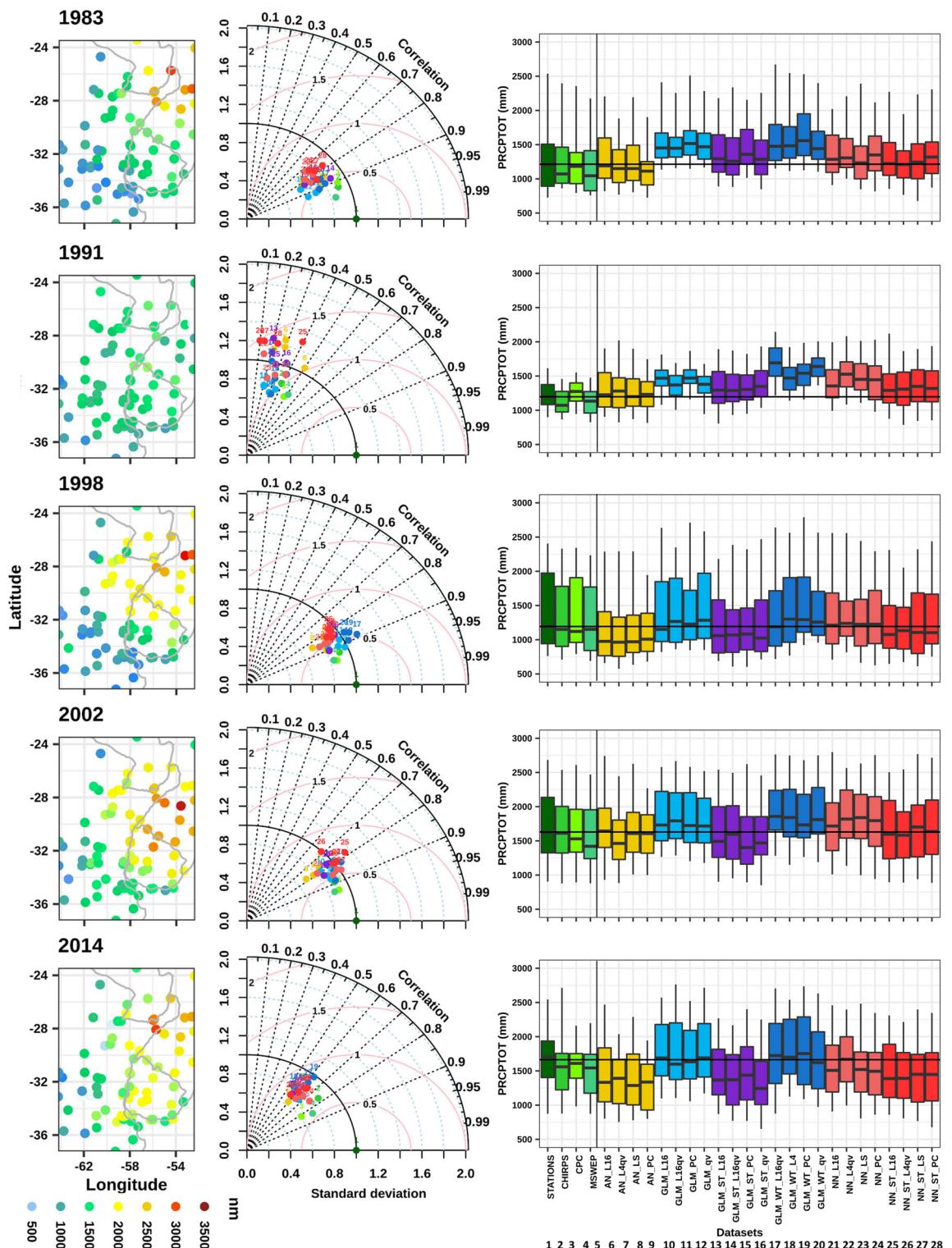


FIGURE 8 Total precipitation amount (PRCPTOT, expressed in millimetres) analysis over SESA during the wet years: from left to right, PRCPTOT spatial patterns of the STATIONS reference; Taylor diagrams of the spatial patterns indicating in colours and numbers the gridded precipitation datasets and ESD models; and boxplots of PRCPTOT intensities over SESA, showing the 25th, 50th and 75th percentiles in their boxes and the 5th and 95th percentiles in their whiskers [Colour figure can be viewed at wileyonlinelibrary.com]

during the wet years, which typically presented maximum intensities up to 3,500 mm in the northeastern portion of the region (Figure 8). Looking into the Taylor

diagrams, the spatial correlations during the years 1998 and 2002 were between 0.7 and 0.9 in all models and precipitation datasets. The ANs and GLMs tended to

underestimate the spatial variability of PRCPTOT, whereas the GLM_WTs and NNs presented normalized standard deviations nearer to one. Particularly, the intense and localized maximums found during 1983 and 2014 were probably responsible for the larger deviations—in terms of standard deviation—from the stations reference as well as the more noticeable difference between the ESD models cloud and the precipitation products in the Taylor diagrams. The exception was the year 1991, which showed reduced PRCPTOT values compared to the other wet years but spanning almost all SESA. This feature affected the models' performance as depicted by the large dispersion among models in the Taylor diagrams and lower spatial correlations detected during this year, when all datasets but most of the analogue models underestimated the spatial variability of PRCPTOT. In this case, AN_PC adequately reproduced the spatial pattern of PRCPTOT.

As depicted by the boxplots of PRCPTOT intensities (Figure 8), consistent results were generally found in the multiple precipitation datasets both in median and extreme values. The ANs reasonably reproduced the observed intensities over SESA, although they underestimated them during 1998 and 2014. GLM_WTs and GLMs tended to overestimate PRCPTOT values such as during 1983 and 1991. This was consistent with the overestimation of R20 (Figure 2) and with the misrepresentation of the precipitation accumulated frequency distribution by the GLMs as seen in Figure S1. Thus, the total annual precipitation was overestimated by these models mainly due to more frequent non-extreme simulated events than observed. The deterministic neural network models typically showed congruent results to the GLMs. However, they adequately simulated the intensities, like in 1983 and 1998. Additionally, the stochastic versions of GLMs and NNs adequately reproduced the PRCPTOT values, similarly to the other models, although some overestimations (underestimations) were detected during 1991 (2014). Hence, despite the differences in the ESD simulations evidenced in these analyses, the statistical models were able to reproduce these wet years in SESA.

Following this evaluation, the ESD multi-model ensemble of the daily regional averaged time series over SESA were illustrated for each wet year and compared to the STATIONS reference (Figure 9). The 24 statistical models used throughout this paper were considered in the construction of the ensemble to account for model uncertainty. It is interesting to mention that the maximum areal intensities occurred—for the different wet years—not only during the warm season (such was the case of 2002, particularly during spring) but also during the cold season. The latter was the case of the year 1991,

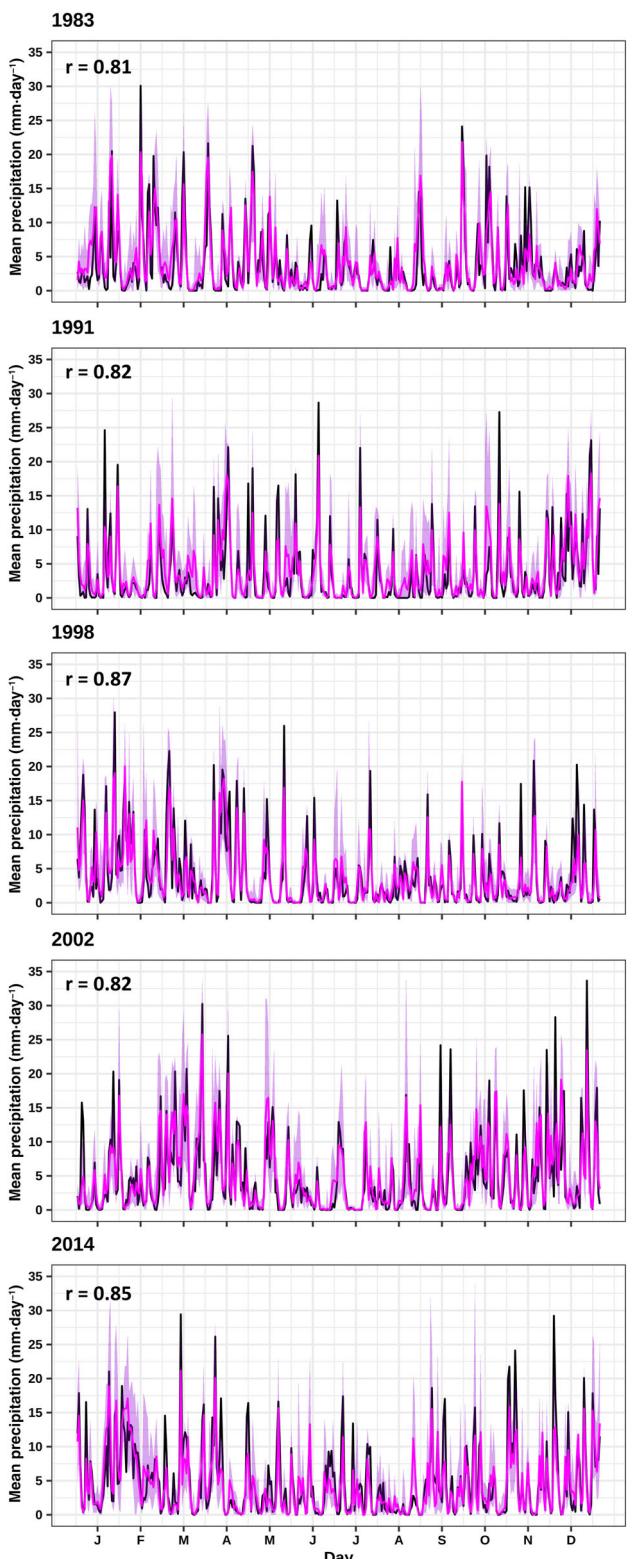


FIGURE 9 Daily areal mean precipitation time series over SESA for each wet year as depicted by STATIONS (black line) and the multi-model ensemble mean, considering all the ESD simulations (pink line). The shading indicates the range of the ESD multi-model ensemble members. Pearson correlation values (r) between the observations and the ESD ensemble mean are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

when the extreme events occurred throughout the year, but the maximum intensities were detected during the winter months. The ESD ensemble successfully captured the temporal variability of daily precipitation quantified by the temporal Pearson correlations between the ESD multi-model ensemble mean and the STATIONS reference above 0.8 and up to 0.87 in 1998 and to 0.85 in 2014. Note that all the ensemble correlations displayed in Figure 9 were statistically significant at the 95% confidence level. Additionally, even though the mean of the multi-model ensemble was expected to underestimate the areal-averaged maximum intensities, these values were usually captured in the spread of the ensemble members (Figure 9). This pointed out the importance of using multiple ESD models when addressing the simulation of daily precipitation extremes in SESA.

In conclusion, the evaluation of the individual years contained in the wet fold evidenced that each of the selected wet years presented different features—such as the spatial distribution and intensity of precipitation—which implies an additional difficulty to any sort of modelling strategy. In this sense, the ESD models were able—despite their limitations—to simulate these features over SESA.

4 | DISCUSSION AND CONCLUSIONS

Regional climate information is one of the primary issues and needs for climate services as they improve coarse resolution global climate models simulations and allow for detailed impact assessments in a climate change scenario. Addressing the study of downscaling methods poses a challenge particularly for designing adaptation and mitigation strategies to precipitation extremes, which are recognized as some of the major threats in a warming climate (IPCC, 2012). In this regard, empirical statistical downscaling tools (ESD) provide detailed information at regional and local scales, taking much less computational demands than other downscaling procedures. In this study, multiple ESD techniques were calibrated and validated for daily precipitation estimates—with special focus on extreme events - during the period 1979–2017 in southeastern South America (SESA), a region with remarkable frequency and intensity of extreme rainfall events (Cavalcanti *et al.*, 2015). The models included in this evaluation involved different configurations—with varied predictors schemes—of the analogue method (ANs), deterministic and stochastic versions of the artificial neural networks (NNs and NN_STs) and generalized linear models (GLMs and GLM_STs) and circulation-conditioned GLMs (GLM_WTs). In the cross-validation

procedure, the years with maximum frequencies of extreme rainfall events were grouped in a separate fold to validate the models' performance in particularly wet conditions. In addition, three gridded precipitation products were included to account for observational uncertainty.

An assessment of several metrics and indices evidenced an overall good performance of the statistical models in reproducing several aspects of daily precipitation over SESA (Figure 2). No particular model configuration was found to perform better in all metrics, while the different model families usually presented congruent results. ESD dispersion was usually contained in the spread exposed by the observational datasets. The simplest models in terms of local predictors information and number of variables within each family of ESDs—such as GLM_qv and NN_l4qv—tended to show poorer performances compared to other model configurations with more predictor variables. Among the configurations with local predictors information, the models using predictor variables at the 16 nearest grid-points (L16) presented better results, indicating that a larger spatial coverage is informative for simulating daily precipitation. The model configuration that employed principal components also managed to capture the different aspects of daily precipitation over SESA. All deterministic models but the analogues underestimated heavy precipitation estimates as depicted by the 98th percentile of daily precipitation, in particular the GLMs, which is a typical limitation of these techniques (Hertig *et al.*, 2018). Regarding this, the GLM_WTs outperformed the simple GLMs as they presented more similar values to the observations. Moreover, the inclusion of stochastic versions of the regression-based methods (GLM_STs and NN_STs) notably improved the model performance especially in high and extreme precipitation indices compared to their deterministic approach. Recall that these deterministic regression-based techniques calibrate the mean value towards the observations and therefore the benefit of introducing a stochastic component in these methods arises when simulating extremes. In addition, the performance during the wet fold was in good agreement with the results during the rest of the period, as depicted by the point representing median wet fold conditions typically contained in the boxplots of almost all metrics for each ESD model. This indicated that most of the models designed in this experiment were skilful in simulating a wetter climate. Note, however, that the analogue models, as implemented in this work, are not able to extrapolate non-recorded values in the calibration period, which is an important limitation to overcome in climate change applications. Although additional analyses are needed to further evaluate the stationarity assumption and ESD

limitations, these results are promising for following applications in future GCMs projections over SESA.

When focusing on extreme precipitation according to the 95th percentile of daily precipitation (P95), the analogue models stood out in the representation of the spatial pattern and the spatially aggregated annual cycle of P95 (Figures 4 and 5, respectively). The GLMs strongly underestimated P95 intensities and its spatial variability over SESA. However, the GLM_WTs exposed much better results, particularly the one using spatial wide predictors (GLM_WT_PC). Thereby, introducing weather typing schemes considerably improved the GLMs performance in most of the aspects analysed throughout this study, also highlighting the utility of the synoptic circulation patterns identified in Olmo and Bettolli (2021) when conditioning the GLMs for the dominant atmospheric states, separately. In the case of the NNs, they presented similar outcomes than the GLM_WTs, although they better captured the shape of the annual cycle, especially the P95 values in the transition seasons. Again, the stochastic GLMs and NNs showed better results in both the spatial pattern and intra-annual variability of P95 than their deterministic versions. The models spread seemed smaller during winter, probably associated with superior representation in the ERA-Interim reanalysis of the large-scale atmospheric systems that lead to precipitation occurrence at this time of the year (Espinoza *et al.*, 2013). ESD models were able to simulate the frequency of extreme rainfall events according to different thresholds of covered area (quantified by percentages of station points), particularly for low thresholds, while larger discrepancies emerged when increasing the number of stations with extreme precipitation (Figure 6). The GLM_WTs family ensemble showed small spread and adequately reproduced the STATIONS, whereas larger differences were detected in the simple GLMs and in the neural networks models.

A spatial threshold of 5% of station points with extreme rainfall in SESA was selected to define days with widespread extreme precipitation events. The inter-annual variability of the frequency of these events in SESA was usually well reproduced by the statistical models (Figure 7). The temporal variability was in good agreement in the different datasets particularly during the warm season, although the analogues exhibited less correspondence with the observations particularly during the warm season. A strong ENSO signal was detected in the observed series, with maximum (minimum) frequencies of extreme rainfall events occurring in years with the El Niño (La Niña) event. These frequencies were captured by most of the ESD models, which seemed to inherit from the reanalysis predictors the main signals presented on the large-scale circulation controlled by this

oscillation. In general, the wet years selected for model calibration and validation in the ESD experiment were strongly influenced by this teleconnection. Prior research has thoroughly investigated the influence of El Niño–Southern Oscillation on precipitation variability in the region. The occurrence of extreme rainfall events in SESA has a powerful connection with El Niño events (Barreiro, 2010; Robledo *et al.*, 2016), and therefore the spatio-temporal variability of this oscillation would probably be related with the future expectations on precipitation extremes. Even more, ENSO variability over the last five decades was about 25% stronger than during the pre-industrial period, which supports the idea that ENSO extremes would be intensified under larger greenhouse gases concentrations (Grothe *et al.*, 2020). Notwithstanding, due to a nonlinear response of the oscillation to surface global warming, there is still large model spread and uncertainty in the results (Power *et al.*, 2013). These statements reinforce the clear influence and modulation of precipitation extremes in SESA by this teleconnection, which should be carefully considered when performing a statistical downscaling of global climate models (GCMs). In this context, the ENSO teleconnection patterns should be well represented by the GCMs to fulfil the perfect prognosis assumption.

The individual evaluation of the wet years showed that most of the models managed to simulate the total annual precipitation both in its spatial pattern and intensities over SESA (Figure 8), generally exhibiting similar results than the precipitation products. Additionally, the multi-model ensemble successfully captured the correct timing of daily precipitation and the spread captured the maximum intensities on the areal-averaged time series (Figure 9).

These results widen our knowledge on the ESD models strengths and limitations in representing daily precipitation and extreme events over SESA and contribute to the still few studies assessing their performance in the region. However, many other aspects of ESD models, such as the reproduction of the spatial and temporal structure of precipitation extremes at sub-regional scales, the long-term changes, the inter-variable dependences and the sensitivity to reanalysis choice, are still open issues. Moreover, these findings based on historical observational records may not be sufficient to determine if these statistical models will correctly represent future climate conditions. Further evaluations of the ESD techniques and their application to historical outputs of GCMs are required to ascertain our confidence in future precipitation changes over SESA, particularly for extremes. In this line, follow-up studies will comprehensively analyse the ESD potential added value on GCMs

simulations and their future projections over the region during the 21st century.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

Matías Ezequiel Olmo: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing - original draft; writing-review & editing. **Maria Bettolli:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; validation; visualization; writing - original draft; writing-review & editing.

ORCID

Matías Ezequiel Olmo  <https://orcid.org/0000-0003-3324-9040>

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

REFERENCES

- Ambrizzi, T., Reboita, M.S., Porfirio da Rocha, R. and Llopard, M. (2018) The state-of-the-art and fundamental aspects of regional climate modeling in South America. *Annals of the New York Academy of Sciences*, 1436(1), 98–120. <https://doi.org/10.1111/nyas.13932>.
- Araya-Osset, 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>.
- 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(4), 2109–2124. <https://doi.org/10.5194/gmd-2019-278>.
- Barreiro, M. (2010) Influence of ENSO and the South Atlantic ocean on climate predictability over southeastern South America. *Climate Dynamics*, 35, 1493–1508. <https://doi.org/10.1007/s00382-009-0666-9>.
- Barros, V., Clarke, R. and Silva Dias, P.L. (2006) Climate change in the La Plata Basin. 1st edn. Buenos Aires: CIMA-CONICET.
- Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. and de Roo, A. (2017) MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences Discussions*, 21, 589–615. <https://doi.org/10.5194/hess-21-589-2017>.
- 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: contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, 13(3), 1711–1735. <https://doi.org/10.5194/gmd-2019-224>.
- Benestad, R. (2010) Downscaling precipitation extremes: correction of analog models through PDF predictions. *Theoretical and Applied Climatology*, 100, 1–21. <https://doi.org/10.1007/s00704-009-0158-1>.
- Berbery, E.H. and Barros, V.R. (2002) The hydrological cycle of the La Plata Basin in South America. *Journal of Hydrometeorology*, 3, 630–645.
- 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., Solman, S., da Rocha, R.P., Llopard, M., Gutiérrez, J. M., Fernández, J., Olmo, M., Lavín-Gullón, A., Chou, S.C., Carneiro Rodrigues, D., Coppola, E., Balmaceda Huarte, R., Barreiro, M., Blázquez, J., Doyle, M., Feijoó, M., Huth, R., Machado, L. and Vianna Cuadra, S. (2021) The CORDEX Flagship Pilot Study in southeastern South America: A comparative study of statistical and dynamical downscaling models in simulating daily extreme precipitation events. *Climate Dynamics*, 56, 1589–1608. <https://doi.org/10.1007/s00382-020-05549-z>.
- Blázquez, J. and Solman, S.A. (2020) Multiscale precipitation variability and extremes over South America: analysis of future changes from a set of CORDEX regional climate model simulations. *Climate Dynamics*, 55, 2089–2106. <https://doi.org/10.1007/s00382-020-05370-8>.
- Casanueva, A., Herrera, S., Fernández, J. and Gutiérrez, J.M. (2016) Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative. *Climatic Change*, 137(3–4), 411–426. <https://doi.org/10.1007/s10584-016-1683-4>.
- Castellano, C.M. and De Gaetano, A. (2016) A multi-step approach for downscaling daily precipitation extremes from historical analogues. *International Journal of Climatology*, 36, 1797–1807.
- Cavalcanti, I. (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, 27–40.
- Cavalcanti, I., Carril, A., Penalba, O., Grimm, A.M., Menéndez, C., Sánchez, E., Cherchi, A., Sörensson, A., Robledo, F., Rivera, J., Pántano, V., Bettolli, M.L., Zaninelli, P., Zamboni, L., Tedeschi, R., Domínguez, M., Ruscica, R. and Flach, R. (2015) Precipitation extremes over La Plata Basin—review and new results from observations and climate simulations. *Journal of Hydrology*, 523, 211–230. <https://doi.org/10.1016/j.jhydrol.2015.01.028>.
- Cavazos, T. (1999) Large-scale circulation anomalies conducive to extreme events and simulation of daily rainfall in northeastern Mexico and southeastern Texas. *Journal of Climate*, 12, 1506–1523.
- Cerón, W.L., Kayano, M.T., Andreoli, R.V., Avila-Diaz, A., Ayes, I., Freitas, E.D., Martins, J.A. and Souza, R.A.F. (2020) Recent intensification of extreme precipitation events in the La Plata Basin in southern South America (1981–2018). *Atmospheric Research*, 249, 105299. <https://doi.org/10.1016/j.atmosres.2020.105299>.
- 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., Källberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>.
- 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. <https://doi.org/10.1007/s10584-009-9738-4>.
- Espinosa, J.C., Ronchail, J., Lengaigne, M., Quispe, N., Silva, Y., Bettolli, M.L., Avalos, G. and Llacza, A. (2013) Revisiting wintertime cold air intrusions at the east of the Andes: propagating features from subtropical Argentina to Peruvian Amazon and relationship with large-scale circulation patterns. *Climate Dynamics*, 41, 1983–2002. <https://doi.org/10.1007/s00382-012-1639-y>.
- Frost, A.J., Charles, S.P., Timbal, B., Chiew, F.H.S., Mehrotra, R., Nguyen, K.C., Chandler, R.E., McGregor, J.L., Fu, G., Kirono, D.G.C., Fernandez, E. and Kent, D.M. (2011) A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. *Journal of Hydrology*, 408, 1–18. <https://doi.org/10.1016/j.jhydrol.2011.06.021>.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. (2014) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066. <https://doi.org/10.1038/sdata.2015.66>.
- Grimm, A.M., Barros, V.R. and Doyle, M. (2000) Climate variability in southern South America associated with El Niño and La Niña events. *Journal of Climate*, 13, 35–58.
- Grothe, P.R., Cobb, K.M., Liguori, G., Di Lorenzo, E., Capotondi, A., Lu, Y., Cheng, H., Edwards, R.L., Southon, J.R., Santos, G.M., Deocampo, D.M., Lynch-Stieglitz, J., Chen, T., Sayani, H.R., Thompson, D.M., Conroy, J.L., Moore, A.L., Townsend, K., Hagos, M., O'Connor, G. and Toth, L.T. (2020) Enhanced El Niño–Southern Oscillation variability in recent decades. *Geophysical Research Letters*, 47, e2019GL083906. <https://doi.org/10.1029/2019GL083906>.
- 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., San Martín, D., Brands, S., Manzanas, R. and Herrera, S. (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, 26, 171–188. <https://doi.org/10.1175/JCLI-D-11-00687.1>.
- Haylock, M.R., Cawley, G.C., Harpham, C., Wilby, R.L. and Goodess, C.M. (2006) Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology*, 26, 1397–1415. <https://doi.org/10.1002/joc.1318>.
- Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J.M., Wibig, J., Casanueva, A. and Soares, P.M.M. (2018) Comparison of statistical downscaling methods with respect to extreme events over Europe: validation results from the perfect predictor experiment of the COST Action VALUE. *International Journal of Climatology*, 39, 3846–3867. <https://doi.org/10.1002/joc.5469>.
- Horton, P. and Brönnimann, S. (2018) 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., 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. (2012) In: Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.-K., Allen, S.K., Tignor, M. and Midgley, P.M. (Eds.) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. New York, USA: Cambridge University Press.
- Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M.D., Manzanas, R., San-Martín, D., Cimadevilla, E., Cofiño, A.S. and Gutiérrez, J.M. (2019) The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling and Software*, 111, 42–54. <https://doi.org/10.1016/j.envsoft.2018.09.009>.
- Lavín-Gullón, A., Feijoó, M., Solman, S., Fernández, J., da Rocha, R.P. and Bettolli, M.L. (2021) Synoptic forcing associated with extreme precipitation events over southeastern South America as depicted by a CORDEX FPS set of convection-permitting RCMs. *Climate Dynamics*, 56, 3187–3203. <https://doi.org/10.1007/s00382-021-05637-8>.
- Li, S., Otto, F.E.L., Harrington, L.J., Sparrow, S.N. and Wallom, D.C. (2020) A pan-South-America assessment of avoided exposure to dangerous extreme precipitation by limiting to 1.5°C warming. *Environmental Research Letters*, 15, 054005.
- 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(1–34), RG3003. <https://doi.org/10.1029/2009RG000314>.
- Maraun, D., Widmann, M. and Gutiérrez, J.M. (2018) 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>.

- Maraun, D., Widmann, M., Gutierrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. I. (2015) VALUE—a framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1), 1–14. <https://doi.org/10.1002/2014EF000259>.
- Menéndez, C.G., de Castro, M., Boulanger, J.P., D'Onofrio, A., Sanchez, E., Sörensson, A.A., Blazquez, J., Elizalde, A., Jacob, D., le Treut, H., Li, Z.X., Núñez, M.N., Pessacq, N., Pfeiffer, S., Rojas, M., Rolla, A., Samuelsson, P., Solman, S.A. and Teichmann, C. (2010) Downscaling extreme month-long anomalies in southern South America. *Climatic Change*, 98, 379–403. <https://doi.org/10.1007/s10584-009-9738-4>.
- 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. <https://doi.org/10.1016/j.wace.2020.100267>.
- Olmo, M.E. and Bettolli, M.L. (2021) Extreme daily precipitation in southern South America: statistical characterization and circulation types using observational datasets and regional climate models. *Climate Dynamics*. <https://doi.org/10.1007/s00382-021-05748-2>.
- Penalba, O.C. and Robledo, F. (2010) Spatial and temporal variability of the frequency of extreme daily rainfall regime in the La Plata Basin during the 20th century. *Climatic Change*, 98(3), 531–550.
- Penalba, O.C. and Vargas, W. (2004) Interdecadal and interannual variations of annual and extreme precipitation over central-northeastern Argentina. *International Journal of Climatology*, 24, 1565–1580. <https://doi.org/10.1002/joc.1069>.
- Power, S., Delage, F., Chung, C., Kociuba, G. and Keay, K. (2013) Robust twenty-first-century projections of El Niño and related precipitation variability. *Nature*, 502, 541–545. <https://doi.org/10.1038/nature12580>.
- Rasmusson, K.L. and Houze, R.A., Jr. (2016) Convective initiation near the Andes in subtropical South America. *Monthly Weather Review*, 144, 2351–2374.
- Re, M. and Barros, V. (2009) Extreme rainfalls in SE South America. *Climatic Change*, 96, 119–136.
- Robledo, F., Vera, C. and Penalba, O. (2016) Influence of the large-scale climate variability on daily rainfall extremes over Argentina. *International Journal of Climatology*, 36, 412–423. <https://doi.org/10.1002/joc.4359>.
- Salio, P., Nicolini, M. and Zipser, E.J. (2007) Mesoscale convective systems over southeastern South America and their relationship with the South American low level jet. *Monthly Weather Review*, 135, 1290–1309. <https://doi.org/10.1175/MWR3305.1>.
- San Martín, D., Manzanas, R., Brands, S., Herrera, S. and Gutiérrez, J.M. (2017) Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *Journal of Climate*, 30, 203–223. <https://doi.org/10.1175/JCLI-D-16-0366.1>.
- Schmidli, J., Goodess, C.M., Frei, C., Haylock, M.R., Hundecha, Y., Ribalaygua, J. and Schmitt, T. (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research*, 112, D04105. <https://doi.org/10.1029/2005JD007026>.
- Sillmann, J., Kharin, V.V., Zwiers, F.W., Zhang, X. and Bronaugh, D. (2013) Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, 118, 2473–2493. <https://doi.org/10.1002/jgrd.50188>.
- Solman, S.A., Bettolli, M.L., Doyle, M.E., Olmo, M.E., Feijoo, M., Martínez, D., Blázquez, J. and Balmaceda-Huarte, R. (2021) Evaluation of multiple downscaling tools for simulating extreme precipitation events over southeastern South America: a case study approach. *Climate Dynamics*. <https://doi.org/10.1007/s00382-021-05770-4>.
- Solman, S.A. and Blázquez, J. (2019) Multiscale precipitation variability over South America: analysis of the added value of CORDEX RCM simulations. *Climate Dynamics*, 53, 1547–1565. <https://doi.org/10.1007/s00382-019-04689-1>.
- Sörensson, A.A., Menéndez, C.G., Ruscica, R., Alexander, P., Samuelsson, P. and Willén, U. (2010) Projected precipitation changes in South America: a dynamical downscaling within CLARIS. *Meteorologische Zeitschrift*, 19, 347–355. <https://doi.org/10.1127/0941-2948/2010/0467>.
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719>.
- Timbal, B., Fernandez, E. and Li, Z. (2009) Generalization of a statistical downscaling model to provide local climate change projections for Australia. *Environmental Modelling & Software*, 24, 341–358.
- Vörösmarty, C.J., de Guenni, L.B., Wolheim, W.M., Pellerin, B., Bjerklie, D., Cardoso, M., D'Almeida, C., Green, P. and Colon, L. (2013) Extreme rainfall, vulnerability and risk: a continental-scale assessment for South America. *Philosophical Transactions of the Royal Society A*, 371, 20120408. <https://doi.org/10.1098/rsta.2012.0408>.
- Xie, P., Chen, M. and Shi, W. (2010) *CPC global unified gauge-based analysis of daily precipitation*. Boulder, CO: NOAA/OAR/ESRL PSL.
- Zorita, E. and von Storch, H. (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, 12, 2474–2489. <https://doi.org/10.1175/1520-0442>.

SUPPORTING INFORMATION

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

How to cite this article: Olmo, M. E., & Bettolli, M. L. (2022). Statistical downscaling of daily precipitation over southeastern South America: Assessing the performance in extreme events. *International Journal of Climatology*, 42(2), 1283–1302. <https://doi.org/10.1002/joc.7303>