



Climate predictability on seasonal timescales over South America from the NMME models

Luciano G. Andrian^{1,2,3} · Marisol Osman^{1,2,3,4} · Carolina S. Vera^{1,2,3}

Received: 21 January 2022 / Accepted: 12 September 2022 / Published online: 24 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This paper evaluates the predictability and skill of the models from the North American Multi-Model Ensemble project (NMME) in South America on seasonal timescales using analysis of variance (ANOVA). The results show that the temperature variance is dominated by the multi-model ensemble signal in the austral autumn and summer and by the inter-model biases in the austral spring and winter. The temperature predictability is higher at low latitudes, although moderate values are found in extratropical latitudes in the austral spring and summer. The predictability of precipitation is lower than that of temperature because noise dominates the variance. The highest levels of precipitation predictability are reached in tropical latitudes with large inter-seasonal variations. Southeastern South America and Patagonia present the highest predictability at midlatitudes. The NMME skill of temperature is better than that of precipitation, and it is better at low latitudes for both variables. At extratropical latitudes, the skill is moderate for temperature and low for precipitation, although precipitation reaches a local maximum in southeastern South America.

Keywords Climate prediction · Seasonal forecast skill · Multi-model Ensemble

1 Introduction

The development of regional climate services requires understanding the predictability and prediction skill of climate variability on different timescales. Nowadays, seasonal climate predictions are routinely produced by operational centers and internationally coordinated activities worldwide, using a multi-model ensemble (MME) of coupled general circulation models (CGCMs) to address the uncertainties associated with the chaotic nature of the atmosphere and

the errors arising due to the initial conditions as well as numerical formulation of the dynamical models used. However, unlike other parts of the world, in South America not many works address the levels of predictability achieved by a MME of CGCMs on seasonal timescales.

Many previous studies on predictability in South America are limited by being performed with Atmospheric General Circulation Models (AGCMs) (e.g. Barreiro et al. 2002; Barreiro 2010; Nobre et al. 2006; Taschetto and Wainer 2008) or with a single coupled model (e.g. Barreiro 2010; Bombardi et al. 2018; Gubler et al. 2020). One of the few studies carried out in South America with an ensemble of CGCM was performed by Coelho et al. (2006) using three models from the Development of a European Multi-model Ensemble system for seasonal to interannual prediction (DEMETER) project. In this work they addressed the seasonal predictability of the precipitation in summer during El Niño Southern Oscillation (ENSO) events. The most predictable regions were found during ENSO events in Southern Brazil, Paraguay, Uruguay, and Northern Argentina. More recently, Osman and Vera (2017) analyzed the predictability in winter and summer with the CGCMs of the Climate Historical Forecast Project (CHFP, Tompkins et al. 2017). They found that the temperature predictability is higher than the

✉ Luciano G. Andrian
luciano.andrian@cima.fcen.uba.ar

¹ Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
² CONICET-Universidad de Buenos Aires, Centro de Investigaciones del Mar y la Atmósfera (CIMA), Buenos Aires, Argentina
³ CNRS-IRD-CONICET-UBA, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL3351 IFAECI), Buenos Aires, Argentina
⁴ Present Address: Karlsruhe Institute of Technology, Karlsruhe, Germany

precipitation predictability in both seasons, especially in the tropics. Precipitation predictability is higher in extratropical regions than temperature predictability, and the values are lower than in the tropics but still useful for potential applications. These regions are the Southern Andes in winter and summer, the extratropical Andes in winter, and Southeastern South America (SESA, 39 °S–17 °S; 296 °E–315 °E) in summer. However, these studies did not include the transitional seasons, of which there is very little knowledge about predictability in South America. Transitional seasons, especially spring, are relevant because ENSO, the main regional climate driver on seasonal timescales, has a strong impact in South America during this season (Cai et al. 2020), and therefore predictability could be high.

The North American Multi-Model Ensemble project (NMME, Kirtman et al. 2014) is an operational multi-model forecast system consisting of CGCMs belonging to the USA and Canada modeling centers available to the community through the International Research Institute for Climate and Society (IRI) Data Library (IRIDL). The NMME has shown promising applications in various parts of the world (Cash et al. 2019; Rodrigues et al. 2019; Slater et al. 2019; Zhao et al. 2019), and is continuously being improved (Becker et al. 2020). Nevertheless, only a few studies and applications exist in South America (Osman et al. 2021), and the study of the climate predictability levels achieved there by the NMME has not yet been performed. The global predictability of the NMME for temperature and precipitation has been assessed by Becker et al. (2014). This study shows that predictability in South America is generally high at low latitudes throughout the year and for both variables (temperature and precipitation). At middle and sub-polar latitudes, temperature predictability is higher in winter and summer, while precipitation has its maximum in SESA in spring and summer and in northern SESA (29 °S–17 °S; 303 °E–315 °E) during winter. The availability of the NMME reforecast allows the documentation of the predictability in spring and autumn, which were less explored in past studies.

The analysis of variance (ANOVA) technique is helpful to understand the different contributors and their interactions in the total variance of a variable's field. ANOVA has been applied previously by Zwiers (1996) and Zwiers et al. (2000) to assess predictability. In both studies, the variance was decomposed into different components: signal, noise, and the differences between ensemble members. The predictability was defined through the signal-to-noise ratio. More recently, this approach was also successfully applied by Hodson and Sutton (2008) in studying the predictability of sea level pressure simulated by an ensemble of Atmospheric Model Intercomparison Project (AMIP) simulations. Osman and Vera (2020) used this technique to assess the predictability associated with the main modes of variability of the Southern Hemisphere circulation. The potential of ANOVA

for identifying the sources of variability depicted by single and multi-model ensemble motivates its application to diagnose the predictability of temperature and precipitation by NMME models.

The main objective of this work is to quantify and describe the seasonal predictability of precipitation and temperature for four seasons (austral fall, winter, spring, and summer) by the NMME forecast models in South America using the ANOVA technique. The paper is organized as follows: Sect. 2 describes the observed and forecast data, Sect. 3 presents the main results and Sect. 4 discusses the main conclusions.

2 Data and methodology

For verification, 2-meter air temperature from the Global Historical Climatology Network and Climate Anomaly Monitoring System (GHCN + CAMS, Fan and van den Dool 2008) with 1.0° × 1.0° resolution and precipitation from the CPC Merged Analysis of Precipitation (CMAP, Xie and Arkin 1997) were used as verification fields. CMAP data was regridded to a 1.0° × 1.0° using the R-package (R Core Team 2020) “fields” (Nychka et al. 2017).

Seasonal hindcasts from eight models participating in the NMME project were used in this study. Each model's features, such as the ensemble size, atmosphere and ocean models, are detailed in Table 1. The predictability analysis is based on 3-monthly mean outputs of 2-m air [°C] temperature and precipitation rate [mm], both predicted for March–April–May (MAM), June–July–August (JJA), September–October–November (SON), and December–January–February (DJF), with initial conditions observed in February, May, August and November, respectively (lead 1 month) over the period 1982–2010. The analysis is confined to South America (60 °S–15 °N; 265 °E–330 °E).

All figures were made using R and contour plots using the metR package (Campitelli 2021).

2.1 Analysis of variance

A two-way ANOVA with interactions (Storch and Zwiers 1999; Hodson and Sutton 2008) was applied to identify the sources of variability in the MME. A detailed description of the methodology can be found in Storch and Zwiers (1999). The goal of ANOVA is to decompose the total variance of the ensemble into contributions from different factors in order to estimate the potential predictability. If we want to examine a variable X (precipitation or temperature) in a multi-model multi-member ensemble then each of the M models has K ensemble members in a period T of years. Unlike Hodson and Sutton (2008), the number of ensemble members varies between the models. Therefore,

Table 1 Models from NMME used in this study

Model	Institution	Atmospheric Component	Oceanic component	Ensemble size	Hindcast period
NCAR-CCSM4	NCAR	CAM4 0.9 × 1.25° L26	POPL60 .25°	10	(1982–2010)
GFDL-CM2p1	NOAA/GFDL	CM2.1 1 × 2.5° L24	MOM4 L50 .3°Eq	10	(1982–2010)
GFDL-FLOR-A06	NOAA/GFDL	CM2.5 C18L32	MOM5 L50 .3°Eq	12	(1982–2010)
GFDL-FLOR-B01	NOAA/GFDL	CM2.5 C18L32	MOM5 L50 .3°Eq	12	(1982–2010)
NASA-GEOS S2S	NASA Goddard Space Flight Center	GEOS5 AGCM 0.5° L72	MOM5 L40 .5°Eq	4	(1982–2017)
NCEP-CFSv2	NOAA/NCEP	GFS T126L64	MOM4 L40 .25°Eq	24	(1982–2010)
ECCC-CanCM4i	Environment Canada	CanAM4 T63L35	CanOM4 L40 .94°Eq	10	(1981–2018)
ECCC-GEM-NEMO	Environment Canada	GEM256 × 128	NEMO 1 × 1 1/3 Eq.	10	(1981–2018)

K values varies according to the model (as shown in Table 1), and we can define K_m with $m = 1, \dots, M$ as the number of k ensemble members for each m model. Following Hodson and Sutton (2008), the analysis begins by proposing a linear decomposition of a certain data value X_{tmk} (where $t = 1, \dots, T$, m and k denote year, model, and ensemble member, respectively) as follows:

$$X_{tmk} = \mu + \alpha_t + \beta_m + \gamma_{tm} + \epsilon_{tmk} \quad (1)$$

Here, μ represents the time-mean ensemble-mean model-mean, α_t is the time-varying behavior that is common to all members from all models, β_m is the bias of a given model relative to the MME climatology, γ_{tm} is the unbiased time-varying difference between the ensemble-mean of the models and captures the difference in the model's response to the same forcing and ϵ_{tmk} is the residual noise. Concerning the structural term γ_{tm} , it is important to note that, unlike Hodson and Sutton (2008), NMME models do not all have the same initial conditions. It can be thought that the forcing is the same in all cases but that each model represents it differently (differences due to the observations or reanalyses and assimilation methods employed to initialize the models). Therefore, this term captures the variability associated with differences in model responses to forcings, as well as the differences between the initial conditions used to initialize the models. However, it could be the case that the models have systematic (i.e. independent of time) differences associated with the initial conditions used to initialize the models, in which case it would be reflected in the β term. In this study we assume that the influence of initial condition is mainly capture by γ .

Due to the definition of μ ; α_t , β_m and γ_{tm} must satisfy:

$$\sum_t \alpha_t = \sum_m \beta_m = \sum_t \gamma_{tm} = \sum_m \gamma_{tm} = 0 \quad (2)$$

To assess the effect of each term in Eq. 1, the total sum of squares (TSS) is defined as:

$$TSS = \sum_{t,m,k}^{T,M,K_m} (X_{tmk} - X_{ooo})^2 \quad (3)$$

where the sub-index \mathbf{o} indicates the average in the missing sub-index. TSS represents the X_{tmk} variance concerning the mean of the MME. This expression can be decomposed as follows:

$$TSS = SS\alpha + SS\beta + SS\gamma + SS\epsilon$$

where

$$\begin{aligned} SS\alpha &= \sum_t^T K_{total} (X_{too} - X_{ooo})^2 \\ SS\beta &= \sum_m^M TK_m (X_{omo} - X_{ooo})^2 \\ SS\gamma &= \sum_{t,m}^{T,M} K_m (X_{tmo} - X_{too} - X_{omo} + X_{ooo})^2 \\ SS\epsilon &= \sum_{t,m,k}^{T,M,K_m} (X_{tmk} - X_{tmo})^2 \end{aligned} \quad (4)$$

with $K_{total} = \sum_m^M K_m$. From these expressions, one can construct estimators of the terms on the right-hand side of equation 1 and then assess the effect of each of them on the MME. For example, for β_m the ratio:

$$\frac{SS\beta}{(M-1)} \frac{T(K_{total} - M)}{SS\epsilon} \quad (5)$$

is an unbiased estimator (Storch and Zwiers 1999) of:

$$\frac{\frac{TK_{total}}{M-1} \sum_m \beta_m^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \quad (6)$$

Then, it is possible to compare the magnitude of the effects of β_m with respect to the noise. Formally, it allows the formulation of a null (H_0) and alternative hypothesis (H_1):

$$\begin{aligned} H_0 &: \sum_m \beta_m^2 = 0 \\ H_1 &: \sum_m \beta_m^2 \neq 0 \end{aligned} \quad (7)$$

and to perform a Fisher-test with $(M - 1)$ and $T(K_{total} - M)$ degrees of freedom. If the effect is significant, it is possible to ask how much of β_m contributes to the X_{imk} variability (TSS). The latter can be calculated through the adjusted coefficient of multiple determination:

$$R_b^2 = \frac{SS\beta - \frac{M-1}{T(K_{total}-M)}SS\epsilon}{TSS}$$

This coefficient is the fraction of the total variance determined by $SS\beta$. Similar expressions for α_t and γ_{im} can also be constructed.

The potential predictability (PP_t) of the MME was assessed using ANOVA. The PP_t is the fraction of variability that could be predicted given knowledge of the forcing. A measure of PP_t was defined as follows:

$$PP_t = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2} \quad (8)$$

where the variances were calculated from α and ϵ . Following Zwiers et al. (2000), an unbiased estimator of PP_t can be defined as:

$$\hat{PP}_t = \frac{1}{1 + \frac{K_{total}}{\left(\frac{SS\alpha}{SS\epsilon} \frac{T(K_{total}-M)}{T-1} - 1\right)}} \quad (9)$$

and its significance can be tested with a Fisher test, with $(T - 1)$ and $T(K_{total} - M)$ degrees of freedom.

We also analyzed the impact on PP_t of removing one model at a time from the MME. To assess whether the observed differences with respect to the total MME were statistically significant, a Monte Carlo test was performed. The test consists of constructing a random ensemble of the eight NMME models, calculating the PP_t of this new random MME and the difference with respect to the original MME. This process was repeated 10,000 times to compute the 95th percentile based on 10,000 sample values. Differences from the original MME obtained by removing a single model from the MME were compared with the 95th percentile of the test to assess the significance.

2.2 Forecast skill

The forecast skill was assessed by the standard-deviation-normalized root-mean-square error (NRMSE) as follows:

$$NRMSE = 1 - \frac{\sqrt{\frac{1}{T} \sum_t (F_t - \bar{F}_t + \bar{O}_t - O_t)^2}}{\sqrt{\frac{1}{T} \sum_t (O_t - \bar{O}_t)^2}} \quad (10)$$

Here, F and O are the forecast and reference data values for any variable (precipitation or temperature), while \bar{F}_t and \bar{O}_t are the climatological means of each without including year t . Moreover, the skill was also assessed by the anomaly correlation coefficient (ACC) between the MME and the reference data (Wilks 1995). The statistical significance of the ACC was estimated using a one-tailed t-test at 95% of confidence. In all calculations a one-year-out cross-validation was used. In addition, theoretical ACC was calculated as the mean ACC between the MME and the individual models, as the skill score of the perfect model. This skill was used as a measure of the agreement between the models as an alternative estimate of predictability. However, it is important to remember that perfect-model skill should not be considered an upper bound on actual skill (Kumar et al. 2014). The differences between the actual and theoretical ACC were tested through a Monte Carlo test as was done for PP_t (see Sect. 2.1).

3 Results

3.1 MME variance decomposition

Figure 1 shows the fraction of the MME temperature variance explained by each term of Eq. 1 for each season. The dots denote the grid points where the contribution of each term is non-significant according to the Fisher test. As was found in Zwiers et al. (2000), the value that delimits the significance level is usually low, so most of the grid points are significant. Overall, the signal (α) and inter-model bias (β) are the dominant contributors to TSS (Eq. 3). In MAM and DJF, the signal dominates the variability (Fig. 1a, e, i, and m) at low latitudes in northeastern Peru, southeastern Colombia, northern Bolivia, and in central and west-central Brazil (0.6–0.8). At midlatitudes, moderate values (0.4–0.5) are observed in SESA and central Argentina in both seasons. On the other hand, in JJA and SON the variability is dominated by the inter-model biases (Fig. 1b, f, j, and n) in almost the entire continent, especially at low latitudes in SON (0.7–0.8). The contribution of the structural term (γ , Fig. 1c, g, k, and o) to TSS is smaller than the other terms but is significant and presents a high inter-seasonal variability with maximum values over Amazonia in MAM, southeast Brazil

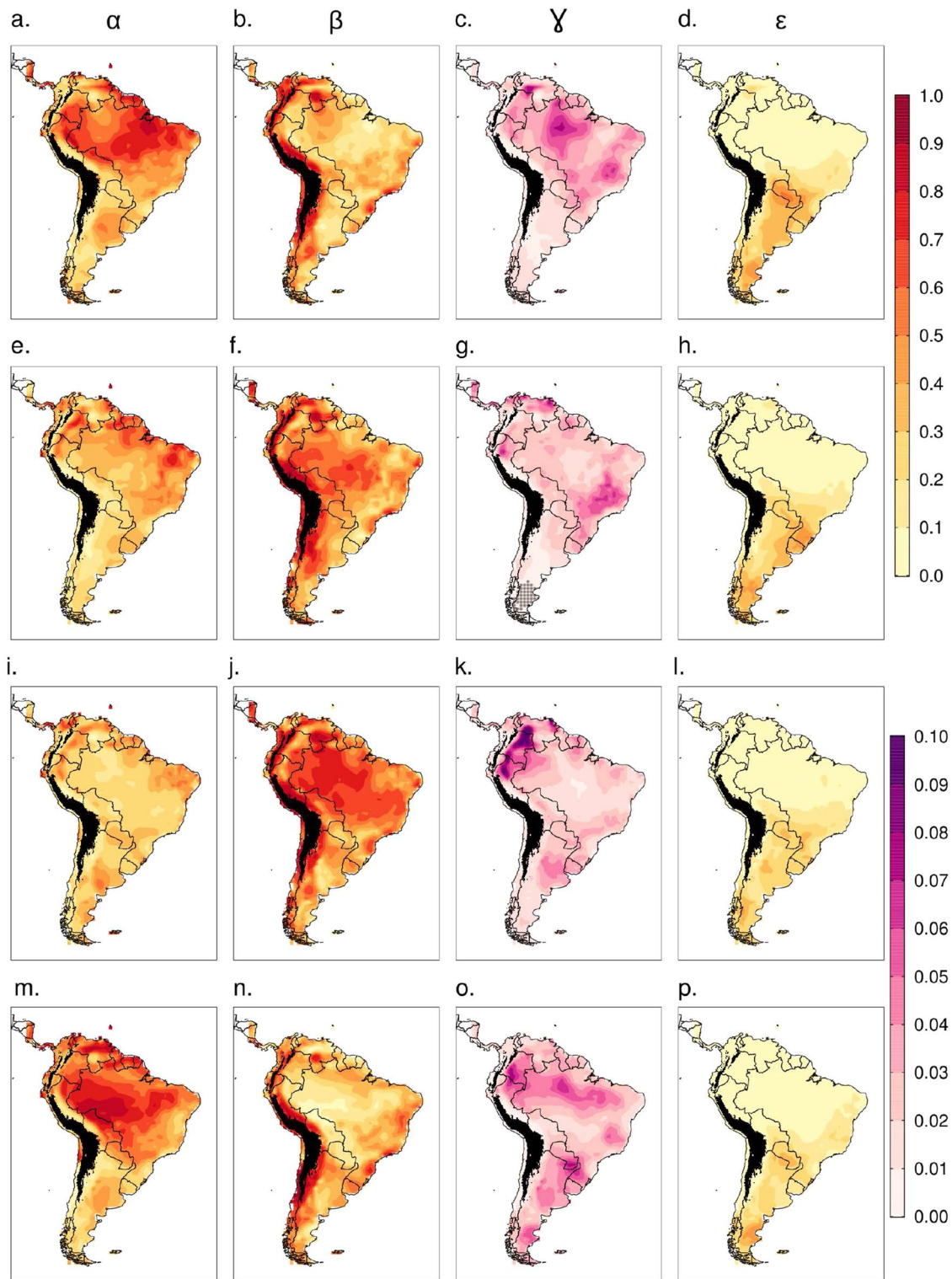


Fig. 1 Fraction of the total variance of the MME forecast temperature explained by each TSS component in MAM, JJA, SON and DJF (row 1, 2, 3 and 4, respectively), represented by α_r (a, e, i, m), β_m (b, f, j, n), γ_m (c, g, k, o) y ϵ_{mk} (d, h, l, p). The upper color bar corresponds to the figures for the α_r , β_m and ϵ_{mk} terms, while the lower color bar

corresponds to the figures for the γ_m term. Dotted areas denote the grid points where the F-test, which compares the variance of the plotted component to the noise term, is not significant at the 90% confidence level

in JJA; over SESA, northern Peru and Colombia in SON; and SESA in DJF. Finally, the fraction of variance associated

with the noise (ϵ , Fig. 1d, h, l, and p) is highest from mid to high latitudes and reaches its maxima in MAM.

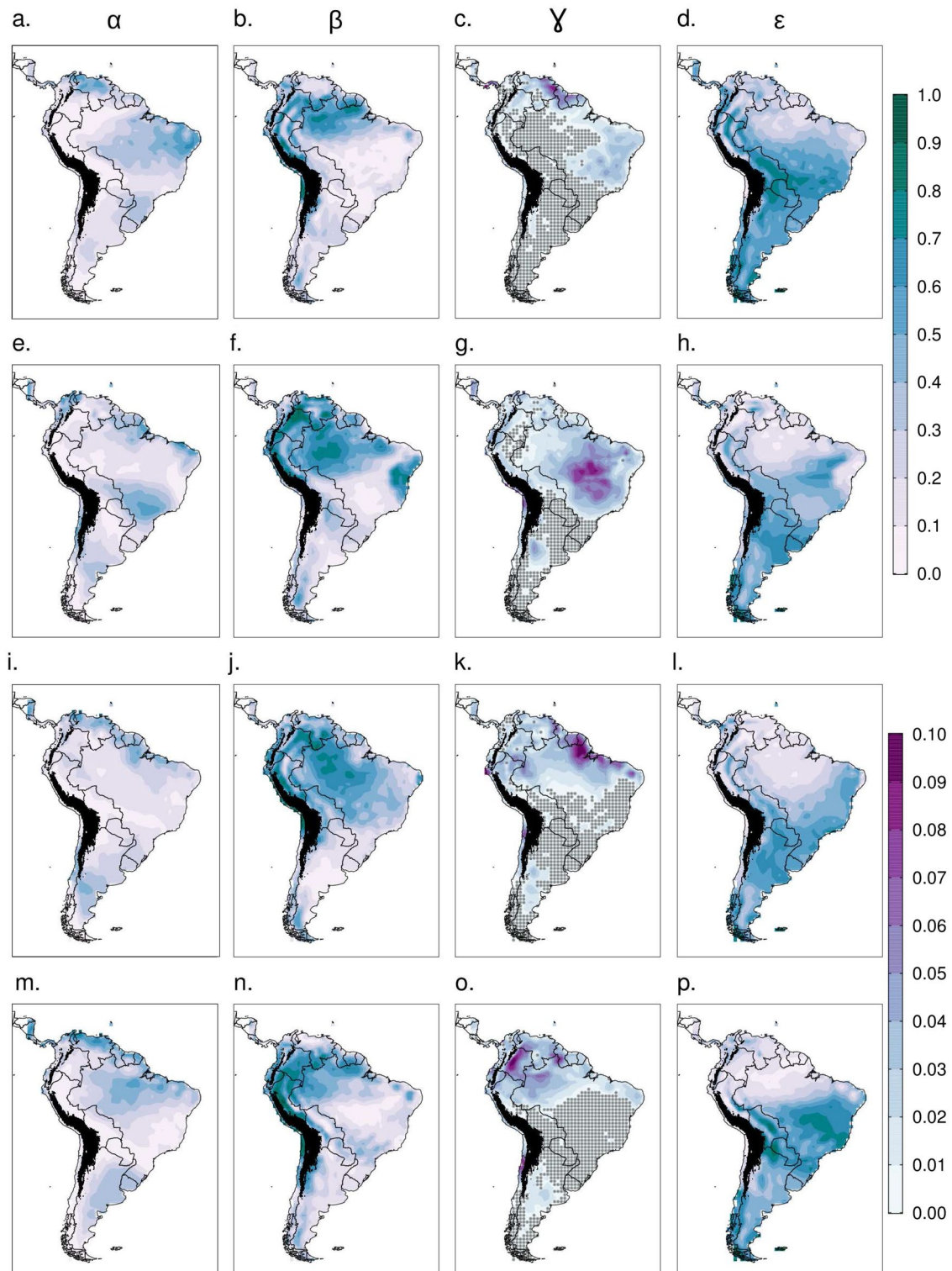


Fig. 2 Same as Fig. 1 but for precipitation

In the case of precipitation, the signal is notably lower than that for temperature across the continent. The regions in which α (Fig. 2a, e, i, and m) explains a large part of the variance (between 0.4 and 0.7) show a high inter-seasonal variation and correspond very well to the areas where precipitation is modulated by ENSO (Cai et al. 2020). However, in SON, a season when ENSO influence on precipitation in SESA is high, the signal does not peak there. The inter-model biases (Fig. 2b, f, j, and n) dominate the precipitation variability at low latitudes over the northern and northwestern portion of the continent and present high inter-seasonal variations over midlatitudes. The structural term (Fig. 2c, g, k, and o) is lower than the other terms and presents non-significant values (dotted areas) in most of the continent and high inter-seasonal variations. The highest γ values are found over eastern Brazil in JJA. Noise (Fig. 2d, h, l and p) dominates the precipitation variability from 10 °S–20 °S towards higher latitudes in all seasons, especially in MAM.

3.2 Model biases and structural term

The bias term of the ANOVA (β) provides information on the consistency between models. A large β means considerable differences in the model's climatologies. In the case of temperature (Fig. 1b, f, j, and n), β presents high values in JJA and SON over central Brazil and throughout the four seasons in regions with high topography, such as the Andes and the Guyana Plateau in southern Venezuela. To assess whether it is possible to attribute the large biases in these regions to specific models, the ANOVA is computed when each model, in turn, is removed from the MME. Based on the analysis of the figures obtained in each case,

we have selected the most outstanding ones. Figure 3a shows the bias term after removing the CanCM4i model from the MME in MAM. The comparison between Figures 1b and 3a reveals that most of the inter-model bias in Amazonia reduces when CanCM4i is removed from the MME: Thus, most of the inter-model bias is attributable to the CanCM4i model. Figure 3b shows the difference between the CanCM4i climatology and the MME (without the CanCM4i model) climatology, where large positive values are observed over the Amazonia, confirming the discrepancies between CanCM4i and the MME without considering it. It is important to note that the bias term provides information about the consistency among models but does not assess any agreement with observations. We then evaluate the extent of such agreement by computing the differences between the climatology of the reference data and the climatology of the CanCM4i model. Figure 3c confirms that the positive bias in the temperature over the Amazonia region is also observed when the model is compared against the reference data, in agreement with previous works (Lin et al. 2020). When the CanCM4i model is removed from the MME, the MME biases in Amazonia decreases (Fig. 3d), suggesting that the model's climate representation in that region may be inadequate. In JJA and SON, a similar reduction of the inter-model biases term across the northern portion of the continent is observed by removing the CM2p1 model from the MME, while in JJA removing the CFSv2 from the MME reduces the bias term in SESA and eastern Brazil (not shown).

For precipitation, the removal of models from the MME has a less consistent impact than for temperature. Removing a given model can reduce the variability explained by the

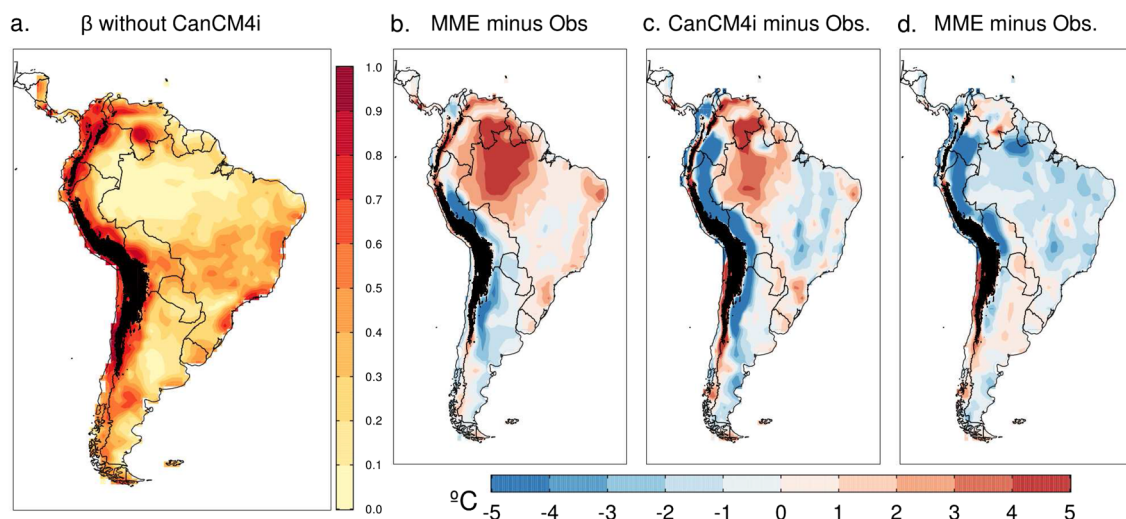


Fig. 3 Fraction of total variance represented by β_m for the MME without CanCM4i model **a** in MAM. Differences in the climatology fields between **(b)** the CanCM4i and the MME without CanCM4i model,

(c) the CanCM4i model and the reference data, and **(d)** the MME without the CanCM4i model and the reference data, in MAM

inter-model bias in the north of the continent and increase it in the south, and vice versa.

The structural term captures the differences in model response to the same forcing and the differences between initial conditions used to initialize the forecasts. To assess each model's impact on this term as well as the inter-model biases, the ANOVA is repeated after each model, in turn, is removed from the MME. For the temperature, the structural term decreases in SON when the CFSv2 model is removed from the MME (Fig. 4a). Most of the changes in the structural term are observed in regions where ENSO modulates the temperature variability, such as central and southeastern Brazil, although not over Colombia and northern Peru. Therefore, the relationship between El Niño 3.4 and temperature in the MME and CFSv2 was explored. For this purpose, the correlation between the sea surface temperature (SST) in the El Niño 3.4 region in August (i.e., the month when the forecasts are issued), and the SON temperature forecast by the MME (Fig. 4b), the SON temperature forecast by CFSv2 model (Fig. 4c), and the observed SON temperature (Fig. 4d), respectively, were calculated. The regions without dots denote the grid points where the correlation is significant according to a t-test with 90% confidence. The model-simulated SST was used for the correlation between the temperature of MME without CFSv2 and the single CFSv2 model, whereas the "NOAA OI V2" SST database (Reynolds et al. 2007) was used for the reference temperature. The correlation between SST and the temperature of the MME without the CFSv2 model is significant on almost the entire continent (Fig. 4b). Conversely, in the north and south of the continent, the correlation between the SST and the CFSv2 model is weaker than that between the

SST and the MME. Moreover, in the center of the continent, the sign of these correlations differs (Fig. 4c). While the MME without the CFSv2 model seems to overestimate the ENSO signal (as observed in Cash and Burls (2019)), the CFSv2 model also differs from the reference data (Fig. 4d). These results suggest that the CFSv2 model contribution to the structural term in the MME may be due in part to an inadequate representation of ENSO impacts on temperature in this season.

In the case of precipitation, the structural term is reduced in central-eastern Brazil in JJA when the GEM-NEMO model is removed from the MME (Fig. 5a). According to previous works, the precipitation in this region is modulated by the SST anomalies of the tropical North Atlantic Ocean (TNA) (Ronchail et al. 2002; Yoon and Zeng 2009; Towner et al. 2020). The correlation between the observed precipitation in central-eastern Brazil and the SST over the TNA (Fig. 5d) is significant and negative. The correlation for the MME without the GEM-NEMO model has the same sign and similar spatial distribution to the observed correlation (Fig. 5b), but this is not the case for the individual model (Fig. 5c). These differences in the TNA signal on precipitation variability may be causing the changes in variability explained by the structural term, but the influence of local factors cannot be ruled out.

3.3 Potential predictability

The potential predictability (PP_t) was estimated from ANOVA as the ratio between the signal variance and the sum of the signal and noise variances (Eq. 8). The estimate for temperature is shown in Fig. 6. At all seasons, PP_t is

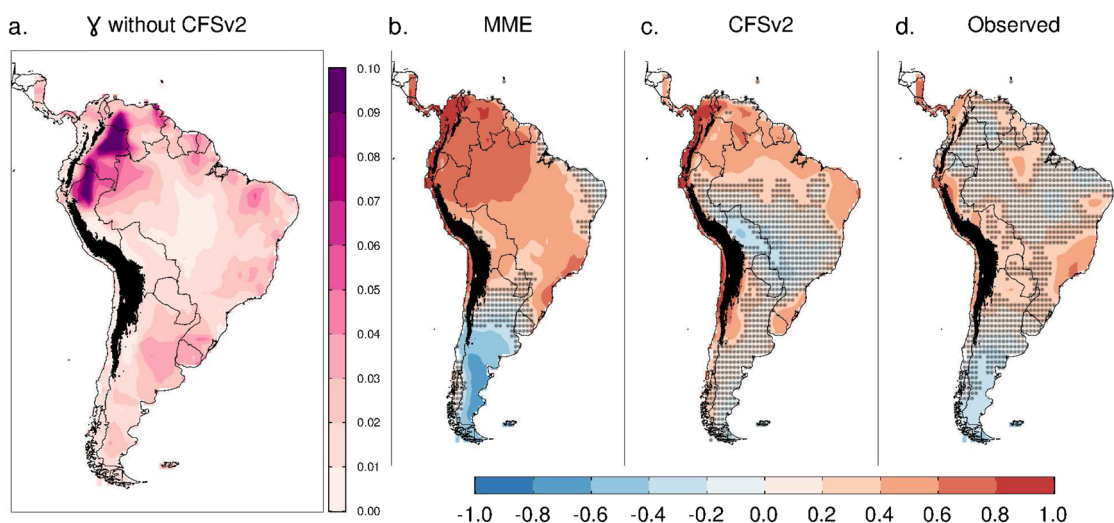


Fig. 4 Fraction of total variance represented by γ_m for the MME without CFSv2 model (a) in SON. Correlation between the observed SST in the Niño3.4 region in August and the forecast mean tempera-

ture in South America in SON for the MME (b), the EMM without the CFSv2 model (c), and the reference data (d)

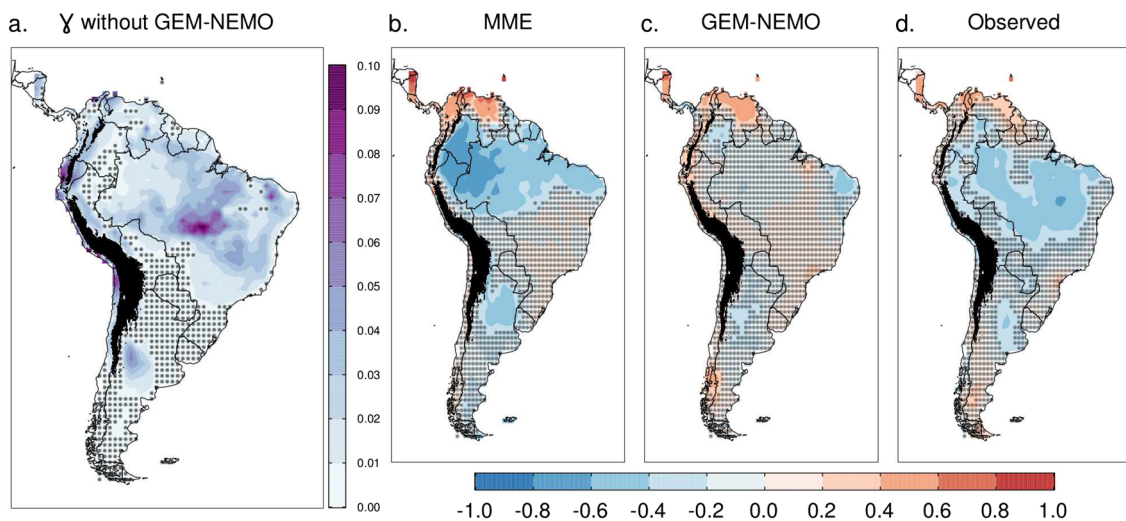


Fig. 5 Same as Fig. 4 but for precipitation, the MME without the GEM-NEMO model in JJA (a), and (b–d) the correlation between the observed SST in the NTA (10°–30° N; 300°–330° E) in May and the

forecast mean precipitation in JJA for the MME (b), the EMM without the GEM-NEMO model (c), and the reference data (d)

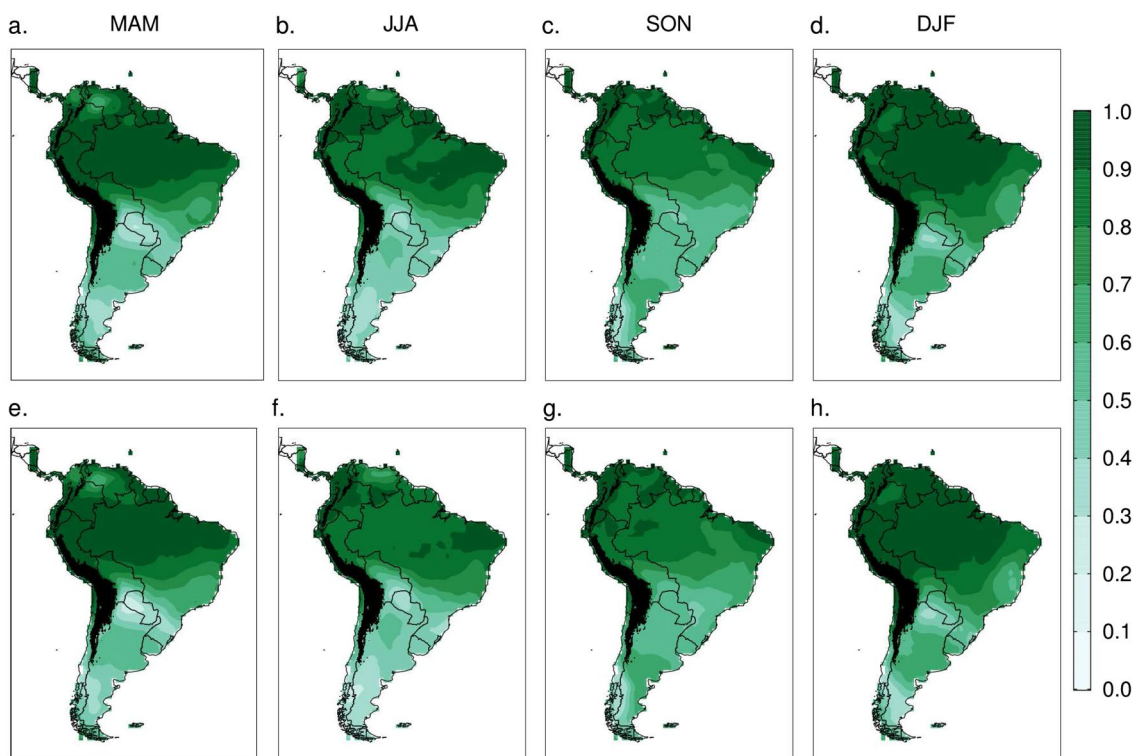


Fig. 6 Temperature potential predictability in MAM, JJA, SON, and DJF for the EMM (a–d) and EMM without the CFSv2 model (e–h)

higher in the tropical region and reaches the maximum values in MAM and DJF, while the lowest values in these regions are observed in SON. At middle and sub-polar latitudes, PP_t is always lower than that found further north. However, relatively high values of PP_t are observed over

central Argentina in MAM, JJA, and DJF, and in particular the highest PP_t value is found in SON and DJF over northern Patagonia and central-eastern Argentina, respectively. Each model impact on PP_t was assessed in the same manner as each term of the variance decomposition. For this

purpose, the PP_t was computed again, but removing each model in turn from the MME. Overall, for temperature PP_t , no important changes were observed when removing individual models from the MME. The most relevant changes, but non significant according to the Monte Carlo test, are observed when the CFSv2 model was removed from the MME (Fig. 6e-h). However, these slight changes do not point in the same direction across the continent. For example, in SON, PP_t increases in Ecuador and northern Peru but decreases in southern Brazil.

Due to the high fraction of variance explained by noise in precipitation, the PP_t is lower than for temperature (Fig. 7). Despite this, highest precipitation PP_t values are observed at low latitudes in all seasons, being maximum for DJF in Amazonia. Unlike temperature PP_t , precipitation PP_t over Bolivia, Peru, Ecuador, and Colombia is among the lowest on the continent in all seasons, with values close to zero in MAM. These lower PP_t values are because the precipitation variability over these regions is dominated by inter-model biases, as in all regions with high topography, and the signal is close to zero (see Fig. 2). At middle and sub-polar latitudes, the PP_t is lower, with relative maxima in southern Brazil in JJA, northern Patagonia in SON, and central-eastern Argentina and Uruguay in DJF. Precipitation PP_t increases when the CFSv2 model is removed from the MME (Fig. 7e-h) because most of the noise in continent-wide

precipitation variability is attributable to this model (not shown). Although these PP_t increases are non significant according to the Monte Carlo test, PP_t values of the MME without the CFSv2 model are higher in all seasons, with the same spatial pattern as when this model is included in the MME.

3.4 Skill

The MME skill was analyzed through the RMSE and ACC. Since the models have systematic errors, the RMSE was corrected by the climatology and normalized by the standard deviation (NRMSE). That gives a measure of how good the approximation of the MME is, compared to the simple approximation of the climatology mean. NRMSE values close to 1 indicate that the errors are lower than the variability and that the error associated with the MME forecast is smaller than the error obtained when using climatology as a forecast. Figure 8 shows the NRMSE of temperature (upper panel) and precipitation (lower panel) for the four seasons. For temperature, negative NRMSE values dominate almost the entire continent, while positive values are found only over northeastern Brazil for all the seasons. For precipitation, NRMSE values are close to zero over almost the entire continent while positive values are found only over

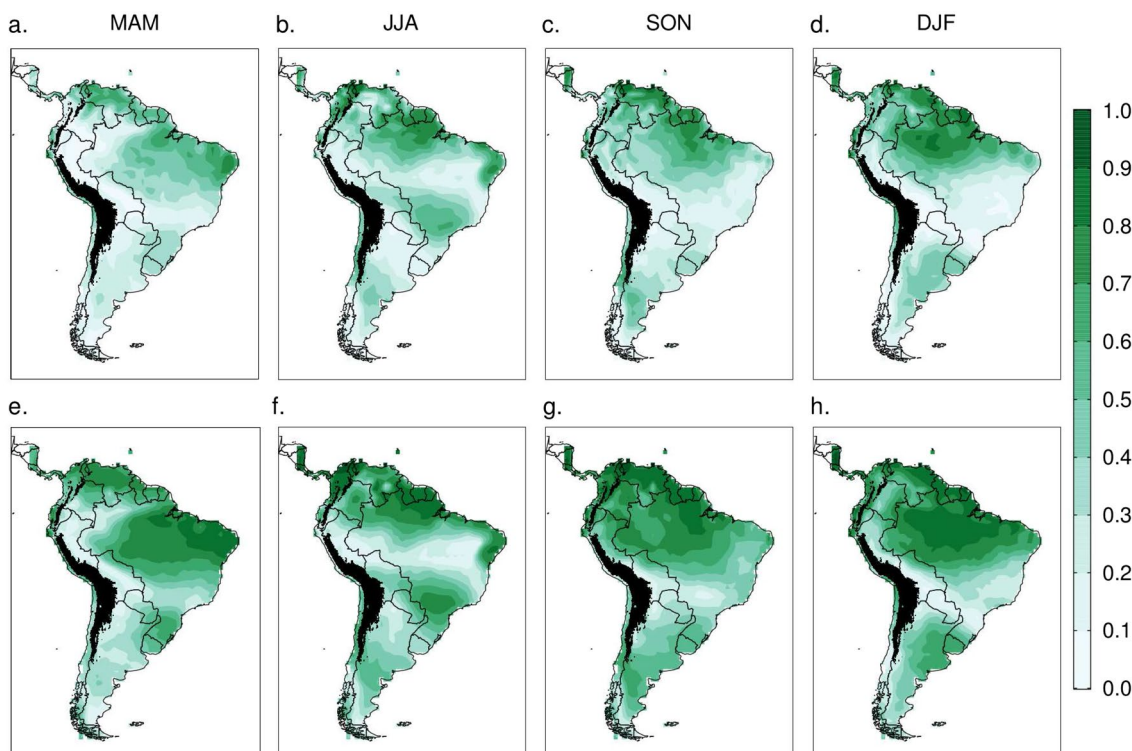


Fig. 7 Same as Fig. 6 but for precipitation

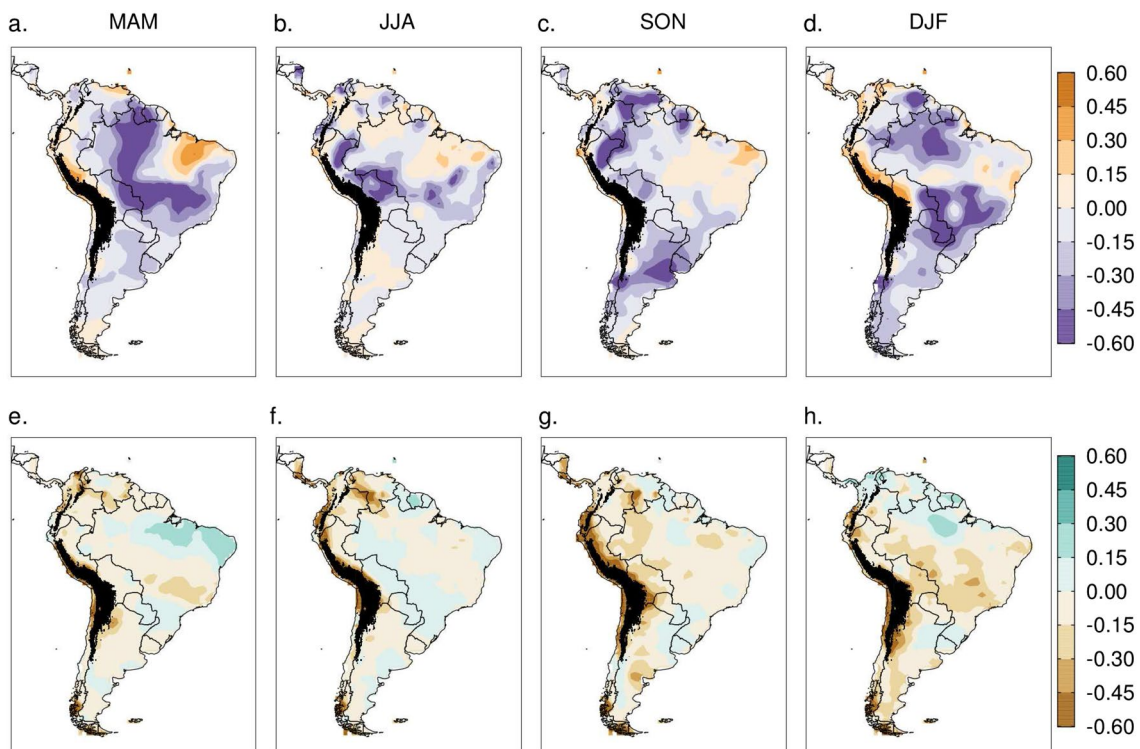


Fig. 8 NRMSE in MAM, JJA, SON, and DJF (from left to right) for temperature (upper panel) and precipitation (bottom panel)

northeastern Brazil in MAM, over SESA in all seasons, and northern Brazil in DJF.

Figure 9 shows the temperature ACC computed between reference data and the MME (skill, top panel); and the theoretical ACC (perfect-model skill, bottom panel) in all seasons. Overall, a better skill (Fig. 9a–d) is observed at tropical regions, which decreases at extratropical latitudes, although with high inter-seasonal variations. Northeastern Brazil shows the highest skill in all four seasons, with a maximum of 0.8 in MAM. It also shows high skill over Peru and northern Bolivia in MAM and DJF. At midlatitudes, modest skill values of 0.5–0.6 are found over SESA in all seasons and are maximum in JJA. At sub-polar latitudes, the highest skill is found in MAM over the southern tip of Patagonia, with values of 0.6–0.7. In JJA and SON at these latitudes, a significant skill is found only over the east coast.

The perfect-model skill (Fig. 9e–h) is higher in the tropical region, and a secondary maximum is found at midlatitudes. Perfect-model skill is always higher than the skill, which means a better agreement between models than between them and observations. However, the differences observed between the skill and perfect-model skill are not significant, according to the Monte Carlo test, over most of the continent (shaded regions). These differences are significant only over Patagonia in MAM and DJF; and over

midlatitudes in JJA and SON. In all seasons, the lowest values of the perfect-model skill are found between 20 °S–30 °S.

For precipitation, the skill (Fig. 10a–d) is lower than for temperature, and most of the continent presents non-significant values. The highest skill is located over the equatorial region, in northeastern Brazil in MAM, which is the maximum of all seasons, and northern Brazil and Colombia in SON and DJF. In JJA the skill is the lowest at these latitudes. Models fail in forecasting precipitation over the Monsoon region in SON and DJF, when the Monsoon develops and peaks, respectively. At midlatitudes, the skill is significant over SESA with values between 0.5–0.7 but with large inter-seasonal variations, and over Bolivia and Paraguay only in JJA. Over Patagonia, the skill is significant over the east coast and is maximum in JJA and SON while, as for temperature, the skill is non-significant in DJF.

The perfect-model skill is higher than the skill (Fig. 10e–h) but with large areas with non-significant values. The region between 10 °S and 20 °S presents non-significant values in all the seasons but JJA. Unlike temperature, it is possible to identify some regions where the skill exceeds the perfect-model skill, such as northeastern Brazil in MAM and SON, but these differences are non-significant (shaded regions) according to the Monte Carlo test.

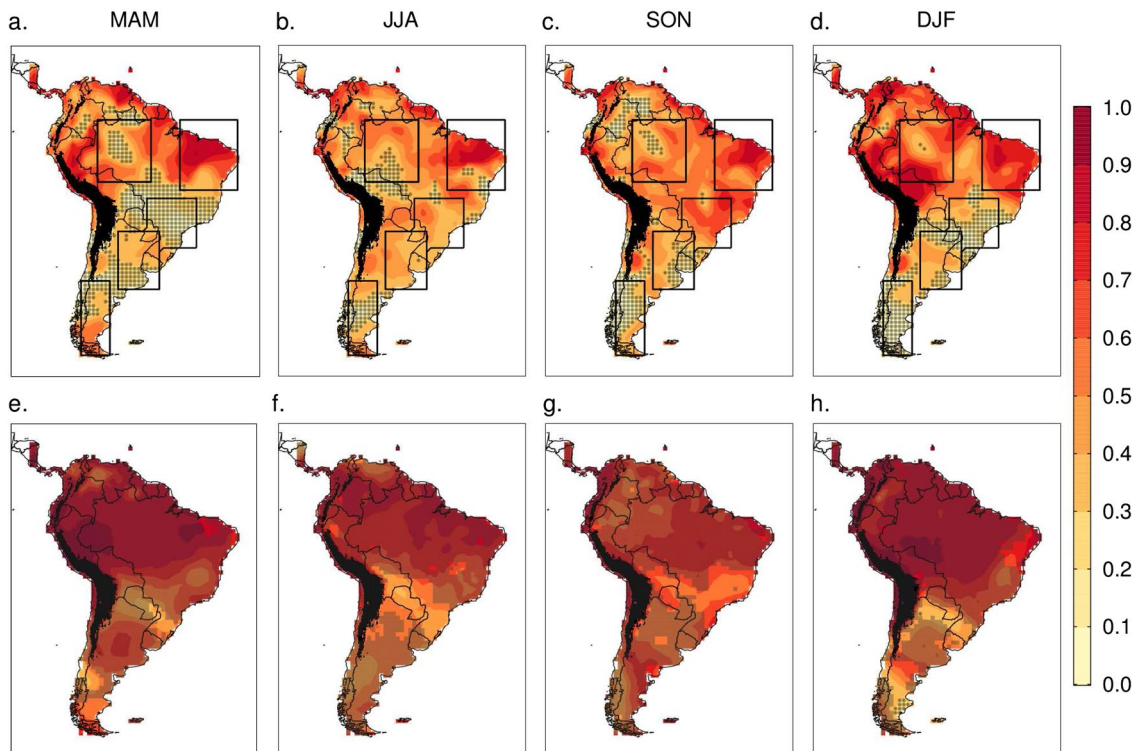


Fig. 9 Anomaly correlation coefficient (ACC) for temperature in MAM, JJA, SON and DJF (from left to right). ACC between the EMM and the reference data (a–d), average of the ACC between MME with each individual model, theoretical ACC (e–h). Dotted areas indicate the grid points where the ACC is negative or non-sig-

nificant positive at the 95% confidence level. Shaded areas denote grid point where the differences between actual and theoretical ACC are not significant according to the Monte Carlo test (see Sect. 2). The rectangles delimit the averaged regions in Fig. 11 (see text)

As in the previous sections, individual MME models were removed to assess their impact on MME skill. In doing so, no major changes in MME skill were observed for either variable. Although removing the CFSv2 model from the MME showed considerable changes in the sources of variability and potential predictability, the MME skill does not show important changes in either direction.

Given that South America encompasses tropical and extratropical regions, the models' skill can be different in each of them. Therefore, the skill was analyzed over regions where potential predictability is maximized according to the previous analysis (rectangles in upper panel of Figs. 9 and 10): (1) Amazonia (Am) (13°S – 2°N ; 291°E – 304°E), (2) northeast Brazil (NeB) (15°S – 2°N ; 311°E – 325°E), (3) north of SESA (N-SESA) (29°S – 17°S ; 303°E – 315°E), 4) south of SESA (S-SESA) (39°S – 25°S ; 296°E – 306°E) y 5) Patagonia (73°S – 37°S ; 287°E – 294°E). Figure 11 shows the MME skill in the four seasons, measured as the average of ACC over the study regions. The horizontal gray line indicates the 95% confidence level according to a t-test. In the case of temperature (Fig. 11a), although on some occasions the MME has lower skill than the individual models, overall the MME skill is among the best. This is important since no

model always has the best or worst skill at all regions and seasons. Moreover, some models may have the best skill in one season and the worst in another, such as the FLOR-B01 model (model number 4 in the figure) in MAM and SON in Patagonia.

For precipitation (Fig. 11b), results are like temperature but with lower values of skill, which results in most of the regions with non-significant values in most of the seasons. The skill of the MME is among the best at almost all regions and seasons, and no model stands out with the worst or best skill.

4 Conclusions and discussion

In this paper, we analyzed the variance, predictability, and skill of the North American Multi-Model Ensemble (NMME) seasonal forecasts of temperature and precipitation in South America. In addition, we studied for the first time in South America the predictability and performance of an ensemble of CGCMs in the transitional seasons. An ANOVA was applied to analyze the contribution of signal,

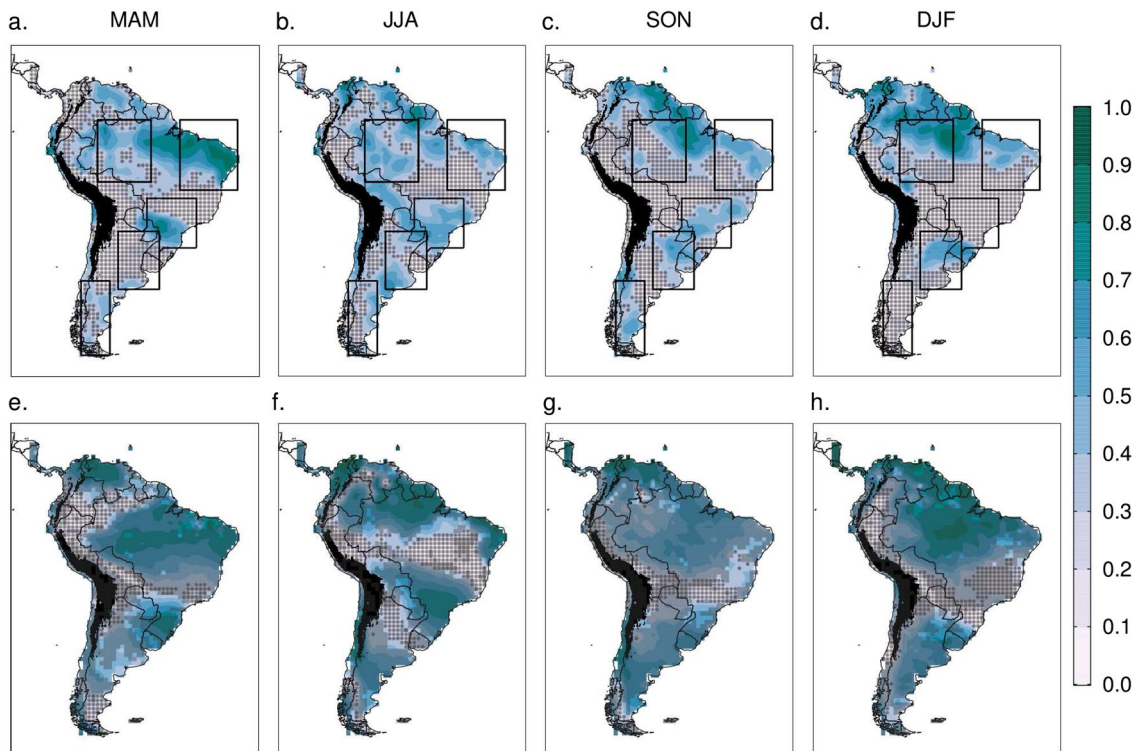


Fig. 10 Same as Fig. 9 but for precipitation

inter-model biases, differences in response to initial conditions, and noise to the total variance of both variables.

For temperature, two different behaviors were identified in the analysis of the sources of variance throughout the year. In the northern part of the continent, the signal explained around 70% of the total variance in MAM and DJF, whereas near 60% of the total variance is explained by the inter-model biases in JJA and SON. On the other hand, in SESA the signal dominates the variance in MAM and JJA, and the inter-model biases dominate it in SON and DJF. The results obtained for DJF and JJA agree with Osman and Vera (2017) using the CHFP models. For the precipitation, as was reported by previous works (Barreiro et al. 2002; Osman and Vera 2017; Bombardi et al. 2018), noise explains around 60% of the variance in all seasons. We also found that the use of the ANOVA technique was adequate for a better understanding of the variability of the models and helped document the effect of model biases and the different response to initial conditions in the MME variability.

The ANOVA was also used to study the potential predictability (PP_t). For both variables, the highest PP_t was found in the tropics throughout the year and decreases towards higher latitudes, while the temperature PP_t is considerably higher than precipitation PP_t . Although the signal peaks in regions where precipitation is modulated by ENSO in MAM, JJA, and DJF (Cai et al. 2020), this is only reflected in PP_t

at low latitudes due to the lower noise in that region. These two results are consistent with those previously found by Wu and Kirtman (2006), Peng et al. (2009) and Osman and Vera (2017).

Finally, the NMME performance was analyzed using the ACC. Similar to previous studies (e.g. Becker et al. 2014; Osman and Vera 2017), the best skill for both variables was found at low latitudes, and the temperature skill was better than precipitation skill. In addition, forecasting precipitation over the Monsoon region is still challenging, as the skill found in this work does not improved from what was previously reported in Becker et al. (2014); Bombardi et al. (2018). Finally, as was observed for DJF in previous works (e.g. Coelho et al. 2006; Gubler et al. 2020), the highest skill in the four seasons was found in regions where both variables are modulated by ENSO.

To our knowledge, the most recent study documenting the predictability of a MME in South America is Osman and Vera (2017) using the CHFP models. We noticed that one major difference between our work and the former is in the position of maximum values of precipitation noise, towards high latitudes for NMME and in the tropical regions for CHFP. In both ensembles the spatial pattern of PP_t is similar although NMME models show higher values for temperature and smaller for precipitation. For both variables, the skill of the NMME models is better

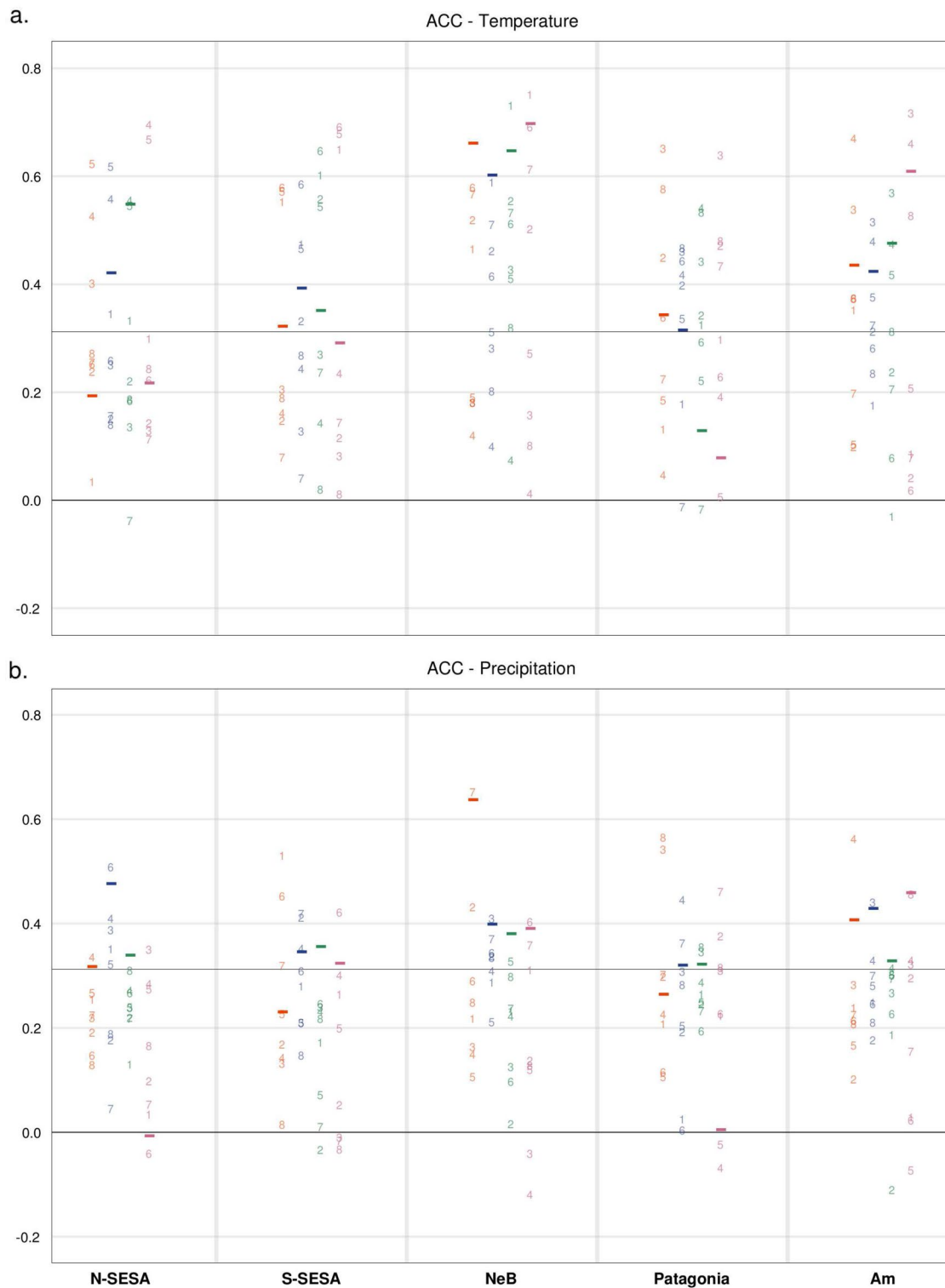


Fig. 11 Average ACC for temperature (upper panel) and precipitation (bottom panel) in north and south SESA (N-SESA, S-SESA), northeast of Brazil (NeB), Patagonia and Amazonia (Am) for the models (numbers) and mean MME (bold hyphen) in MAM, JJA, SON, and

DJF (from left to right). The horizontal gray line indicates the significance limit at the 95% confidence level. CCSM4(1), CM2p1(2), FLOR-A06(3), FLOR-B01(4), GEOS2S(5), CFSv2(6), CanCM4i(7) and GEM-NEMO(8)

than that achieved by the CHFP models in JJA and similar in DJF. Furthermore, NMME presents a higher theoretical ACC than CHFP, which indicates a better agreement between the models in NMME than in CHFP. Comparing the NMME version used in this work to the previous one (Becker et al. 2014), the temperature skill is better in JJA and SON and similar for precipitation in all seasons. In addition, since no model was found to have the best or worst performance in all regions and seasons, the MME skill proves to be one of the best compared to the individual model skill. Finally, as in Becker et al. (2014), we find that the dispersion between the individual model skill is much smaller for precipitation than for temperature.

We conclude that the NMME represents an advance over other model ensembles such as CHFP, not only because it is an operational model ensemble which has retrospective and real-time forecasts produced every month with a lead time of up to 1 year, but also because it exhibits a better skill, at least for JJA and SON temperature. Moreover, considering that NMME provides access to all data in near real-time, it has potential as an operational seasonal forecasting tool in South America.

Although this study has documented the predictability and performance of NMME models, there are still further investigations that could be carried out in the future. These include the assessment of the predictability and skill of the models when they are initialized several months in advance and the evaluation the probabilistic prediction performance of this ensemble.

Acknowledgements We thank the two anonymous reviewers for their comments. We acknowledge the agencies that support the NMME-Phase II system, and we thank the climate modeling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP, and University of Miami) for producing and making available their model output. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system.

Author Contributions LGA: data curation; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. MO: conceptualization; data curation; supervision; methodology; writing—review and editing. CSV: conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing—review and editing.

Funding The research was supported by UBACyT20020170100428BA, PDE 46 2019, PICT-2018-03046, CLIMAT-AMSUD 21-CLIMAT-05 and the CLIMAX Project funded by Belmont Forum/ANR-15-JCL/0002-01. LGA is supported by a fellowship grant from CONICET.

Data availability Raw data can be found at <http://iridl.ldeo.columbia.edu/SOURCES/Models/NMME/>. Data supporting the findings of this article is available upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interests that could have influenced the work reported in this paper.

Ethics approval Not applicable.

Code availability All codes used in this work can be found at <https://github.com/LucianoAndrian/tesis>

References

- Barreiro M (2010) Influence of ENSO and the south Atlantic ocean on climate predictability over southeastern South America. *Clim Dyn* 35:1493–1508
- Barreiro M, Chang P, Saravanan R (2002) Variability of the south Atlantic convergence zone simulated by an atmospheric general circulation model. *J Clim* 15(7):745–763
- Becker E, van den Dool H, Zhang Q (2014) Predictability and forecast skill in nmme. *J Clim* 27(15):5891–5906
- Becker E, Kirtman B, Pegion K (2020) Evolution of the north American multi-model ensemble (nmme). *Geophys Res Lett* 47
- Bombardi RJ, Trenary L, Pegion K, Cash B, DelSole T, Kinter JL (2018) Seasonal predictability of summer rainfall over south America. *J Clim* 31(20):8181–8195
- Cai W, McPhaden M, Grimm A, Rodrigues R, Taschetto A, Garreaud R, Dewitte B, Poveda G, Ham Y-G, Santoso A, Ng B, Anderson W, Wang G, Geng T, Jo H-S, Marengo J, Alves L, Osman M, Li S, Vera C (2020) Climate impacts of the el niño-southern oscillation on south America. *Nat Rev Earth Environ* 1:215–231
- Campitelli E (2021) metR: tools for easier analysis of meteorological fields. R package version 0.10.0
- Cash B, Burls N (2019) Predictable and unpredictable aspects of us west coast rainfall and el niño: understanding the 2015–2016 event. *J Clim* 32(10):2843–2868
- Cash B, Manganello J, Kinter J (2019) Evaluation of nmme temperature and precipitation bias and forecast skill for south Asia. *Clim Dyn* 53:7363–7380
- Coelho CAS, Stephenson DB, Doblas-Reyes FJ, Balmaseda M (2006) The skill of empirical and combined/calibrated coupled multi-model south American seasonal predictions during ENSO. *Adv Geosci* 6:51–55
- Fan Y, van den Dool H (2008) A global monthly land surface air temperature analysis for 1948–present. *J Geophys Res*. <https://doi.org/10.1029/2007JD008470>
- Gubler S, Sedlmeier K, Bhend J, Avalos G, Coelho CAS, Escajadillo Y, Jacques-Coper M, Martinez R, Schwierz C, de Skansi M, Spirig C (2020) Assessment of ecmwf seas5 seasonal forecast performance over south America. *Weather Forecast* 35(2):561–584
- Hodson D, Sutton R (2008) Exploring multi-model atmospheric gcm ensembles with ANOVA. *Clim Dyn* 31:973–986
- Kirtman BP, Min D, Infanti JM, Kinter James LI, Paolino DA, Zhang Q, van den Dool H, Saha S, Mendez MP, Becker E, Peng P, Tripp P, Huang J, DeWitt DG, Tippett MK, Barnston AG, Li S, Rosati A, Schubert SD, Rienecker M, Suarez M, Li ZE, Marshak J, Lim Y-K, Tribbia J, Pegion K, Merryfield WJ, Denis B, Wood EF (2014) The north American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull Am Meteorol Soc* 95(4):585–601
- Kumar A, Peng P, Chen M (2014) Is there a relationship between potential and actual skill? *Mon Weather Rev* 142:2220–2227

- Lin H, Merryfield W, Muncaster R, Smith G, Markovic M, Dupont F, Roy F, Lemieux J.-F, Dirkson A, Kharin V, Lee W-S, Charron M, Erfani A (2020) The Canadian seasonal to interannual prediction system version 2 (cansipsv2). *Weather Forecast*. <https://doi.org/10.1175/WAF-D-19-0259.1>
- Nobre P, Marengo JA, Cavalcanti IFA, Obregon G, Barros V, Camilioni I, Campos N, Ferreira AG (2006) Seasonal-to-decadal predictability and prediction of south American climate. *J Clim* 19(23):5988–6004
- Nychka D, Furrer R, Paige J, Sain S (2017) *fields: tools for spatial data*. R package version 12:3
- Osman M, Vera C (2017) Climate predictability and prediction skill on seasonal time scales over south America from chfp models. *Clim Dyn* 49:2365–2383
- Osman M, Vera CS (2020) Predictability of extratropical upper-tropospheric circulation in the southern hemisphere by its main modes of variability. *J Clim* 33(4):1405–1421
- Osman M, Coelho CAS, Vera CS (2021) Calibration and combination of seasonal precipitation forecasts over South America using Ensemble Regression. *Clim Dyn* 57:2889–2904
- Peng P, Kumar A, Wang W (2009) An analysis of seasonal predictability in coupled model forecasts. *Clim Dyn* 36:637–648
- R Core Team (2020) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria
- Reynolds R, Smith T, Liu C, Chelton D, Casey K (2007) Daily high-resolution-blended analyses for sea surface temperature. *J Clim* 20(22):5473–5496
- Rodrigues LRL, Doblas-Reyes FJ, Coelho CAS (2019) Calibration and combination of monthly near-surface temperature and precipitation predictions over Europe. *Clim Dyn* 53(12):7305–7320
- Ronchail J, Cochonneau G, Molinier M, Guyot J-L, De Miranda Chaves AG, Guimarães V, de Oliveira E (2002) Interannual rainfall variability in the amazon basin and sea-surface temperatures in the equatorial pacific and the tropical atlantic oceans. *Int J Climatol* 22(13):1663–1686
- Slater LJ, Villarini G, Bradley AA (2019) Evaluation of the skill of north-American multi-model ensemble (nmme) global climate models in predicting average and extreme precipitation and temperature over the continental USA. *Clim Dyn* 53(12):7381–7396
- Storch HV, Zwiers FW (1999) *Analysis of variance. Statistical analysis in climate research*. Cambridge University Press, Cambridge, pp 171–192
- Taschetto A, Wainer I (2008) Reproducibility of south American precipitation due to subtropical south Atlantic ssts. *J Clim* 21:2835–2851
- Tompkins AM, Zárata MIOD, Saurral RI, Vera C, Saulo C, Merryfield WJ, Sigmond M, Lee W-S, Baehr J, Braun A, Butler A, Déqué M, Doblas-Reyes FJ, Gordon M, Scaife AA, Imada Y, Ishii M, Ose T, Kirtman B, Kumar A, Müller WA, Pirani A, Stockdale T, Rixen M, Yasuda T (2017) The climate-system historical forecast project: Providing open access to seasonal forecast ensembles from centers around the globe. *Bull Am Meteorol Soc* 98(11):2293–2301
- Towner J, Cloke HL, Lavado W, Santini W, Bazo J, Coughlan de Perez E, Stephens EM (2020) Attribution of amazon floods to modes of climate variability: A review. *Meteorol Appl* 27(5):e1949
- Wilks DS (1995) *Statistical methods in the atmospheric sciences*. Academic Press, San Diego
- Wu R, Kirtman B (2006) Changes in spread and predictability associated with ENSO in an ensemble coupled GCM. *J Clim* 19(17):4378–4396
- Xie P, Arkin PA (1997) Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull Am Meteorol Soc* 78(11):2539–2558
- Yoon J-H, Zeng N (2009) An Atlantic influence on amazon rainfall. *Clim Dyn* 34:249–264
- Zhao T, Zhang Y, Chen X (2019) Predictive performance of nmme seasonal forecasts of global precipitation: a spatial-temporal perspective. *J Hydrol* 570:17–25
- Zwiers F (1996) Interannual variability and predictability in an ensemble of amip climate simulations conducted with the ccc gcm2. *Clim Dyn* 12:825–847
- Zwiers FW, Wang XL, Sheng J (2000) Effects of specifying bottom boundary conditions in an ensemble of atmospheric gcm simulations. *J Geophys Res Atmos* 105(D6):7295–7315

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.